

**Proteome, transcriptome and metabolome  
plasticity in closely related strains of  
*Escherichia coli*- K12 and its diversity  
during molecular evolution processes.**

Chandran Vijayendran

A Thesis submitted for the degree of  
Doctor of Natural Sciences (Dr. rer. nat.)

Faculty of Technology  
Bielefeld University

Bielefeld, March 2007

This thesis is dedicated to *M.K. Bellie* for years of encouragement and motivation.



---

# CONTENTS

---

1	Prologue	vii
I	The plasticity of global proteome and genome expression analyzed in closely related W3110 and MG1655 strains of a well studied model organism, <i>Escherichia coli</i> - K12.	x
2	Introduction (I)	1
3	Materials and Methods (I)	4
3.1	Strains and medium . . . . .	4
3.2	Two-dimensional SDS-PAGE gel electrophoresis . . . . .	4
3.3	In-gel tryptic digestion and mass spectrometry . . . . .	4
3.4	Microarray experiments . . . . .	5
3.5	DNA microarray analysis . . . . .	5
3.6	Microarray data analysis . . . . .	5
3.7	Real-time RT-PCR . . . . .	5
3.8	Determination of siderophores . . . . .	6
3.9	Cloning and sequencing of <i>rpoS</i> . . . . .	6
4	Results and Discussion (I)	7
4.1	Expression of RpoS-dependent genes . . . . .	7
4.2	Expression of energy metabolism genes . . . . .	10
4.3	Expression of iron-related genes . . . . .	14
4.4	Verification of microarray results . . . . .	14
4.5	Effect of <i>rpoS</i> gene amber mutation in gene expression . . . . .	16
4.6	Expression of cellular proteins . . . . .	17
5	Conclusion (I)	18
5.1	Plasticity of gene and protein expression . . . . .	18
II	Perceiving molecular evolution processes in <i>Escherichia coli</i> by comprehensive metabolite and gene expression profiling.	20
6	Introduction (II)	22
7	Materials and Methods (II)	24
7.1	Strains and culture conditions . . . . .	24
7.2	Metabolite profiling . . . . .	24
7.3	Gene expression profiling . . . . .	25

7.4	Microarray data analysis . . . . .	25
7.5	Network analysis . . . . .	25
7.6	Network functional analysis . . . . .	26
7.7	Outer-membrane protein analysis . . . . .	26
<b>8</b>	<b>Results (II)</b>	<b>27</b>
8.1	Metabolome profiling . . . . .	27
8.2	Gene expression profiling . . . . .	31
8.3	Extent of changes . . . . .	34
8.4	Direction of the observed extent of changes . . . . .	34
8.5	Metabolite to gene correlation network analysis . . . . .	38
8.6	Evaluation of evolution specific networks . . . . .	39
8.7	Evolution specific intersection networks . . . . .	40
8.8	Parallelism and functional relevance of molecular evolution . . . . .	40
8.9	Parallelism in outer-membrane protein expression . . . . .	42
<b>9</b>	<b>Discussion (II)</b>	<b>44</b>
<b>10</b>	<b>Conclusion (II)</b>	<b>50</b>
<b>III</b>	<b>Integration of proteome with transcriptome and metabolome profile during molecular evolution processes in <i>Escherichia coli</i>.</b>	<b>51</b>
<b>11</b>	<b>Introduction (III)</b>	<b>53</b>
<b>12</b>	<b>Materials and Methods (III)</b>	<b>55</b>
12.1	Strain and culture conditions . . . . .	55
12.2	Two-dimensional SDS-PAGE gel electrophoresis . . . . .	55
12.3	In-gel tryptic digestion and mass spectrometry . . . . .	56
12.4	Analysis of two-dimensional protein gels . . . . .	56
12.5	Network analysis . . . . .	57
<b>13</b>	<b>Results (III)</b>	<b>58</b>
13.1	Proteome profiles of ancestral and evolved lines . . . . .	58
13.2	Changes in transcript and metabolite levels . . . . .	59
13.3	Changes in protein expression levels in experimental evolutionary conditions . . . . .	61
13.4	Direction of the observed protein expression changes . . . . .	64
13.5	Protein correlation network analysis . . . . .	64
13.6	Comparison of transcript, metabolite and protein levels . . . . .	68
13.7	Comparison of parallelism in the three functional levels . . . . .	70
<b>14</b>	<b>Discussion (III)</b>	<b>72</b>
<b>15</b>	<b>Conclusion (III)</b>	<b>74</b>

<b>IV 2DBase: 2D-PAGE Database of <i>Escherichia coli</i>.</b>	<b>75</b>
16 Introduction (IV)	77
17 Description (IV)	79
18 Implementation & Conclusion (IV)	81
19 Epilogue	84
Bibliography	95
Supplementary Materials	96
Abbreviations	123
Acknowledgments	124

---

# LIST OF FIGURES

---

1.1	Scanning electron microscopy image of <i>E. coli</i> . . . . .	vii
2.1	Origins of <i>E. coli</i> K-12 strains MG1655 and W3110. . . . .	2
2.2	Growth pattern and genome structure analyses. . . . .	3
4.1	Frequency histogram of gene expression data from W3110 and MG1655 based on $\log_2$ ratios. . . . .	7
4.2	Comparative proteome analyses of the <i>E. coli</i> sub-strains W3110 and MG1655. . . . .	9
4.3	Two-dimensional hierarchical clustering of gene expression datasets from W3110 and MG1655 . . . . .	11
4.4	Development of gene expression during cultivation of <i>E. coli</i> sub-strains W3110 and MG1655 . . . . .	12
4.5	Principal components analysis (PCA) of RpoS-regulated genes in W3110 and MG1655. . . . .	13
4.6	Comparison of the genome sequence fragment of W3110 and MG1655. . . . .	13
4.7	Additional confirming analyses. . . . .	15
4.8	Real-time RT-PCR analyses. . . . .	16
4.9	Comparative OsmY protein analyses of the <i>E. coli</i> sub-strains W3110 and MG1655. . . . .	17
5.1	Global comparison of gene expression in <i>E. coli</i> sub-strains W3110 and MG1655. . . . .	18
7.1	Gene expression among the biological replicates. . . . .	26
8.1	Overlay heat map of the metabolite profiles. . . . .	28
8.2	Typical examples of metabolite differential levels among the ancestral and evolved lines. . . . .	29
8.3	Venn diagrams comparing number of genes and metabolites changes. . . . .	30
8.4	The levels of metabolites involved in the TCA cycle and the diversion of key intermediates to biosynthetic pathways. . . . .	31
8.5	Broad functional annotations of the transcriptome profiling data. . . . .	32
8.6	Pentose phosphate pathway in excess-nutrient adapted strains. . . . .	33
8.7	Integration of transcriptome and metabolome data during the comparison of MG and MGAdp strains in excess-nutrient adaptive evolution. . . . .	35
8.8	Integration of transcriptome and metabolome data during the comparison of DH and DHAdp strains in excess-nutrient adaptive evolution. . . . .	36
8.9	The extent of changes in experimental evolution among the strains. . . . .	37
8.10	Metabolite-to-gene correlation network analyses. . . . .	38

8.11	Parallelism and functional relevance of molecular evolution. . . . .	41
8.12	Parallelism and functional significance in the outer-membrane protein expression. . . . .	43
9.1	Histidine biosynthesis pathway in excess-nutrient adapted strains.	45
9.2	Integration of transcriptome and metabolome data during the comparison of MG and MGStat strains in prolonged stationary phase evolution. . . . .	46
9.3	Integration of transcriptome and metabolome data during the comparison of DH and DHStat strains in prolonged stationary phase evolution. . . . .	47
13.1	Flowchart of the experimental scheme. . . . .	58
13.2	Broad functional annotations of the transcriptome and metabolome profiling data. . . . .	60
13.3	Typical examples of differential protein expression among the ancestral and evolved lines. . . . .	62
13.4	Hierarchical clustering of protein expression levels of ancestral and evolved strains. . . . .	63
13.5	Comparison of direction of changes in the experimental evolution among the strains. . . . .	65
13.6	Protein correlation network analyses. . . . .	67
13.7	Protein, gene and metabolite levels in the central metabolic routes and the diversion of key intermediates to biosynthetic pathways.	69
13.8	Evolutionary trends of protein, transcript and metabolite levels. .	71
16.1	Data flow diagram of 2DBase of <i>Escherichia coli</i> . . . . .	77
17.1	CAS-PLUS and CAS-MINUS 2Dgels. . . . .	79
17.2	MultiFun- the classification system for cellular functions of gene products of <i>Escherichia coli</i> . . . . .	80
18.1	MultiFun- metabolism- functional category. . . . .	81
18.2	MultiFun- transport- functional category. . . . .	82
18.3	Multiple gels comparison. . . . .	83
19.1	Integration of transcriptome and metabolome data during the comparison of MGGal and MGAdpGal strains in environmental shift evolutionary condition. . . . .	97
19.2	Integration of transcriptome and metabolome data during the comparison of DHGal and DHAdpGal strains in environmental shift evolutionary condition.. . . .	98

---

# LIST OF TABLES

---

1.1	Taxonomy of <i>Escherichia coli</i> . . . . .	vii
6.1	Strains and their evolved conditions. . . . .	23
11.1	Strains and their evolved conditions. . . . .	54
19.1	Primers used for amplification. . . . .	96
19.2	Expression data of RpoS dependent and osmotic genes during exponential phase. . . . .	101
19.3	Expression data of energy metabolism genes during exponential phase. . . . .	103
19.4	Expression data of iron related metabolism genes during exponential phase. . . . .	104
19.5	Expression data of RpoS dependent genes during exponential phase versus early stationary phase. . . . .	106
19.6	Transcriptome profiling data. . . . .	107
19.7	Metabolome profiling data. . . . .	115
19.8	Proteome profiling data. . . . .	116

---

# CHAPTER 1

## PROLOGUE

---

*Escherichia coli*, originally known as *Bacterium coli* commune, was identified in 1885 by the German pediatrician, Theodor Escherich. They are Gram-negative, straight and rod-shaped bacteria from the *Enterobacteriaceae* family (**Figure 1.1** and **Table 1.1**).

<b>Phylum:</b>	Proteobacteria
<b>Class:</b>	Gamma Proteobacteria
<b>Order:</b>	Enterobacteriales
<b>Family:</b>	Enterobacteriaceae
<b>Genus:</b>	<i>Escherichia</i>
<b>Species:</b>	<i>E. coli</i>

Tab. 1.1: Taxonomy of *Escherichia coli*.

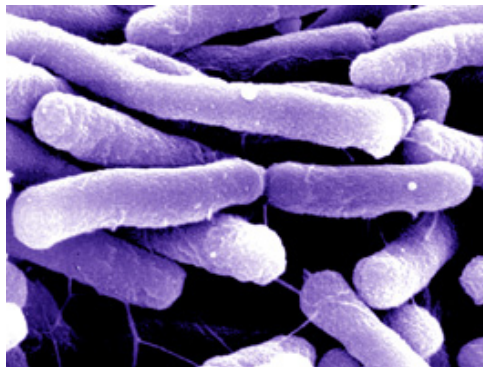


Fig. 1.1: Scanning electron microscopy image of *E. coli*.

Due to its rapid growth rate, simple nutritional requirements and established genetic manipulation techniques, *E. coli* has become a model organism in the field of basic biomolecular sciences for understanding various biological phenomena. As a result, well established information about *E. coli*'s genetics and several completed genome sequences are available. Among them, the genome sequences of two closely related K-12 non-pathogenic strains, MG1655 and W3110 have been accurately determined [1, 2]. Re-sequencing of PCR products of selected regions indicates that there are only eight true insertion/deletion or base differences between the two strains in addition to the 13 sites where differences are due to insertion sequences, defective prophages and two sites due to the W3110 inversion between the ribosomal RNA genes *rrnD* and *rrnE* [3]. The rate of nucleotide changes between both the strains is estimated to be relatively low with almost identical genome structures [3, 4]. Hence the following important questions arise:

- *Is this high degree of similarity at the nucleotide level reflected in the metabolic phenotype?*
- *Do sub-strains with almost identical genome structures exhibit similar behaviour in cellular metabolism?*
- *How do global aspects of cell metabolism, protein synthesis and gene expression differ among closely related sub-strains of the same species, revealing possible complexities of cellular metabolism?*

To address these queries, we analyzed the growth behaviour in strictly controlled conditions and analysed the global proteome and transcriptome pools of these closely related *E.coli* sub-strains W3110 and MG1655. We applied the conventional 2-dimensional polyacrylamide gel electrophoresis for global proteomic profiling which is still the major method for global proteome analysis. Global changes in the gene expression levels were analysed using microarrays, thus providing quantitative information about the gene expression levels.

Being the most extensively studied model organism, *E.coli* is frequently used in molecular evolutionary studies. A few potential reasons for this are: its capacity to propagate and reproduce quickly, facilitating the evolution experiments for many generations in a short time span, and capability to store the evolved and ancestor strains, allowing for direct comparison between them. As a result, several studies have used gene expression [5, 6, 7, 8] and proteome profiling methods [9, 10, 11, 12] to study molecular evolution, but these studies were confined to a single type of evolution process and were focused on a single molecular aspect that characterizes a cell (transcript or protein abundance). Metabolome profiling has been frequently applied for obtaining quantitative information on metabolites for studies on mutational or environmental effects, but not in an evolutionary context [13, 14].

In our study, we depicted a complete picture of molecular evolution processes in the laboratory among the two strains MG1655 and DH10B under three different evolutionary conditions in all three functional levels of the cell (transcriptome, proteome and metabolome). These data sets obtained from the three functional levels would be of vital importance for viewing a global picture of the experimental sample in question. To eliminate the possibility of the strain-dependent phenomenon of evolution and to examine the parallelism of the laboratory evolution processes, we examined all the evolutionary processes in two strains. The major questions that arose during our study were:

- *What are the transcriptome, proteome and metabolome changes occurring during the excess-nutrient adaptive evolution process?*



- *Which genes, proteins and metabolites are vitally involved in the prolonged stationary phase evolution process?*
- *What are the transcript, protein and metabolite changes occurring due to the pleiotropic effects due to environmental shift?*
- *To what extent are the changes occurring during these evolutionary processes seen in both strains?*
- *Among both the strains, is the path of evolution similar in these evolutionary processes (parallelism)?*

By global protein profiling technologies and integrating the multidimensional datasets generated, we aimed to find vital genes, proteins and metabolites involved in the evolutionary processes in three conditions in two *E. coli* K-12 strains. The importance of making the proteome data available to the public was taken care of, by constructing a relational database having the ability to query and to re-analyse that data with valid functional classification through a web interface. Metabolome data will be available for download from the project web page. Transcriptome data will be submitted to a public repository database of gene expression data. These generated datasets from all the three functional levels would be an initial resource for the systems biology of microbial evolution.

# Part I

The plasticity of global proteome and genome expression analyzed in closely related W3110 and MG1655 strains of a well studied model organism, *Escherichia coli* - K12.

## Abstract

The use of *Escherichia coli* as a model organism has provided a great deal of basic information in biomolecular sciences. Examining trait differences among closely-related strains of the same species addresses a fundamental biological question: how much diversity is there at the single species level? The main aim of our research was to identify significant differences in the activities of groups of genes between two laboratory strains of an organism closely related in genome structure. We demonstrate that despite strict and controlled growth conditions, there is high plasticity in the global proteome and genome expression in two closely-related *E. coli* K-12 sub-strains (W3110 and MG1655), which differ insignificantly in genome structure. The growth patterns of these two sub-strains were very similar in a well-equipped bioreactor, and their genome structures were shown to be almost identical by DNA microarray. However, detailed profiling of protein and gene expression by 2-dimensional gel electrophoresis and microarray analysis showed many differentially-expressed genes and proteins, combinations of which were highly correlated. The differentially regulated genes and proteins belonged to the following functional categories: genes regulated by RpoS (sigma subunit of RNA polymerase), enterobactin-related genes, and genes involved in central metabolism. Genes involved in central cell metabolism - the glycolysis pathway, the tricarboxylic acid cycle and the glyoxylate bypass - were differentially regulated at both the mRNA and proteome levels. The strains differ significantly in central metabolism and thus in the generation of precursor metabolites and energy. This high plasticity probably represents a universal feature of metabolic activities in closely-related species, and has the potential to reveal differences in regulatory networks. We suggest that unless care is taken in the choice of strains for any validating experiment, the results might be misleading.

**Authors:** Chandran Vijayendran, Tino Polen, Volker F. Wendisch, Karl Friehs, Karsten Niehaus and Erwin Flaschel.

**Author's contribution:** CV conducted all the experiments cited in this study, analysed the results and wrote this report. TP was involved in microarray experiments. VFW is the scientist in whose laboratory microarray experiments were conducted. KF was involved in manuscript preparation. KN was involved in experimental design. EF is the scientist in whose laboratory these experiments were conducted and was involved in experimental design and manuscript preparation.

---

# CHAPTER 2

## INTRODUCTION (I)

---

Currently, more than 300 completely sequenced genomes are publicly available and more than four times that number are in the process of being sequenced [15, 16]. The genomes of closely-related species are of particular interest since they lead to the elucidation of genome diversity, evolution and pathogenicity. In the case of the proteobacterium *Escherichia coli*, five genomes have been determined and thirteen projects are in progress [15, 16]. Among the completed *E. coli* genomes, two are from non-pathogenic strains, MG1655 and W3110, which are our loci of interest in this study.

Many subtly different strains are in daily use in numerous laboratories for validating new analytical techniques, provoking critical and unanswered questions such as: do sub-strains with almost identical genome structures exhibit similar behaviour in cellular metabolism? Although *E. coli* is such a well-studied model organism, and many of the strains in use around the world differ from the strain of origin only in known genetic characteristics, many more differences remain to be explored. A justifiable question is: why do closely-related sub-strains of *E. coli* grow with different efficiencies on the same media, when the characterized genetic differences between them do not affect the growth pattern? Furthermore, there is a surprising lack of information on how global aspects of cell metabolism, protein synthesis and gene expression differ among closely related sub-strains of the same species, which may reveal the complexities of cellular metabolism. To address this lack of information we analyzed the growth, the proteomes and the transcriptome pools of the frequently-used laboratory *E. coli* sub-strains W3110 and MG1655, which originated from W1485, a derivative of *E. coli* K-12 (**Figure 2.1**) [17]. The whole genome sequences of both sub-strains are publicly available [1, 2]. Their growth patterns and genome structures are almost identical (**Figure 2.2**).

Most of the major differences in genome structures between these two sub-strains result from recombination events mediated by insertion sequences. However, the rate of nucleotide changes between them is estimated to be relatively low [3, 4]. Hence the major question arises: is this high degree of similarity at the nucleotide level reflected in the metabolic phenotype? A further query focused on unravelling the changes resulting from the aforementioned insertion sequences, and subsequently on obtaining a clear view of the regulation of metabolic processes in these closely-related laboratory sub-strains. In order to reveal the patterns of protein and gene expression at different levels of growth in the same medium under the same conditions, we analysed these expression patterns

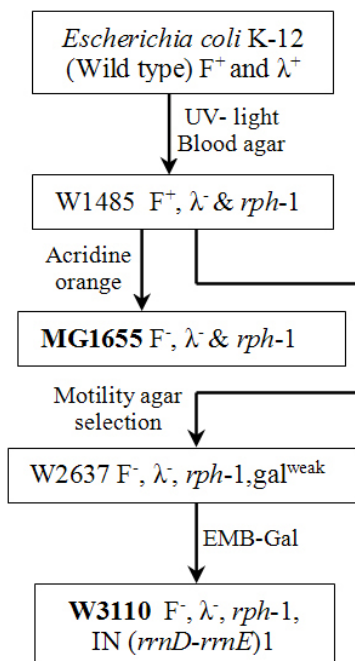
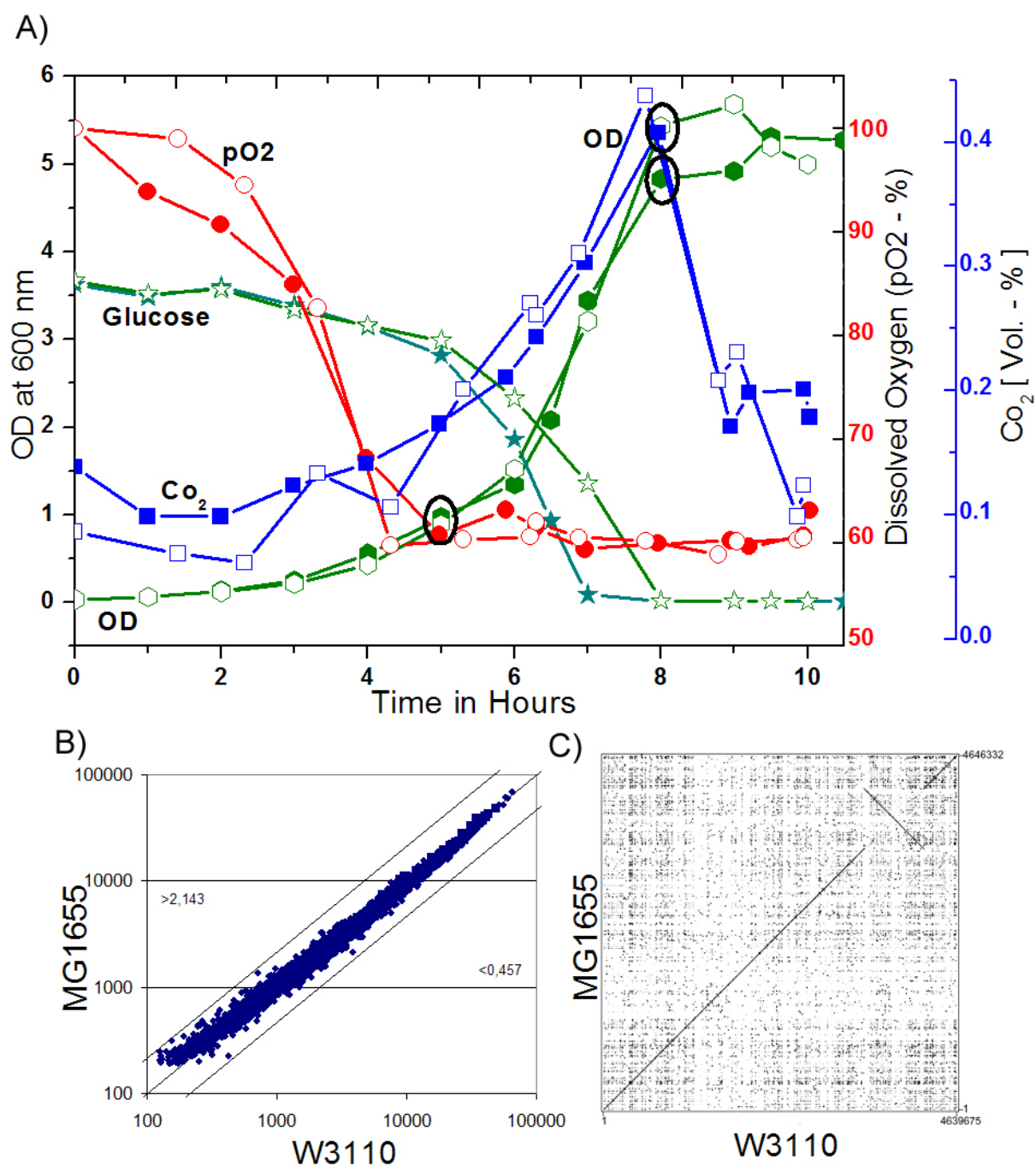


Fig. 2.1: **Origins of *E. coli* K-12 strains MG1655 and W3110.**  
The two strains are closely related and are derivatives of Lederberg strain W1485 [17]

during the exponential and early stationary phases of growth. A detailed study of the relevant proteomics and genomics showed that the patterns of global protein and gene expression differed considerably between the sub-strains despite their close relationship and absence of significant genomic differences, and despite the provision of strict and controlled growth conditions. Previous studies have demonstrated that the expression levels of many genes show abundant natural variation in species from yeast to humans [18, 19, 20, 21, 22, 23]. There is emerging evidence to suggest that mRNA expression patterns are necessary, but not by themselves sufficient, for describing the state of a biological system [24, 25]. Hence we included global proteomics analyses. In this study, the genes and proteins showing differential regulation belonged to the functional categories of RpoS (sigma subunit of RNA polymerase)-regulated genes, osmotic stress-related genes, enterobactin-related genes, and genes related to the energy metabolism of *E. coli*: the glycolysis pathway, the tricarboxylic acid cycle and the glyoxylate bypass.



**Fig. 2.2: Growth pattern and genome structure analyses.**

The graphs show: **(A)** growth pattern of strain W3110 (open symbols) and MG1655 (full symbols). Levels of dissolved oxygen ( $pO_2$ ), carbon dioxide in the emitted gas ( $CO_2$ ) and glucose concentration in the medium (glucose) are shown along with the optical densities (OD) of the bacterial cultures. The times of harvesting the samples are indicated by encircled symbols. **(B)** Scatter plot of a DNA microarray analysis shows a clear-cut gene-to-gene comparison of both genomes. **(C)** Dot plot comparing the whole genome sequence of W3110 against that of MG1655, showing the inversion of ribosomal genes in W3110.

---

# CHAPTER 3

## MATERIALS AND METHODS (I)

---

### *3.1 Strains and medium*

*Escherichia coli* K-12 strain W3110 was obtained from DSMZ (The German National Resource Centre for Biological Material); MG1655, obtained from the laboratory of A. Nishimura, had a culture background originating from the laboratory of M. Singer having no physiological defects, such as deletion around the *fnr* (fumarate-nitrate respiration) regulatory gene [26]. Both sub-strains were grown on M9 minimal medium with 4 gL<sup>-1</sup> glucose as carbon source in a 7 litre bioreactor (MBR Bioreactors). Cultivation was at pH 7, an operating temperature of 37 °C and a dissolved oxygen level of 60% saturation, automatically controlled by the stirrer frequency.

### *3.2 Two-dimensional SDS-PAGE gel electrophoresis*

Approximately 4.1x10<sup>10</sup> cells were harvested from the exponential (OD 600 ≈ 1) and early stationary phases (OD 600 ≈ 4) in all experiments. Isolated proteins were resuspended in 500 μl of rehydration buffer and 300 μg of protein sample was dissolved in 1.5 μl (IPG)-buffer, pH 4-7, and loaded on to a dry 24 cm Immobiline strip (Amersham Biosciences), pI range 4-7. The first dimension was developed on an IPG-phor (Amersham Biosciences) electrophoresis apparatus over 75,000 Vh. After equilibration of the strips, the second dimension was developed on a 12.5% polyacrylamide gel for 30 min at 3 W per gel, followed by a further run at 20 W per gel until the end. The second dimension was performed on an EttanDalt (Amersham Biosciences) electrophoresis unit. For comparative analysis, the gels were stained with Coomassie blue.

### *3.3 In-gel tryptic digestion and mass spectrometry*

Tryptic digestion was performed as reported previously [27]. Mass spectra were obtained on a Biflex III MALDI-TOF-MS (Bruker). The peptide mass fingerprints were annotated using the MASCOT search engine (Matrix Science). The parameters used were: Taxonomy, all entries; Enzyme, trypsin; Missed cleavages, 1; ppm, 100; Database, *E. coli*.

### ***3.4 Microarray experiments***

Genomic *E. coli* K-12 DNA microarrays were produced by spotting the PCR products robotically [28]. Total RNA was isolated using an RNeasy kit (Qiagen) according to the manufacturer's instructions with slight modifications. RNA concentration and quality were checked photometrically and on formaldehyde gels. The same amount of total RNA (25  $\mu\text{g}$ ) was used for each random hexamer-primed synthesis of fluorescence-labeled cDNA incorporating the fluorescent nucleotide analogues Cy3-dUTP or Cy5-dUTP.

### ***3.5 DNA microarray analysis***

Hybridization of fluorescence-labeled cDNA to the microarrays and the washing protocol were as described previously [29, 28]. The fluorescence was detected at 532 nm (Cy3) or 635 nm (Cy5) at a resolution of 10  $\mu\text{m}$  using a GenePix 4000 laser scanner (Axon). Microarray images were analyzed by GenePix Pro 3.0 software (Axon). The normalised Cy5/Cy3 ratio for the median was taken to reflect relative changes in RNA levels [29].

### ***3.6 Microarray data analysis***

For the exponential growth phase experiments, four replicates were used (two biological and two technical); for exponential and early stationary growth phase experiments on both sub-strains, we used two biological replicas. The normalised  $\log_2$  ratios were used to represent the data graphically on a frequency histogram using MapMan software [30] with functional classifications based on MultiFun and Gene Ontology terms, a cell function assignment scheme, with slight modification [31]. The normalised  $\log_2$  ratios were used for unsupervised analysis (two-dimensional hierarchical clustering) using Cluster 3.0 and TreeView [32], where average-linkage clustering was performed using an uncentered correlation metric. Principal components analysis (PCA) was performed using Matlab 6.5 (The MathWorks, Inc.).

### ***3.7 Real-time RT-PCR***

RNA was quantified with a LightCycler (Roche Diagnostics) using the QuantiTech SYBR Green RT-PCR detection system (Qiagen) according to the manufacturer's instructions. The specificity of amplification was determined by melting-curve analysis. For each sample, a straight line representing a log-linear increase in fluorescence was fitted automatically by selecting three points above the threshold level. The intersection of the extrap-



olated line with the threshold band was used to determine the fractional cycle number of the crossing point (Cp), calculated automatically by the LightCycler software. These crossing point values were used to calculate the fold differences between the samples. The 20 primers used for amplification are summarised in **Table 19.1** of Supplementary Material.

### 3.8 *Determination of siderophores*

Chrome Azurol S and Arnow's assays were performed as described previously [33, 34].

### 3.9 *Cloning and sequencing of rpoS*

The *rpoS* gene from both MG1655 and W3110 was amplified by PCR using the primers rp-1 (5'- ATGTTCCGTCAAGGGATCACG) and rp-2 (5'- GCGGATCCCTCGAGTTACTCG). Genomic DNAs were extracted and purified from the sub-strains using appropriate Wizard kits (Promega). Equal amount of genomic DNA were amplified using GoTaq DNA Polymerase (Promega) for 30 cycles; denaturation at 94 °C for 1 min, primer annealing at 65 °C for 1 min and extension at 72 °C for 2 min. A fragment of the predicted size was amplified and cloned into the vector pGEM-T (Promega) according to the manufacturer's directions. The plasmid was transformed into *E. coli* JM109 competent cells (Promega) and screened for white colonies on LB agar plates containing 50 µg/ml ampicillin, 30 µg/ml IPTG and 20 µg/ml X-gal. White colonies were picked and streaked on to LB/amp plates, then the plasmid DNA was purified using appropriate Wizard kits (Promega). The plasmid DNA was screened for the presence of *rpoS* by PCR and restriction digestion experiments before DNA sequencing. Sequencing was performed commercially by Agowa (Berlin, Germany). JM109 clones having the *rpoS* gene sequence from W3110 and MG1655 were utilized for studying the effect of *rpoS* amber mutation by examining the gene expression levels of various candidate genes.

# CHAPTER 4 RESULTS AND DISCUSSION (I)

## 4.1 Expression of *RpoS*-dependent genes

A set of 76 genes known to be regulated by the sigma factor RpoS [35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48] were considerably down-regulated in W3110 in comparison to MG1655 during exponential growth phase (**Figure 4.1A**). Sigma factor RpoS, a subunit

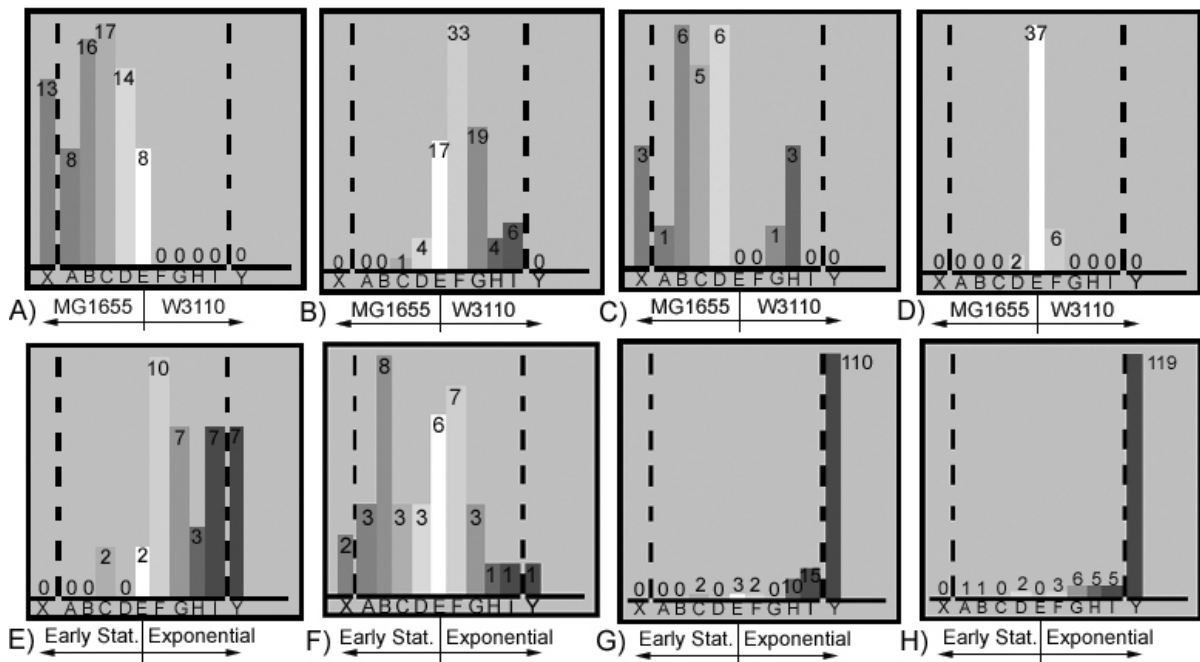


Fig. 4.1: Frequency histogram of gene expression data from W3110 and MG1655 based on  $\log_2$  ratios.

Uppercase letters denote the range of ratios: X=  $<-2.0$ ; A=  $-2.0$  to  $-1.55$ ; B=  $-1.55$  to  $-1.11$ ; C=  $-1.11$  to  $-0.66$ ; D=  $-0.66$  to  $-0.22$ ; E=  $-0.22$  to  $0.22$ ; F=  $0.22$  to  $0.66$ ; G=  $0.66$  to  $1.11$ ; H=  $1.11$  to  $1.55$ ; I=  $1.55$  to  $2.0$ ; Y=  $> 2.0$ . The number on a bar indicates the number of genes in that range. A-D shows the comparison between exponentially growing cells (W3110 versus MG1655); E-H shows the comparison of exponential and early stationary phases. (A) 76 RpoS-dependent genes (Supplementary Material; **Table 19.2**); (B) 85 central metabolism genes (Supplementary Material; **Table 19.3**); (C) 25 enterobactin and iron related genes (Supplementary Material; **Table 19.4**); (D) DNA replication genes, shown as control; (E, F) 38 RpoS-dependent genes from (E) W3110 and (F) MG1655 (Supplementary Material; **Table 19.5**); (G, H) building block biosynthesis genes from (G) W3110 and (H) MG1655, both shown as control.

of RNA polymerase, is known to be a master regulator of a complex regulatory network involved in the control of many genes during growth, especially during stress and the

early onset of stationary phase [49, 50, 51, 52]. Microarray analysis of exponential phase versus early stationary phase samples from these two sub-strains confirmed the difference in expression patterns for these RpoS dependent genes (**Figure 4.1 E,F**). Classification of these 76 RpoS-regulated genes according to functional category revealed considerable differences in expression patterns between the two sub-strains, implying differences in various functional categories of the cell. When comparing the microarray data of Lacour & Landini [42] denoting 41 genes which are considered to be regulated by RpoS, 50% of genes overlapped with the identified genes which are considered as differentially expressed from our microarray data. Likewise, when comparing the data reported by Weber *et al.* [47], out of 140 genes being reported to be controlled by RpoS under three different growth and stress conditions, 49% of the genes overlapped with our data which were considered as differentially expressed. Among these known RpoS-dependent genes, *osmY* (osmotically inducible protein) and *treF* (trehalase) showed higher expression levels in MG1655, and their protein product levels were also higher in a typical total cell proteomic analysis (**Figure 4.2 A,B**). All the major genes involved in osmotic stress - *otsA*, *otsB*, *bolA*, *treA*, *osmB*, *dps*, *osmC* and *osmY* - were up-regulated in MG1655 during the exponential growth phase (**Figure 4.1 A**). RpoS acts as a global regulator for the osmotic control of gene expression throughout growth [53]. This is apparently the case in exponentially growing cells as well as stationary phase cells [54, 55]. It is known that during osmotic upshift, the cells accumulate potassium ions and synthesize glutamate as a counter-ion in order to restore turgor pressure. Eventually, potassium ions and glutamate are replaced by suitable osmoprotectants such as glycine, betaine, proline and trehalose [54]. Among the major genes in the osmotic stress functional category that are up-regulated in MG1655, *otsA* and *otsB*, which code for trehalose-6-phosphate synthase and trehalose phosphatase respectively, are involved in trehalose biosynthesis. Similarly, the *treA* expression level was high in MG1655; the enzyme it encodes can utilize trehalose at high osmolarity by splitting it into glucose molecules. Intracellular metabolite analysis revealed a 1.14 (exponential phase) and 4.52 (early stationary phase) fold increased level of glutamate in MG1655 cells when compared to W3110. Subsequently, the expression levels of *gadA* and *gadB*, coding for glutamate decarboxylase alpha and beta respectively, were higher in MG1655; this enzyme is involved in converting glutamate to gamma-aminobutyrate. However, detailed comparison of the genomes of these two sub-strains in the neighborhood of the RpoS-dependent and osmotic stress genes revealed sequence differences in the area of *treA*, *otsA* and *otsB*. MG1655 contains an IS1 (insertion sequence 1) element downstream of *otsA* and *otsB*, but W3110 contains an IS5 element. Likewise, there is an adjacent IS5 element just upstream of *treA* in W3110 but not in MG1655. Previous studies demonstrated that insertion elements upstream or downstream of a gene or an operon can activate or diminish

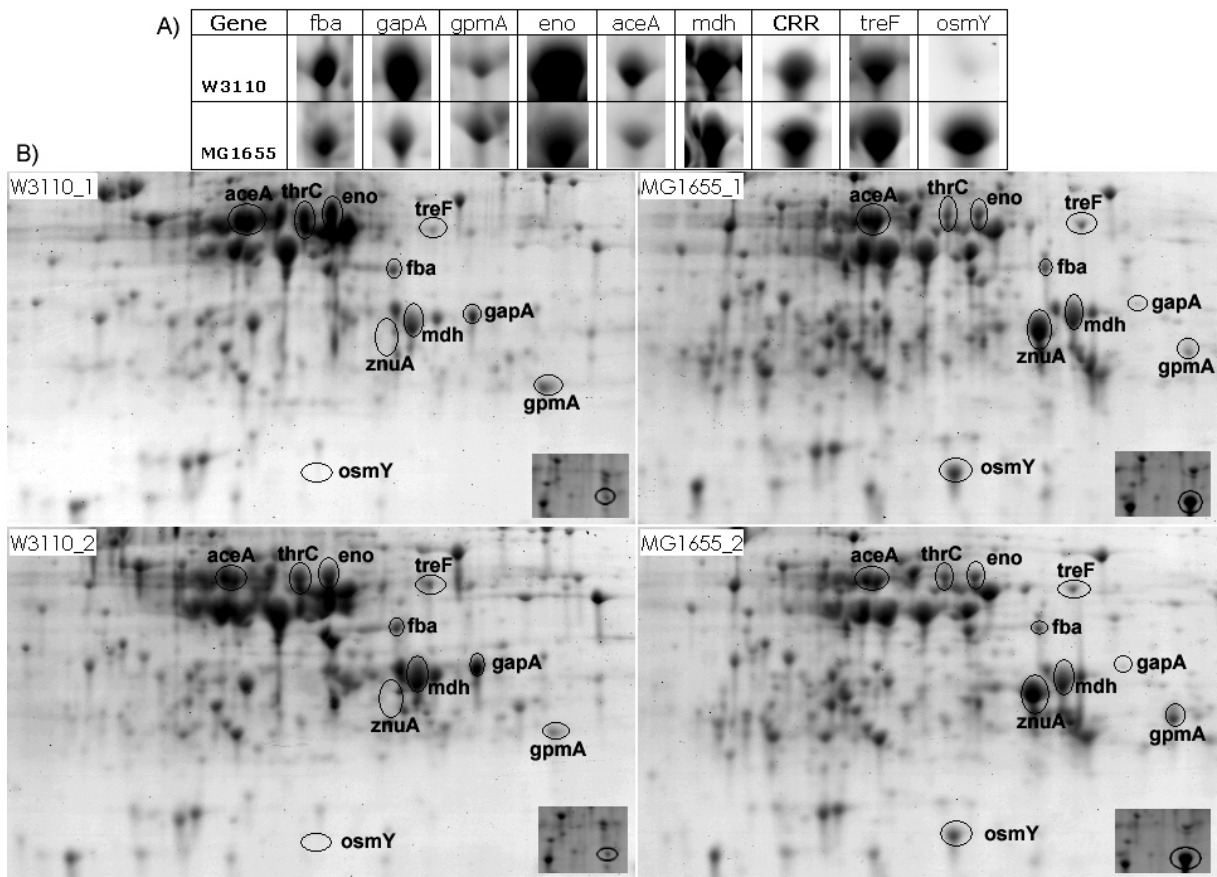


Fig. 4.2: Comparative proteome analyses of the *E. coli* sub-strains W3110 and MG1655.

Variable protein expression patterns were found for central metabolism genes (*fba*, *gapA*, *gpm* and *eno* involved in glycolysis; *aceA* and *mdh* involved in the TCA cycle) and genes involved in the osmotic stress response (*treF* and *osmY*). (A) Differential expression pattern during the exponential growth phase. The invariant expression pattern of CRR, responsible for the transport and availability of glucose, is shown as control. (B) Expression patterns for both the sub-strains during exponential growth phase (1) and early stationary phase (2). The insets show the dissimilarity in the expression patterns of OsmY.

the expression of that gene or operon [56, 57]. Being one of the primary global regulators of a complex regulatory network in the cell, RpoS plays a major role in regulating the genes involved in stress responses [52]. It associates with core RNA polymerase to determine recognition of specific promoter sites, thereby regulating this complex network. It is believed that under stress conditions, RpoS is induced by ppGpp and some unidentified factor(s) that successively stimulate its interaction with the core polymerase [58]. In addition to this regulation of RpoS synthesis, rapid degradation of RpoS is also controlled by the ClpXP protease [58, 59]. Application of unsupervised two-dimensional hierarchical clustering to the microarray data revealed that the RpoS-regulated genes showed clear differential regulation between the sub-strains and at different growth stages (**Figure**

**4.3**). Average linkage clustering was performed for this purpose, using an uncentered correlation metric. Hierarchical clustering of RpoS-dependent and DNA-replication genes showed that those in the former category were clearly over-expressed in MG1655, while those in the latter category showed no significant differences in expression (shown as control) between the sub-strains during exponential phase. When the exponential and early stationary phases were compared, MG1655 showed the commonly-expected response of RpoS-dependent genes. Thus, these genes were more highly expressed during early stationary phase than during exponential growth. The converse difference was observed in W3110 (**Figure 4.4A**). The RpoS-regulated genes were more highly expressed during exponential growth than during the early stationary phase. For multivariate data analysis, which reduces the dimensionality and complexity of the dataset without vitiating the accurate calculation of distance metrics representing the variability between samples, we used principal components analysis (PCA). Detailed PCA analysis of the genes known to be regulated by RpoS shows clear differences in expression levels between the sub-strains (**Figure 4.5**). The expression levels of this small set of RpoS-regulated genes sufficed to differentiate the sub-strains. When we investigated the reason for this differential expression, we found no regulatory elements such as Crp, Cya, and Hfq that differed in their expression levels. We therefore cloned *rpoS* from both sub-strains and determined its DNA sequence. This revealed a single base change in the *rpoS* sequence from W3110, a C-T transition at position 97, which is a commonly-reported point mutation in K12 isolates [60, 61]. This base change causes an amber mutation at codon 33 (CAG - TAG). Hence, the synthesis of RpoS is prematurely terminated. Several strains carrying this amber mutation show reduced RpoS activity when analyzed by traditional catalase assays [61]. Therefore, the down-regulation of the set of 76 known RpoS-regulated genes in W3110 relative to MG1655 during exponential growth phase could be attributed to this amber mutation at codon 33 in *rpoS* in W3110. However, the differences in expression (more or less reciprocal) of these genes between exponential and early stationary growth phase in the two sub-strains remains much more complicated to explain (**Figure 4.4A, and 4.3**).

## 4.2 *Expression of energy metabolism genes*

During exponential growth phase, most of the 85 genes in this functional category were up-regulated in W3110 (**Figure 4.1B**). In this 85-gene set, one group belongs to the *nuoA-N* operon and another to the hydrogenases. *nuoA-L* belonging to the same operon codes for NADH dehydrogenase-I, which shuttles electrons from NADH via FMN and iron-sulfur centers to quinones in the respiratory chain. Similarly, some genes for the tri-carboxylic acid (TCA) cycle such as *acnB* (aconitase B), *sucD* (succinyl-CoA), *sdhA* and

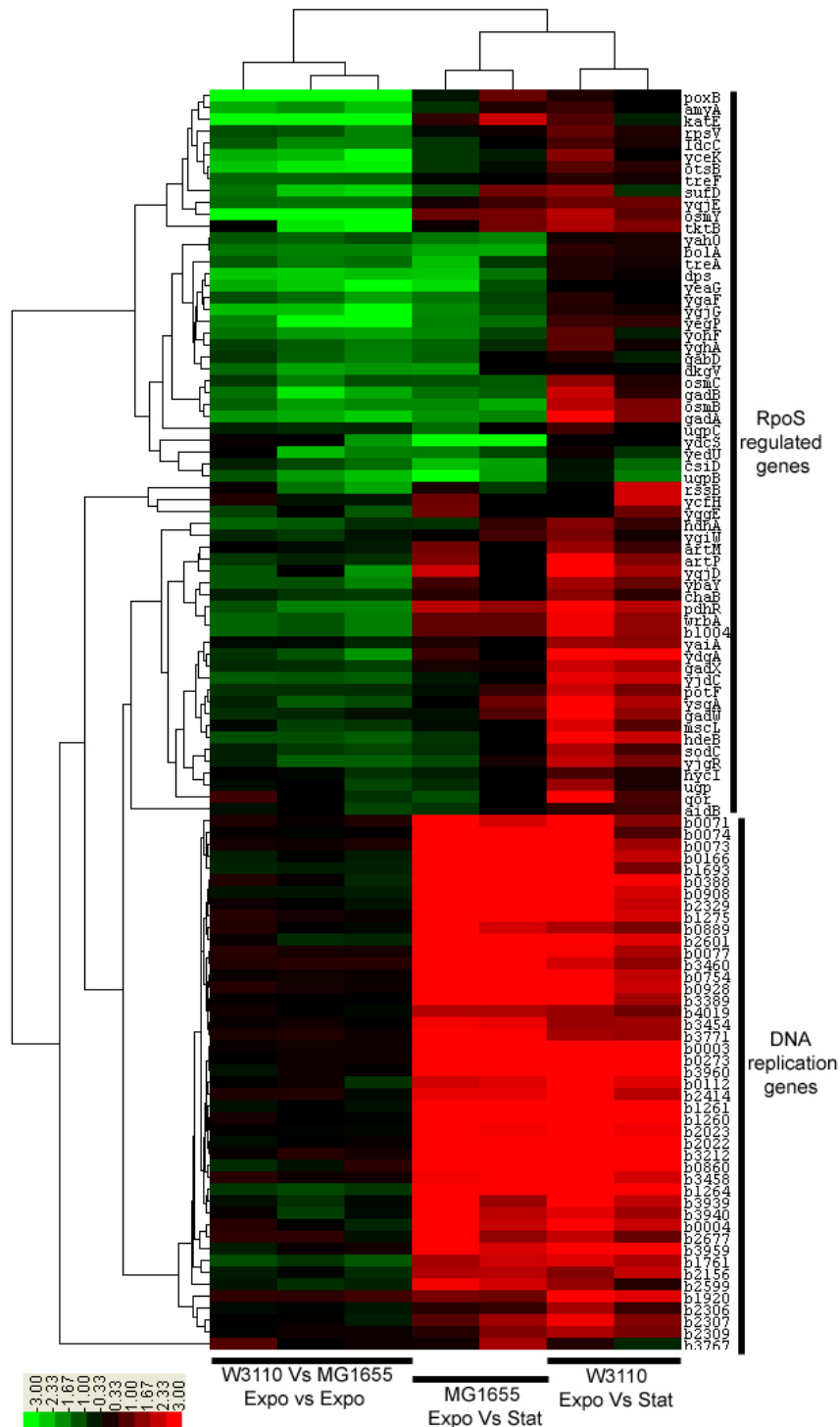


Fig. 4.3: **Two-dimensional hierarchical clustering of gene expression datasets from W3110 and MG1655**

The array data from each strain or growth condition are arranged in columns and rows corresponding to single genes. Dendrograms indicate the degree of similarity between clusters of genes, strain or growth conditions. Red indicates up-regulation and green indicates down-regulation in gene expression, according to the  $\log_2$  ratio scale on the lower left. Clustering shows prominent differential expression of RpoS regulated genes in both sub-strains as well as growth conditions; DNA replication gene expression levels are shown as control. Expo: Exponential phase; Stat: Early stationary phase.

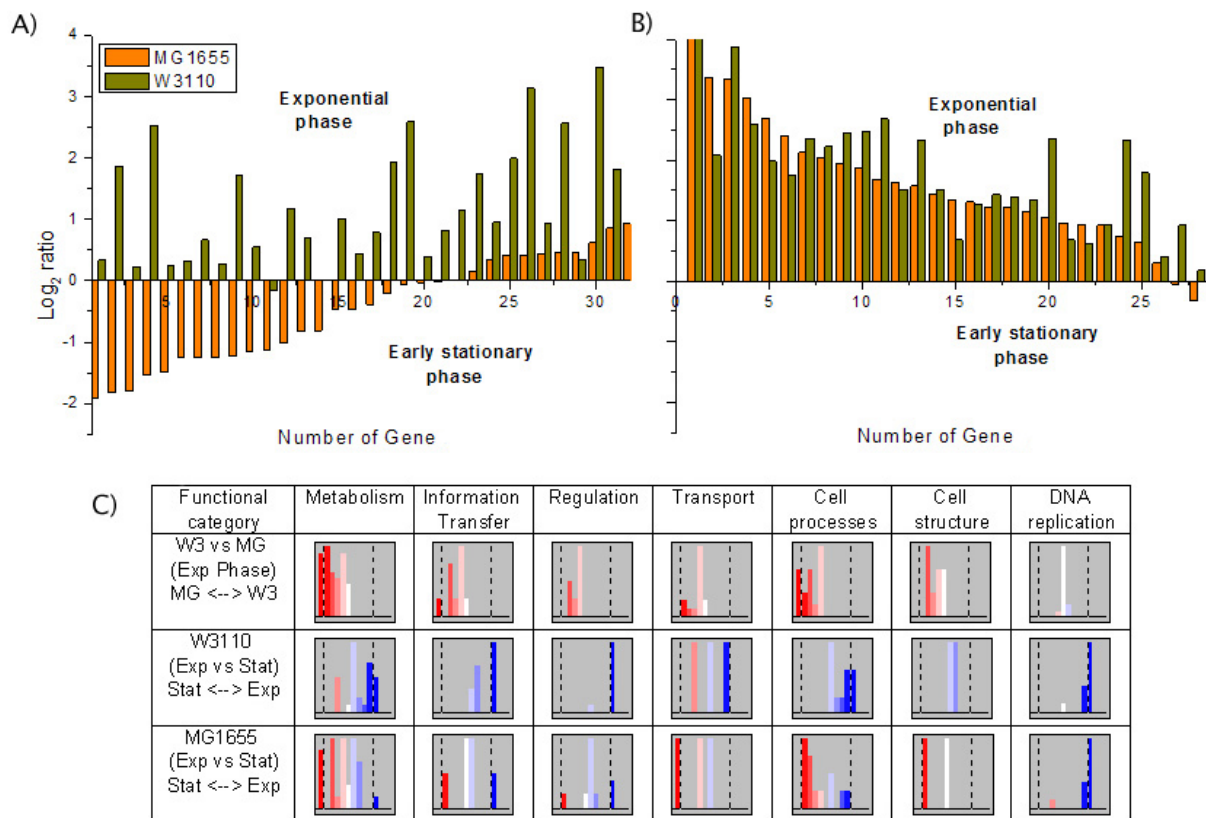


Fig. 4.4: **Development of gene expression during cultivation of *E. coli* sub-strains W3110 and MG1655**

(A, B) Expression levels of genes that are (A) RpoS dependent and (B) involved in cell division (shown as control) from exponential and early stationary phases of growth. (C) Expression levels of Rpos-dependent genes in relation to functional category. Shades of red denote higher expression levels in MG1655 or early stationary phase; shades of blue denote higher expression levels in W3110 or exponential phase. The first, second and third rows depict exponential phase and exponential versus early stationary phase of W3110 and likewise MG1655. The last column shows DNA replication genes (presented as control) obtained from the same dataset.

*sdhB* (succinate dehydrogenases) as well as *mdh* and *mgo* (malate dehydrogenases) were also up-regulated in W3110. This could have been expected, since previous studies have demonstrated that RpoS negatively regulates TCA cycle genes as well as those involved in aerobic and anaerobic respiration, fermentation and electron transport [43], consistent with the previous findings. A detailed comparison of the two sub-strains showed differences in the genome sequences in the neighbourhood of the *nuoA-N* operon, attributable to an insertion element. This was confirmed as IS1 by re-sequencing this region from both sub-strains. *lrhA*, which lies immediately upstream of the *nuoA-N* operon in both sub-strains, is located between *nuoA* and IS1 in W3110, but between *nuoA* and *yfbQ* in MG1655, where IS1 is absent (**Figure 4.6**). Interestingly, most of the genes downstream of *lrhA*, i.e. the *nuoA-N* operon coding for NADH dehydrogenase-I, were highly

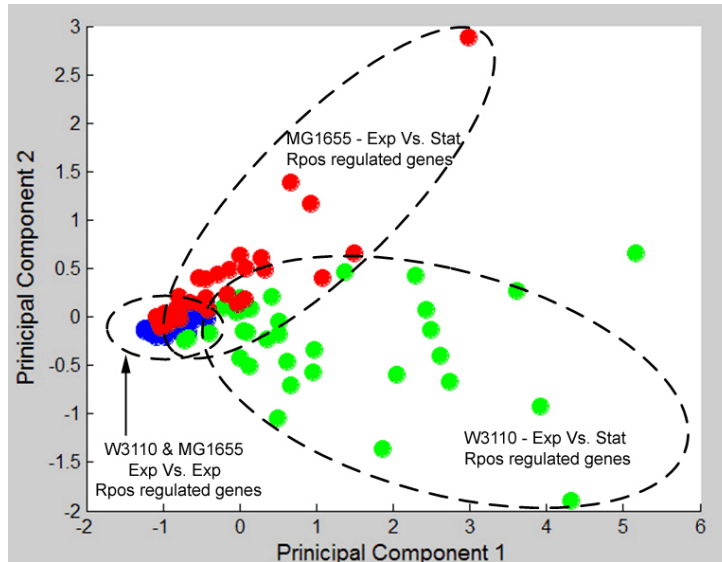


Fig. 4.5: Principal components analysis (PCA) of RpoS-regulated genes in W3110 and MG1655.

The figure shows a PCA of the distribution of expression ratios of RpoS-regulated genes in both sub-strains and growth conditions. The first two principal components are shown. Blue elements are the gene expression ratios of the two sub-strains compared during exponential growth. Red and green elements denote the gene expression ratios for MG1655 and W3110 when the exponential and early stationary growth phases are compared.

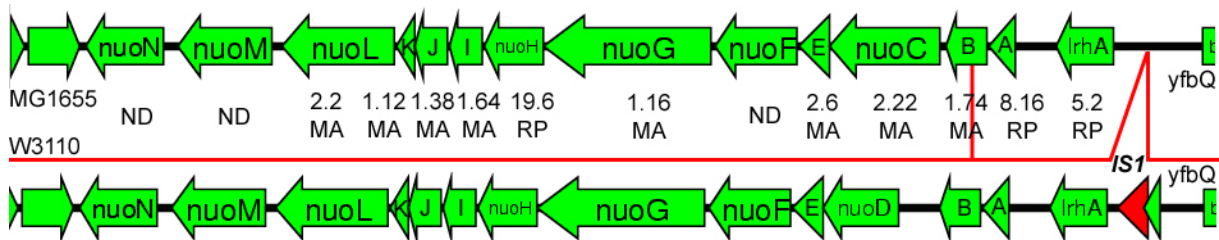


Fig. 4.6: Comparison of the genome sequence fragment of W3110 and MG1655. Genome alignment showing the IS1 element upstream of *lrhA* and the *nuoA-N* operon in W3110 but not MG1655; this was confirmed by re-sequencing the region. Numbers represent the fold change for each gene in W3110, and the sources of data are represented (RT: Real-time RT-PCR; MA: Microarray; ND: Not determined).

expressed in W3110 (**Figure 4.6**). The expression levels of some candidate genes in this operon were also confirmed by real-time RT-PCR. Therefore, it may be assumed that the insertion element upstream of *lrhA* in W3110 enhances the expression of both *lrhA* and the *nuoA-N* operon. Modulation of gene expression due to adjacent mobile elements has previously been reported for the *bgl* operon [57] and the cryptic gene *yicP* [56]. The *lrhA* gene is involved in the RpoS regulation at the translation level [62]. Simultaneously examining the mRNA and protein levels of *rpoS* and *lrhA* genes in W3110 could reveal more information on the regulatory role of *lrhA* gene.



### 4.3 *Expression of iron-related genes*

Precisely 21 of the 25 genes in the functional category of enterobactin and iron-related metabolism were significantly up-regulated in MG1655 compared to W3110 during exponential growth (**Figure 4.1C**). Among the products of these 21 genes, EntA-F proteins are involved in the biosynthesis of the catecholate siderophore enterobactin (Ent), responsible for the transport of iron from the bacterial environment into the cytoplasm. The products of *fepA-D*, *fepG* and *exbB* are involved in the transport of ferric enterobactin [63, 64, 65]; they also belong to this group, as does *fes* (ferric enterobactin esterase), the product of which renders iron available for metabolic use [66]. In this category, only 4 genes showed higher expression levels in W3110 than in MG1655: the ferric ion transport proteins A and B (FeoA, FeoB), and ferritin (FtnA, FtnB), an iron storage protein. Ferric ions are taken up into the cells for storage by means of ferritin-like molecules [67]. Siderophore (iron chelating compound) assays showed that the amounts of siderophore produced and released into the culture supernatants were higher in W3110 than in MG1655. Once ferric enterobactin is internalised, reductases catalyze the reduction of complexed ferric ( $\text{Fe}^{3+}$ ) to complexed ferrous ( $\text{Fe}^{2+}$ ) ions, providing iron for metabolic use. The ferric reductase activities in W3110 and MG1655 remained more or less equal throughout the growth period. Iron is required by virtually all organisms for essential biological processes; its predominant form is ferric ( $\text{Fe}^{3+}$ ). This ion is extremely insoluble; hence, *E. coli* secrete siderophores to ensure their iron supply. The catecholate-type siderophore enterobactin is produced by all strains of *E. coli* [68]. Genes involved in enterobactin uptake were up-regulated in MG1655, whilst operons involved in ferrous ion uptake were up-regulated in W3110. The main regulatory genes for iron metabolism such as *fnr* and *fur* and the ferric reductase gene *fre* showed no differential expression. The levels of siderophores in the culture supernatant of W3110 were shown by Arnow's assay (**Figure 4.7A**) and the Chrome Azurol S assay (**Figure 4.7B**) to be elevated. Interestingly, a search for commonality among the genes with more than two-fold higher expression levels in W3110 revealed that 21% of them code for proteins using iron as prosthetic group or cofactor, proteins with iron-sulfur clusters and iron storage proteins (**Figure 4.7C**). These observations indicate that the W3110 sub-strain should have easier access to reduced iron than MG1655.

### 4.4 *Verification of microarray results*

The expression levels of a significant set of candidates in our microarray data were verified quantitatively by real-time PCR (**Figure 4.8A**). All these expression levels differed considerably between W3110 and MG1655, and the relative differences were in complete

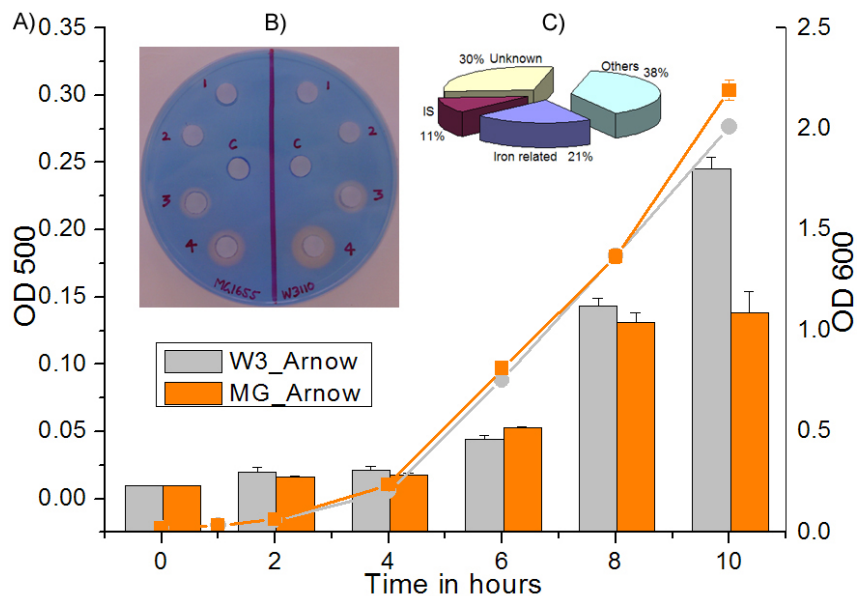


Fig. 4.7: **Additional confirming analyses.**

Siderophores (enterobactin) were analysed by (A) Arnow's assay (measured at OD 500), which showed higher siderophore concentrations in W3110 during late cultivation times (10 h) measured at OD 600. Error bars represents the standard deviation of three replicates. (B) Chrome Azurol S (CAS) assay. The haloes around the wells in the CAS plate show siderophore production in sample supernatant obtained during different periods of growth. Numbers 1-4 represents 4, 6, 8 and 10 hours, C denotes control (iron-supplemented culture supernatant). (C) Of all the genes up-regulated more than twofold in strain W3110, 21% were iron-related and 11% were related to insertion elements because of the high copy number in W3110.

agreement with our microarray data. *nuoA* and *nuoH*, which belong to a cluster of 13 genes encoding NADH dehydrogenase I, were expressed 8.2 and 19.7 higher in W3110, respectively; the expression levels of all genes in this cluster were high in W3110 according to our microarray data. *gapA* and *talA* genes involved in the central metabolism of the cell were expressed 1.4 and 2.1 fold higher in W3110, in line with this the GapA protein was also over expressed in W3110 (Figure 4.2A, B). Similarly, *lrhA* is expressed 5.2 times higher in W3110, where this gene lies between *nuoA* and IS1; in MG1655 it is upstream of *nuoA*, which lacks IS1. *entD*, *fepB* and *fepE* exhibited higher relative changes in MG1655: *entD* (enterobactin synthetase component D), which is involved in the biosynthesis of enterobactin, underwent a 427.6 fold change; *fepB* and *fepE* (ferric enterobactin transport proteins), which are involved in ferric enterobactin transport, underwent 20 and 6.9 fold changes, respectively. Similarly, all 21 genes in the functional category of enterobactin and iron related metabolism were substantially up-regulated in MG1655 according to our microarray data. The change of the ferrous iron uptake operon containing *feoA* and *feoB* were relatively high, 14.1 and 27.1 times those in W3110, in agreement with our microarray data. The relative changes in *rpoS* and *dnaA* expression

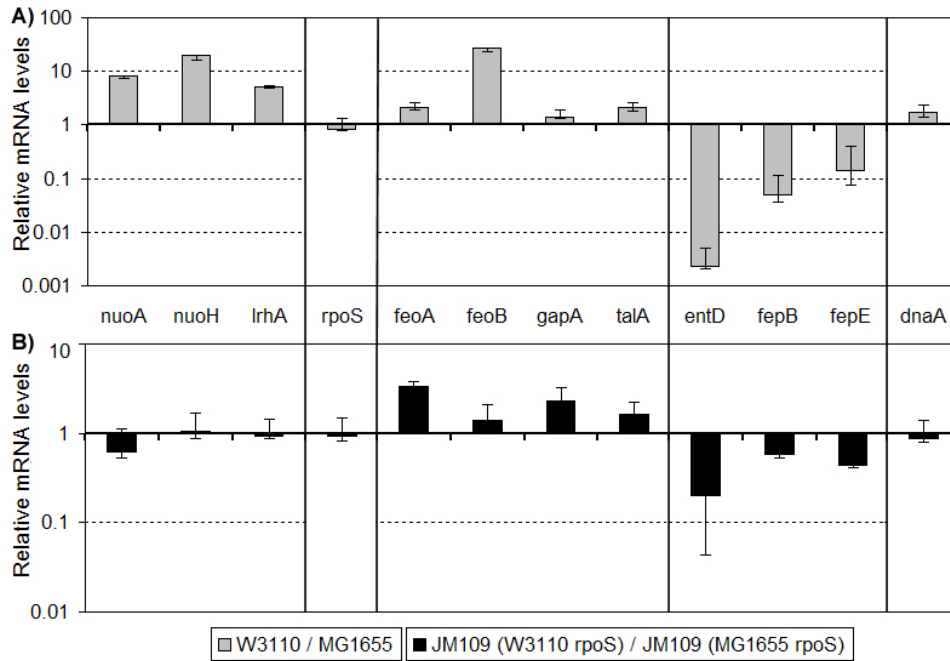


Fig. 4.8: **Real-time RT-PCR analyses.**

(A) The graphs verify the microarray results by quantitative real-time RT-PCR analyses. The bars represent the relative change of mRNA levels extracted from exponentially growing cells of W3110 and MG1655. (B) The relative changes of mRNA levels in JM109 strain harbouring plasmid containing *rpoS* gene from W3110 (JM109 -W3110 *rpoS*) and MG1655 (JM109 -MG1655 *rpoS*). Error bars represent standard deviation.

were insignificant, as expected.

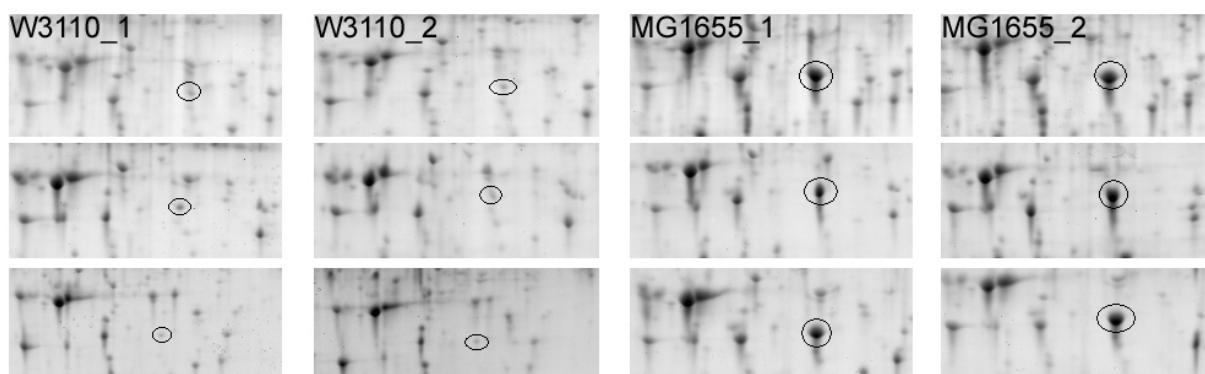
#### 4.5 *Effect of rpoS gene amber mutation in gene expression*

For the entire real-time RT-PCR validated genes from the microarray data, we examined the expression of these genes in the presence (W3110 *rpoS*) or absence (MG1655 *rpoS*) of the *rpoS* gene sequence harbouring amber mutation at codon 33 (Figure 4.8B). We determined this experiment by transforming the full length *rpoS* gene from W3110 and MG1655 into a strain having different genotype (JM109) and thus the strain dependent gene expression variation has been avoided. When analysed, most of the gene expression levels were in the same direction demonstrating the significant role of amber mutation dependent *rpoS* regulation among these genes. In the presence or absence of amber mutation in *rpoS* gene transformed in to a different strain, *nuoA*, *nuoH* and *lrhA* genes were not differentially regulated, unlike in W3110 and MG1655 comparison, indicating strain specific gene expression as assumed that the insertion element upstream of *lrhA* in W3110 enhances the expression of both *lrhA* and the *nuoA-N* operon (Figure 4.6).

The expression of *feoA*, *feoB*, *gapA*, and *talA* genes were higher in the strain bearing the plasmid harbouring the *rpoS* gene containing the amber mutation (from W3110). In line with this behaviour, these genes were also over-expressed in W3110 when compared with MG1655 demonstrating potential effect of the amber mutation in the *rpoS* gene. A similar direction of gene expression behaviour was also observed among *entD*, *fepB* and *fepE* genes in both the comparisons. Overall, these results suggest that the observed differences in the microarray data are largely due to the *rpoS* amber mutation in W3110 represented in this study.

#### 4.6 Expression of cellular proteins

The expression level of enzymes involved in glycolysis and the TCA cycle was higher in W3110 during exponential growth (**Figure 4.2A**). To achieve a reliable overview of the regulation of these enzymes in relation to the growth period, we analysed the protein fractions from early stationary phase, which demonstrated that these enzymes were up-regulated in W3110 (**Figure 4.2B**). Among them, Fba (fructose biphosphate aldolase), GapA (glyceraldehyde phosphate dehydrogenase), GpmA (phosphoglycerate mutase) and Eno (enolase) are involved in glycolysis; AceA (isocitrate lyase) and Mdh (malate dehydrogenase) are involved in the TCA cycle. In contrast, the TreF and OsmY proteins, which are involved in osmotic stress, were induced in higher amounts in MG1655 (**Figure 4.2A, B**). These findings from the proteome studies were in excellent agreement with the microarray results. In particular, OsmY protein was consistently expressed at higher levels in MG1655 during both the exponential and the early stationary phases of growth (**Figure 4.2B inset, 4.9**).



**Fig. 4.9: Comparative OsmY protein analyses of the *E. coli* sub-strains W3110 and MG1655.**

Dissimilar expression patterns of OsmY for both the sub-strains during exponential growth phase (1) and early stationary phase (2).

### 5.1 *Plasticity of gene and protein expression*

The growth behaviour (**Figure 2.2A**) and genome-based data (**Figure 2.2B, C**) point to the conclusion that both *Escherichia coli* sub-strains examined, W3110 and MG1655, may have similar metabolic phenotypes. However, the global transcriptome and proteome data lead to the conclusion that the global proteomes and the genome expression patterns exhibit high plasticity in these closely related sub-strains of *E. coli*- K-12, notwithstanding the strict and controlled growth conditions used. The genomes being highly similar at the nucleotide level, was not reflected in cellular metabolism, showing that complex regulatory patterns can differ even between closely-related sub-strains. In MG1655, genes involved in osmotic stress and enterobactin were effectively up-regulated (**Figure 4.1**); in W3110, genes involved in central metabolism were up-regulated and

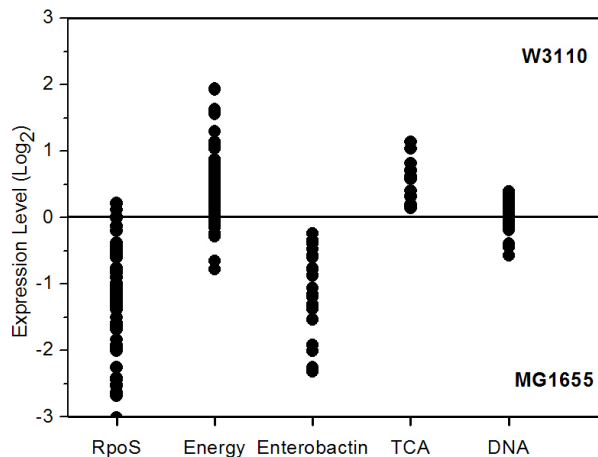


Fig. 5.1: **Global comparison of gene expression in *E. coli* sub-strains W3110 and MG1655.**

The graph shows a global comparison with respect to different functional gene categories, namely, RpoS-dependent, Energy metabolism, Enterobactin related, TCA cycle and DNA replication (shown as control).

RpoS-dependent genes were significantly down-regulated (**Figure 4.1**), with implications for various functional categories (**Figure 4.4C**). In contrast to experiments that measure correlations among transcript and protein levels under different conditions and in strains bearing knockout mutations, our approach led to the general conclusion that gene expression and protein patterns differ between closely-related sub-strains, a phenomenon that

can be largely ignored in the former types of experiments. Our results also emphasize that complex regulatory pathways are altered by such modifications at the genome level. Functional differences in central metabolism, which generates precursor metabolites and energy, were also apparent between these two closely-related and commonly-used sub-strains of *E. coli* K-12. These metabolic differences may indicate differences in regulatory networks. If such differences are not taken into account in critical experiments, the results obtained may easily be interpreted misleadingly.

## Part II

Perceiving molecular evolution processes in *Escherichia coli* by comprehensive metabolite and gene expression profiling.

## Abstract

The mechanism of evolution in different conditions can be examined from various molecular aspects that constitute a cell, namely, transcript, protein or metabolite abundance. We have analyzed transcript and metabolite abundance changes in evolved and ancestor strains in three different evolutionary conditions, namely, excess-nutrient adaptation, prolonged stationary phase adaptation and adaptation due to environmental shift, in two different strains of *Escherichia coli* K-12 (MG1655 and DH10B). Metabolite profiling of 84 identified metabolites revealed that most of the metabolites involved in TCA cycle and nucleotide metabolism were found to be altered in both the excess-nutrient evolved lines. Gene expression profiling using whole genome microarray with 4288 ORFs revealed the overrepresentation of transport functional category in all evolved lines. Excess-nutrient adapted lines were found to exhibit higher degree of positive correlation indicating parallelism between ancestor and evolved lines when compared to prolonged stationary phase adapted lines. Gene-to-metabolite correlation network analysis revealed the overrepresentation of membrane associated functional categories. Proteome analysis revealed the major role of outer-membrane proteins in adaptive evolution. GltB, LamB and YaeT proteins in excess-nutrient lines and FepA, CirA, OmpC and OmpA in prolonged stationary phase lines were found to be differentially over-expressed. These studies show that adaptive evolution in excess nutrient conditions are appropriate for examining the extent of parallelism in the evolutionary process whereas prolonged stationary phase conditions are useful to understand evolution of microbial diversity among evolved populations and the dynamic state of the evolved condition. These studies aid in providing a better understanding of the process of evolution and also provide a valuable initial source of data for systems biology of microbial evolution.

**Authors:** Chandran Vijayendran, Aiko Barsch, Karl Friehs, Karsten Niehaus, Anke Becker and Erwin Flaschel.

**Author's contribution:** CV conducted all the experiments cited in this study, analysed the results and wrote this report. AB1 was involved in metabolomics experiments. KF was involved in experimental guidance. KN was involved in experimental design. AB2 is the scientist in whose laboratory microarray experiments were conducted. EF is the scientist in whose laboratory all the experiments were conducted and was involved in experimental design.



---

## CHAPTER 6 INTRODUCTION (II)

---

Most microorganisms grow in environments that are not conducive to their growth. The level of nutrients available to them is rarely optimal. These microbes must adapt to environmental conditions which consist of either excess, sub-optimal (limiting), or fluctuating levels of nutrients or in famine. The study of evolution can be facilitated by observing the process and consequences of evolution in feasible laboratory conditions by culturing a micro-organism in varying nutrient environments [69, 70, 71, 72]. The extensively studied microbial evolution processes are nutrient-limited adaptive evolution [73, 74, 75] and famine-induced prolonged stationary phase evolution [76, 77, 78]. During prolonged carbon starvation, micro-organisms can undergo rapid evolution with mutants expressing a growth advantage in stationary phase (GASP) phenotype [70]. These mutants harbouring a selective advantage out-compete their siblings and takeover the culture through their progeny [79, 80, 81]. Adaptive evolution of micro-organisms is a process in which specific mutations result in phenotypic attributes responsible for fitness in a particular selective environment [69]. Laboratory studies under these evolutionary conditions can address fundamental questions regarding adaptation process and selection pressures thereby explaining modes of evolution.

In this study, we have used *Escherichia coli* K-12 strains (MG1655 and DH10B) from (1) serial passage system for excess nutrient adaptive evolution studies, (2) constant batch culture for prolonged stationary phase evolution studies and (3) culture with nutrient alteration following adaptation to a particular nutrient for examining pleiotropic effects due to environmental shift. During adverse conditions, micro-organisms are known to exploit the limiting resource more quickly and adapt themselves efficiently to assimilate various metabolites. Some of these residual metabolites comprise an alternative resource which the organism can metabolize on [70]. Continual assimilation of various metabolites and various compounds being metabolized by the organism offers a specific niche which allows the organism to evolve with a genetic capacity to utilize those assimilated metabolites [70]. Hence a detailed metabolite analysis of these evolved populations would guide towards a better understanding of these evolution processes. Along with the data generated from transcriptomics approaches, metabolomics data would be of vital importance for viewing a global picture of an organism in a particular time point, where metabolite behaviour closely reflects the actual cellular environment and the observed phenotype of that organism.

We have applied metabolome and gene expression profiling approaches to understand excess nutrient adaptive evolution, prolonged stationary phase evolution and pleiotropic effects due to environmental shift in two strains harbouring different genotypes. To eliminate the possibility of the strain-dependent phenomenon of evolution and to examine the parallelism of the laboratory evolution process, we examined the hitherto mentioned evolutionary processes in two different strains. As a result we had to compare the metabolite and gene expression profiles under various conditions for these strains, namely,

<b>Strain abbreviations</b>	<b>Evolved condition</b>
MG and DH	MG1655 and DH10B grown in glucose
MGGal and DHGal	MG1655 and DH10B grown in galactose
MGAdp and DHAdp	MG1655 and DH10B adapted about 1000 generations in glucose
MGAdpGal and DHAdpGal	MGAdp and DHAdp (glucose evolved strains) grown in galactose
MGStat and DHStat	MG1655 and DH10B grown in prolonged stationary phase (37 days)

Tab. 6.1: **Strains and their evolved conditions.**

In this study, we depicted a whole picture of laboratory molecular evolution processes among two different strains by integrating multidimensional metabolome and gene expression data to find vital metabolites and genes involved in evolution process.

---

# CHAPTER 7

## MATERIALS AND METHODS (II)

---

### 7.1 *Strains and culture conditions*

Both the bacterial strains, MG1655 and DH10B used in this study are derivatives of *E.coli* K-12. All the experiments were conducted in 250 mL of M9 minimal medium supplemented with 4 gL<sup>-1</sup> of glucose or galactose in covered 1 L Erlenmeyer flasks at 37 °C. Adaptation to excess nutrient experiments was carried out in the presence of 4 gL<sup>-1</sup> glucose through serial passage at exponential phase for about 1000 generations. The cells were grown overnight and were diluted by passage into fresh medium. Passage of each culture into fresh medium was conducted in a laminar flow station using standard sterile technique practices. Serial passage were conducted for 37 days at exponential phase for about 1000 generations. For adaptation due to environmental shift experiments, the strains which were adapted to excess nutrient (glucose) condition for about 1000 generations were grown in 4 gL<sup>-1</sup> galactose. For prolonged stationary phase adaptation experiments, both the strains were incubated for 37 days in M9 minimal medium with 4 gL<sup>-1</sup> glucose as initial source of carbon. The evolved populations were frozen using liquid nitrogen and stored in a freezer at -80 °C.

### 7.2 *Metabolite profiling*

Approximately equal number of cells ( $7 \times 10^9$ ) were taken from the exponential phase of growth for all the experiments. Extracted cells were disrupted using acid washed glass beads at maximum speed in a Ribolyser (*Q-BIOgene*) at a setting of 6.5 ms, two-times for 45 seconds in the presence of 80 % methanol. Subsequently, metabolites were derivatized using methoxylamine hydrochloride and N-methyl-N-[trimethylsilyl]trifluoroacetamide in the presence of ribitol as the internal standard. Sample volumes of 1  $\mu$ l were analysed with a TraceGC gas chromatograph coupled to a PolarisQ ion trap mass spectrometer (Thermo Electron). Derivatized metabolites were evaporated at 250 °C in splitless mode and separated on a 30 m x 0.25 mm Equity-5 column with 0.25  $\mu$ m coating (Supelco). Metabolites were identified by comparison to purified standards, the NIST 2005 database (NIST) and the Golm Metabolome Database [82]. Selected metabolite peak areas were automatically quantified using the processing setup implemented in the Xcalibur 1.4 software (Thermo Electron). The relative response ratios calculated from the peak areas were

normalized by the internal standard ribitol and dry weight of the sample. For both the strains in all the biological experiments, six replicates were used which consisted of three biological replicates (per strain).

### 7.3 *Gene expression profiling*

*E. coli* K12 V2 OciChip arrays containing 4,288 gene specific oligonucleotide probes representing the complete *E. coli* K12 genome were utilized in this study (Ocimum Biosolutions). Total RNA was isolated using RNeasy kit (Qiagen) according to the manufacturer's instructions. Reverse transcription, labelling, and scanning were performed as described previously [83]. Hybridization was carried out according to the manufacturer's instructions (Ocimum Biosolutions).

### 7.4 *Microarray data analysis*

Mean signal and mean local background intensities were determined for each spot of the microarray images by using the ImaGene 6.0 software for spot detection, image segmentation and signal quantification (BioDiscovery). After subtraction of the local background intensities from the signal intensities, the average intensity in both channels was subsequently normalized by the LOWESS method using the GeneSight 4.0 software package (BioDiscovery). The normalised  $\log_2$  ratios were used to represent the data graphically and to calculate Wilcoxon rank sum test  $P$ -values using MapMan software [30] with functional classifications based on MultiFun and Gene Ontology terms, a cell function assignment scheme, with slight modification [31, 84]. For both the strains in all the biological experiments, three or more replicates were used which consisted of three biological replicates. The variation among the biological replicates were estimated to be relatively low (**Figure 7.1**).

### 7.5 *Network analysis*

All the networks reported in this study were constructed based on Pearson correlation coefficient ( $PCC$ )  $r \geq 0.9$  measure i.e. nodes which correspond to genes or metabolites with  $r \geq 0.9$  were linked by edge. All-against-all metabolite and gene expression profile  $r$ -values of evolution-specific matrices were used to generate evolution-specific coexpression network. Strain and evolution specific matrices were used to generate evolution-specific intersection coexpression network. Intersection coexpression networks are the network

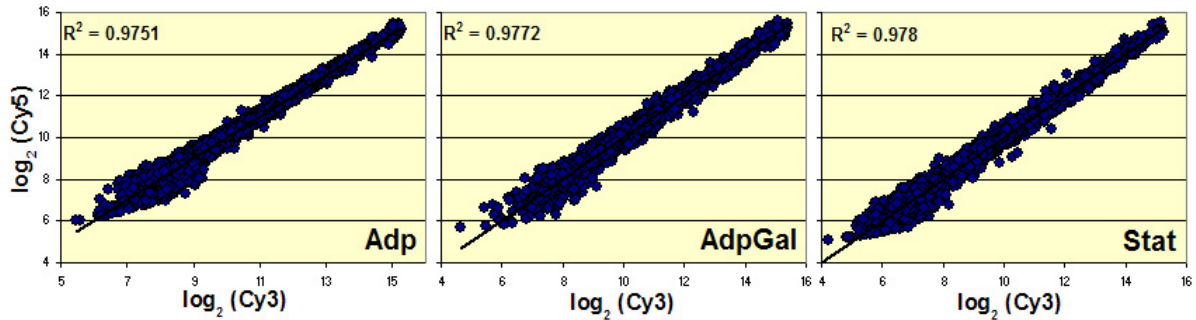


Fig. 7.1: Gene expression among the biological replicates.

The gene expression levels among the evolved strain replicates indicating relatively low level of variation among the replicates used for the evolutionary studies.

over the set of nodes  $N$  where there is a link ( $r \geq 0.9$ ) between two nodes  $i$  and  $j$  if they are connected in both the strains in the particular evolutionary condition in context. Topological properties of the networks were analyzed using the pajek program [85].

## 7.6 Network functional analysis

Network visualization and functional analysis was done using Cytoscape [86]. Networks were screened for highly linked clusters of genes or metabolites using MCODE [87]. Genes in the networks were functionally categorized using their Gene Ontology (GO) biological process annotation terms [88] and the overrepresented GO terms were identified with BINGO [89]. The Hypergeometric test was used to do this with the Benjamini and Hochberg false discovery rate correction ( $P$ -value  $\leq 0.05$ ).

## 7.7 Outer-membrane protein analysis

Approximately equal number of extracted cells ( $7 \times 10^9$ ) were disrupted by ultrasonication with 5 mL of 50 mM Tris/HCl, pH 7.3 containing 0.7 mg of DNase I (Sigma) and 0.5 mM protease inhibitor (pefabloc SC). After the unbroken cells were removed by centrifugation the supernatant was treated with ice cold 0.1 M sodium carbonate (pH 11). Eventually the carbonate treated membranes were collected and subsequently analysed by SDS 1D gel electrophoresis. Excised protein bands were subjected to tryptic digestion and mass spectra were obtained on a Biflex III MALDI-TOF-MS (Bruker). Peptide masses were searched against the *E. coli* database located on our local server using the MASCOT search engine (Matrix Science) with a mass cut-off of 100 p.p.m..

---

# CHAPTER 8

## RESULTS (II)

---

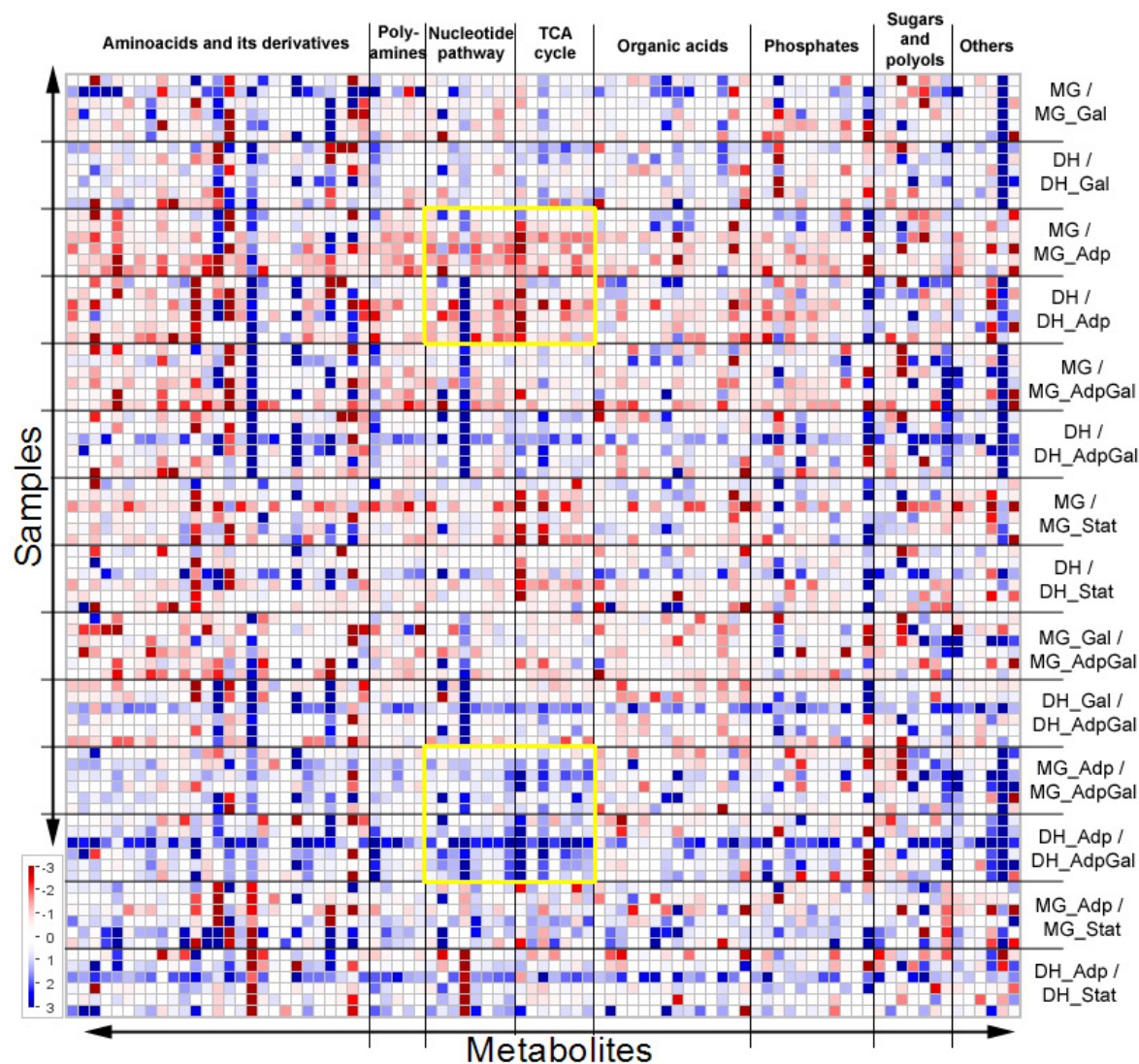
The Adp line cultures (MGAdp and DHAdp) were maintained in prolonged exponential growth phase by daily passage into fresh medium for about 1000 generations. The Stat line cultures (MGStat and DHStat) were maintained in constant batch culture for 37 days where no nutrients were added after initial inoculation and no cells were removed, unlike the former setup. For the AdpGal line cultures (MGAdpGal and DHAdpGal), Adp lines (glucose adapted) were grown in medium containing galactose as carbon source, thus creating an environmental shift for the cells with respect to the standard nutrient source. During this period of adaptation, both Adp lines (evolved) showed increased fitness in their growth, whereas Stat lines (evolved) showed growth behaviour similar to their ancestors (**Figure 13.1B**). The samples of MG, DH, MGGal, DHGal, MGAdp, DHAdp, MGAdpGal, DHAdpGal, MGStat and DHStat lines grown in respective carbon sources were harvested in the mid-exponential phase of growth for both metabolome and transcriptome analysis.

In metabolome analysis, from about 200 peaks in each chromatogram,  $\approx 100$  metabolites were identified by gas chromatography - mass spectrometry (GC-MS) and in transcriptome analysis, a whole genome microarray consisting of 4,288 open reading frames (ORFs) of *Escherichia coli* K-12 was used. To examine the multivariate measures of the variability of the metabolite and gene expression profiles for the obtained data, and for clustering the biological samples, we applied principal components analysis (PCA). In order to identify parallel metabolite accumulation and gene expression, we applied pairwise correlation plot analysis. To examine the extent of parallelism among the evolved lines, metabolite-to-gene correlation networks were constructed and their topological properties were studied. By mapping the correlation networks to Gene ontology (GO) functional annotations, functional relevance of the networks was obtained. Subsequently, functional modules which were statistically significantly overrepresented in respective evolution processes were elucidated.

### 8.1 *Metabolome profiling*

Till date, metabolome profiling has been frequently applied for obtaining quantitative information on metabolites for studies on mutational or environmental effects, but not in an evolutionary context [13, 14]. Here, for our evolutionary studies, we used an approach that

combined metabolomics and transcriptomics which offers whole genome coverage. In total, 84 metabolites of known chemical structure were quantified in every chromatogram. The full datasets (Supplementary Material; **Table 17.9**) from the metabolite profiling study are presented in an overlay heat map in **Figure 8.1**. This map shows the re-

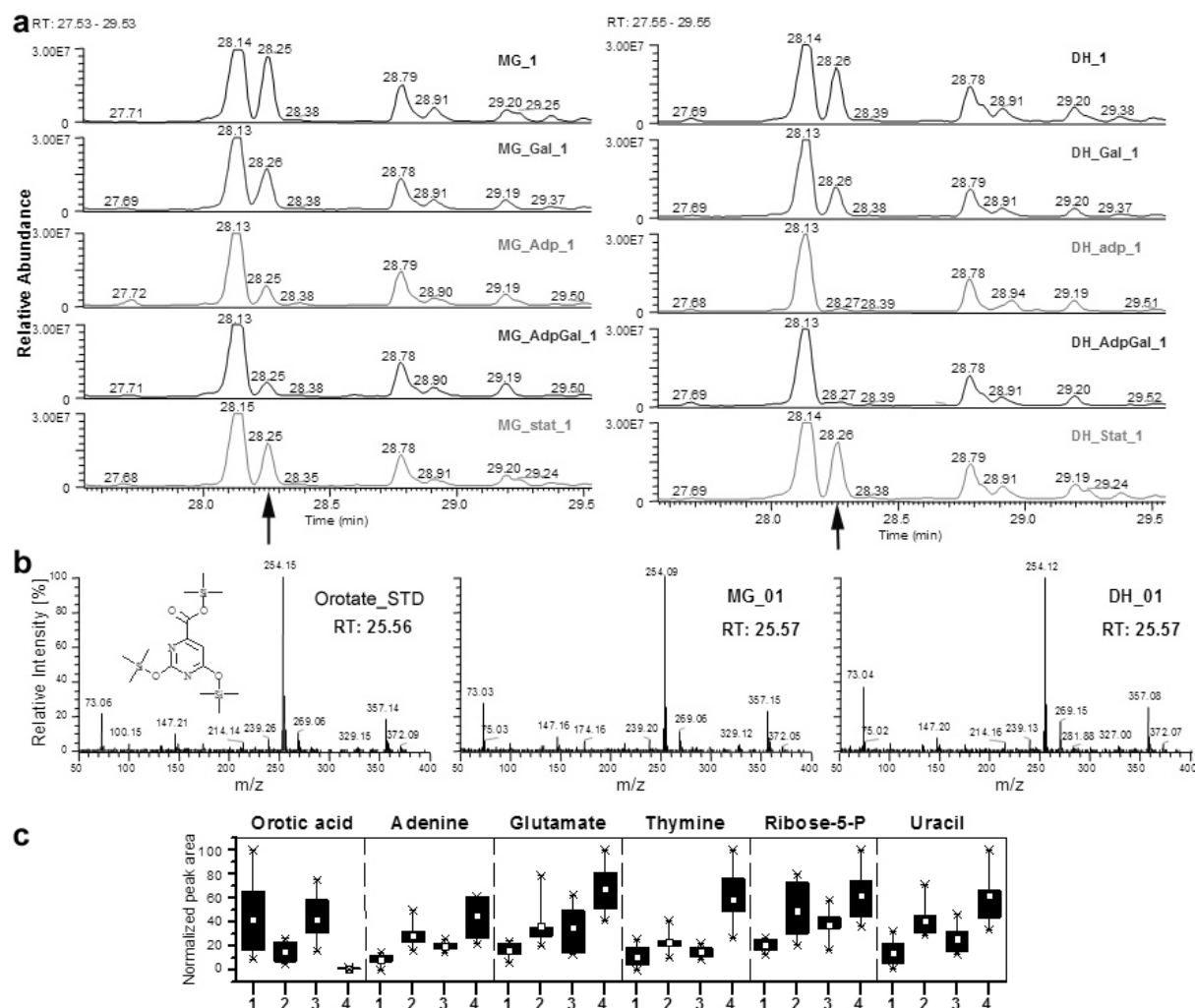


**Fig. 8.1: Overlay heat map of the metabolite profiles.**

Logarithmically transformed (to base 2) response ratios were used to plot the heat map. Red or blue color indicates that the metabolite content is decreased or increased, respectively. For each sample, gas chromatography/mass spectrometry was used to quantify 84 metabolites (non-redundant), categorized into amino acids and their derivatives, polyamines, metabolites involved in nucleotide related pathways, TCA cycle, organic acids, phosphates, and sugar and polyols. Highlighted yellow boxes indicate significant changes in the metabolite level in the TCA cycle and the nucleotide related pathways of the evolved lines.

sponse ratios of relevant comparisons of the samples analyzed. In most cases the levels of metabolites are significantly changed in evolved lines and their directional behavior is

more or less constant in both the ancestral strains and in their evolved strains as shown in **Figure 8.2**.



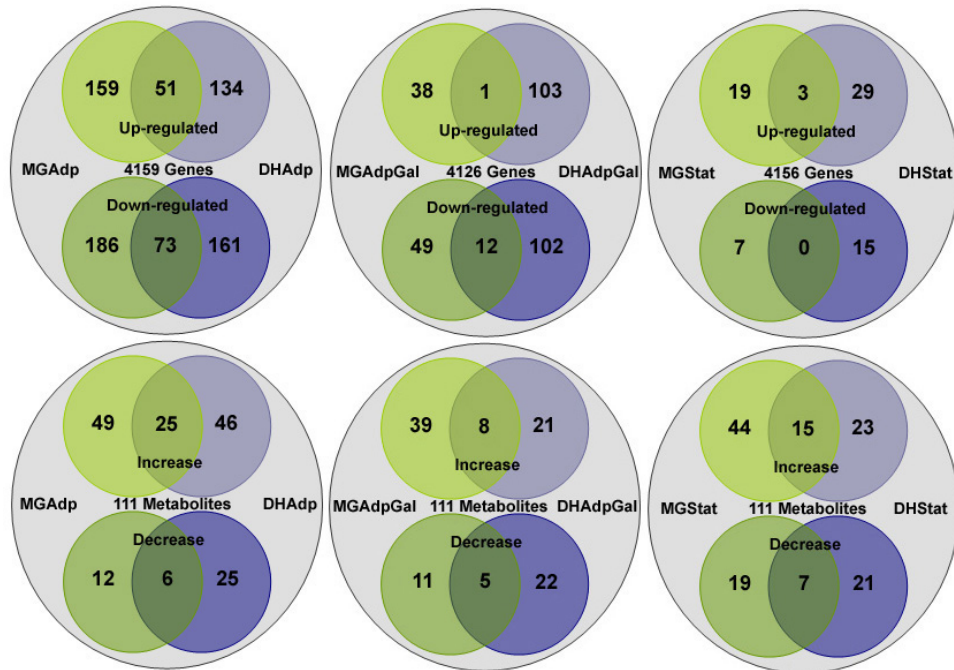
**Fig. 8.2: Typical examples of metabolite differential levels among the ancestral and evolved lines.**

**(a)** Sections of chromatograms showing orotate or orotic acid (denoted by an arrow) abundance among all the lines. **(b)** Mass spectrum of orotate purified standard and mass spectrum of the identified peak as orotate in both strains. **(c)** Box and Whisker plots of metabolites involved in nucleotide related pathways. 1 and 3 represent MG and DH lines (ancestors), 2 and 4 represent MGAdp and DHAdp lines (evolved). The top and bottom of each box represent the 25th and 75th percentiles, the center square indicates the mean, and the extents of the whiskers show the extent of the data. For each metabolite, the maximal measured peak area was normalized to a value of 100.

In MGAdp and DHAdp strains, among all the metabolites, 55% and 64% of metabolites respectively had higher or lower than one fold level change, of which 24% of metabolites were common among both the strains. MGAdpGal and DHAdpGal strains were observed to have 45% and 39% respectively, where 14% of the metabolites were



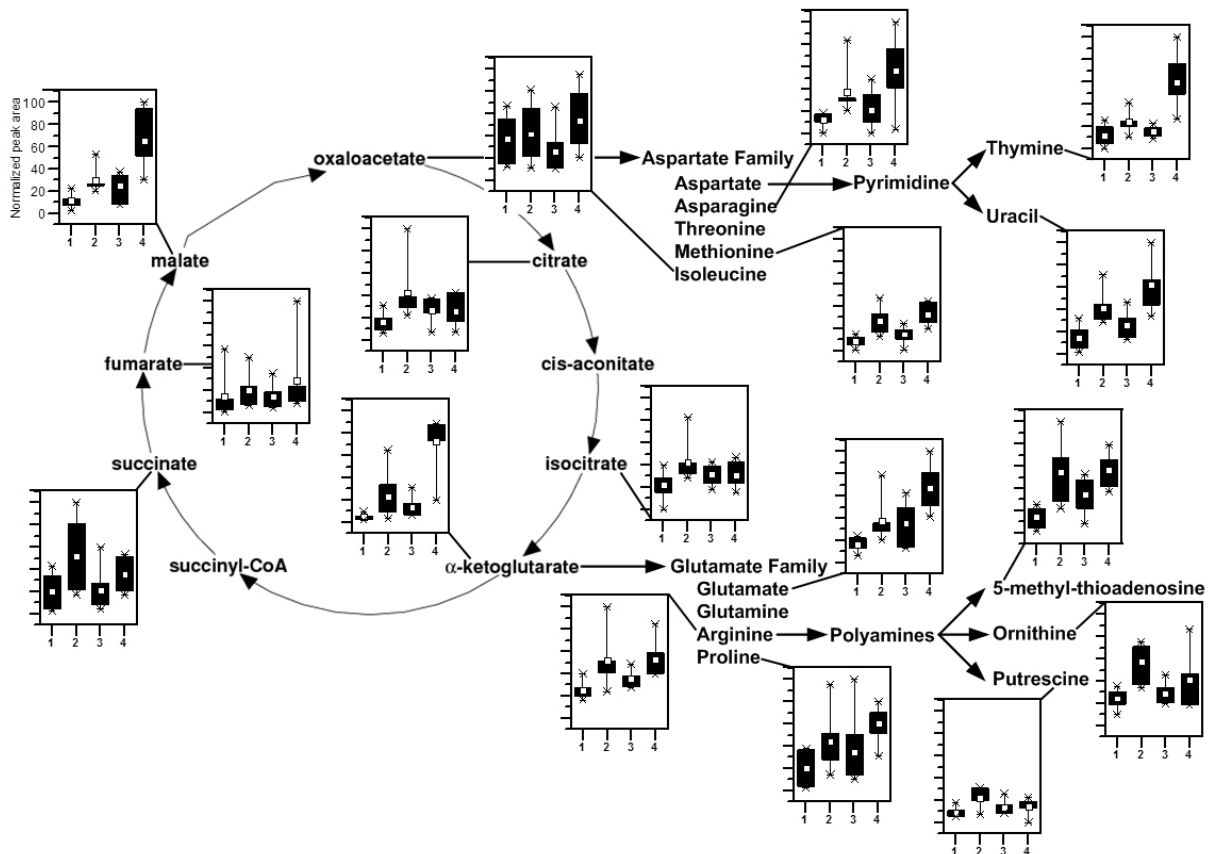
common among both these strains. Likewise MGStat and DHStat showed 57% and 40% and 21% of metabolites were common in both the strains (**Figure 8.3**).



**Fig. 8.3: Venn diagrams comparing number of genes and metabolites changes.** Venn diagrams showing commonalities in metabolites and genes in experimental evolutionary conditions. Venn diagrams demonstrating the distribution of metabolites and genes having one fold higher or lower level at each evolutionary condition are shown for both strains (Green- MG1655, Blue- DH10B).

Those metabolites that showed differences in between the ancestral and the evolved strains fell into groups of metabolites involved in TCA cycle, nucleotide metabolism, aminoacids and their derivatives and polyamine biosynthesis (**Figure 8.1**). For example, metabolites involved in the nucleotide pathway were significantly different between both the ancestral and the evolved strains, MG/MGAdp  $P=0.007$  and DH/DHAdp  $P=0.038$  (Wilcoxon rank sum test; Benjamini-Hochberg corrected). Nucleic acids - adenine, thymine and uracil along with ribose-5-phosphate and orotate (orotic acid) metabolite levels significantly differed in both the Adp evolved strains (**Figure 8.2c**). Orotate is an intermediate in *de novo* biosynthesis of pyrimidine ribonucleotides the level of which were high in the ancestor strains unlike other metabolites which were not the intermediates (**Figure 8.2a,b,c**). Likewise, the levels of metabolites involved in the TCA cycle were significantly different for both the ancestral and the evolved strains, MG/MGAdp  $P=3.70E-06$  and DH/DHAdp  $P=0.026$  (Wilcoxon rank sum test; Benjamini-Hochberg corrected). An overview of the TCA cycle and the diversion of its key intermediates display clear differences in the metabolite levels among the Adp evolved strains and its

ancestors in both the strains (**Figure 8.4**). Since the TCA cycle is the first step in gener-



**Fig. 8.4: The levels of metabolites involved in the TCA cycle and the diversion of key intermediates to biosynthetic pathways.**

In the Box and Whisker plots, 1 and 3 represents MG and DH lines (ancestors), 2 and 4 represents MGAdp and DHAdp lines (evolved). The top and bottom of each box represent the 25th and 75th percentiles, the center square indicates the mean, and the extents of the whiskers show the extent of the data. For each metabolite, the maximal measured peak area was normalized to a value of 100.

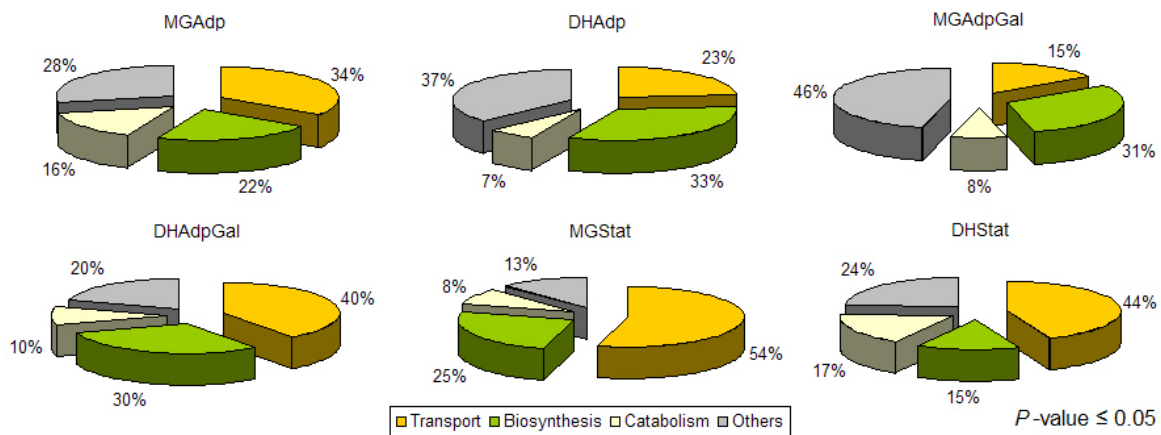
ating precursors for various biosyntheses and one of the main energy producing pathways in a cell, these metabolite level changes can be expected to play a vital role in adaptive evolution of these evolved strains which exhibit increased fitness in growth when compared with their ancestor strains.

## 8.2 Gene expression profiling

Several studies have used gene expression profiling to study molecular evolution, but these studies were confined to a single type of evolution process and were focused on a single molecular aspect that characterizes a cell (transcript abundance) [5, 6, 7, 8]. In our report, we focused on three evolutionary conditions of two strains and two molecular aspects in

a cell (transcript and metabolite abundance). This approach provided us a chance to integrate both metabolome and transcriptome data sets to better understand the process of evolution in laboratory conditions.

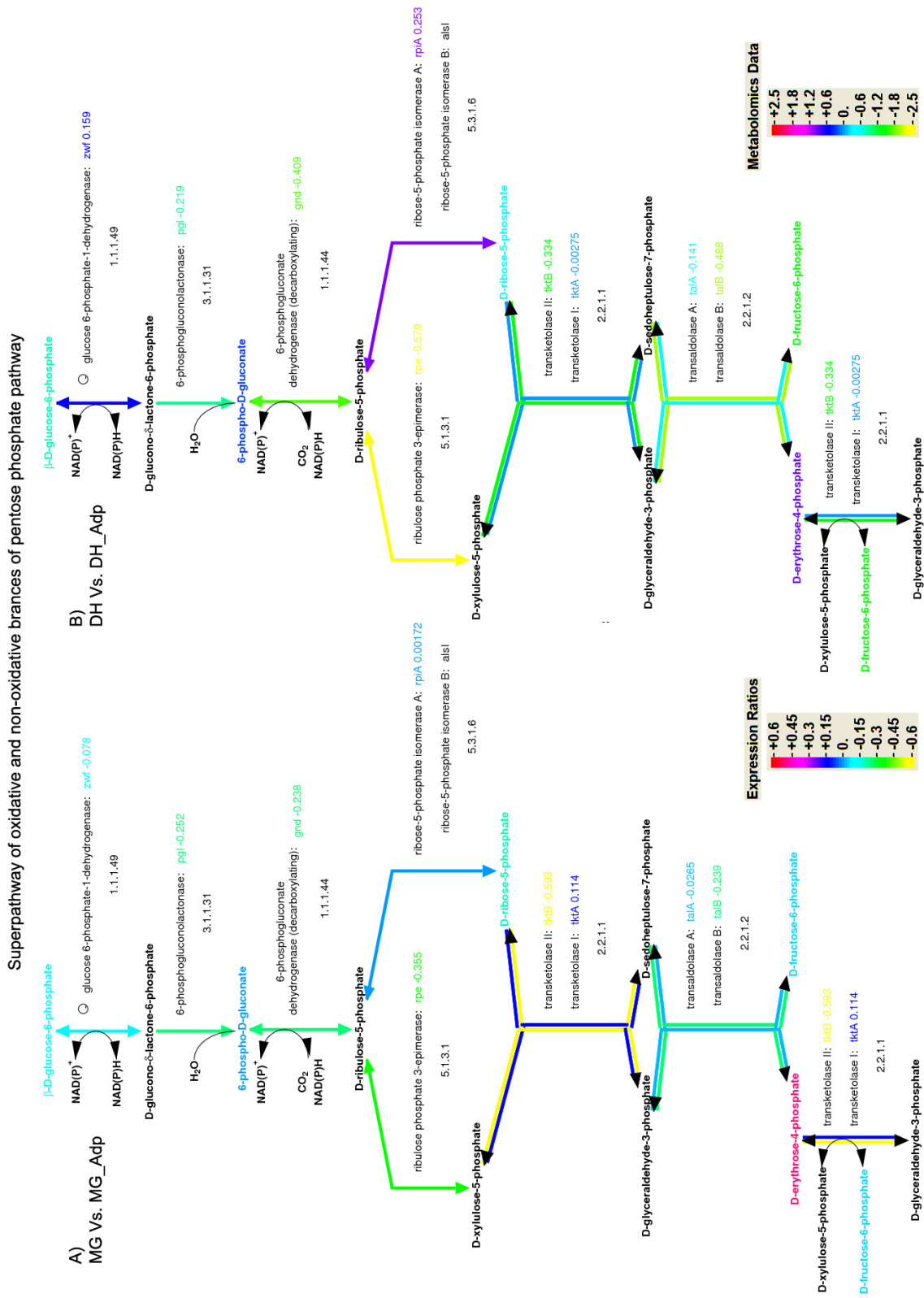
Using the whole genome microarray consisting of 4,288 ORFs, we compared expression levels of the transcripts in MG/MGAdp and DH/DHAdp lines. Among 4159 genes, 8% and 7% showed altered expression levels respectively, among these 19% of the genes were common to both the strains. In MGGal/MGAdpGal and DHGal/DHAdpGal comparison, of 4126 genes it was observed that there were 2% and 5% change, respectively, where 4% of these genes were common among both these strains. Likewise, it was observed in comparing MG/MGStat and DH/DHStat that 1% of the 4156 genes had altered expression levels in both cases, among that 4% of genes were common in both the strains (**Figure 8.3**). In all comparisons, statistically significant functional categories with a  $P \leq 0.05$  (Wilcoxon rank sum test) that did show differences between the ancestral and the evolved strains fell into broad groups of genes involved in transport, biosynthesis and catabolism (**Figure 8.5**). The gene expression changes associated with these main and



**Fig. 8.5: Broad functional annotations of the transcriptome profiling data.** The pie charts of individual evolutionary experimental conditions show the distribution of differentially regulated functional modules, having  $P \leq 0.05$  of Wilcoxon rank sum test. For each evolutionary condition the details of all functional modules and its significant values are provided in supplementary information (**Table 19.6**).

broad functional categories consist of sub-functional categories emphasizing specific functions. (Supplementary Material; **Table 19.6**). For example, genes involved in the pentose phosphate pathway were significantly differentially expressed between the ancestral and the evolved strains of the Adp lines (MG/MGAdp  $P = 0.036$ ; DH/DHAdp  $P = 0.019$ ) (**Figure 8.6, 8.7, 8.8**).

The pentose phosphate pathway produces the precursors (pentose phosphates) for ribose and deoxyribose in the nucleic acids. By the accumulation of nucleic acid metabo-



**Fig. 8.6: Pentose phosphate pathway in excess-nutrient adapted strains.**

The gene expression and metabolite abundance level in evolved strains compared to their ancestor lines are shown according to the log<sub>2</sub> ratio scale shown in the inset. Reaction lines are color-coded according to the relative values of the expression level of the gene that codes for the enzyme that catalyzes that reaction step. Metabolite names are color-coded according to the relative abundance of the metabolite.

lites (**Figure 8.1, 8.2 and 8.4**) and over-expression of pentose phosphate pathway genes in the Adp lines, the involvement of the pentose phosphate pathway in excess nutrient adaptive evolution can be assumed.

### 8.3 *Extent of changes*

To examine the level of metabolite and gene expression changes among all the evolutionary conditions, we applied principal components analysis (PCA). PCA is a technique for multivariate data analysis which reduces the dimensionality and complexity of the dataset without losing the ability to calculate accurate distance metrics. It transforms the metabolome and transcript expression data into a more manageable form in which the number of clusters might be discriminated. When applied to ancestor and Adp lines, both ancestors (MG and DH) cluster together; Adp lines (MGAdp and DHAdp) cluster separately from their ancestor lines, denoting substantial adaptive changes. This pattern was seen in both metabolite and gene expression data as summarized in **Figure 8.9a, d**. When PCA was applied to MGGal, DHGal and AdpGal lines, MGGal and DHGal clustered together; AdpGal lines clustered separately from their ancestor lines, denoting considerable environmental shift due to pleiotropic changes in both metabolite and gene expression data (**Figure 8.9b, e**). Unlike Adp and AdpGal lines, Stat lines displayed dissimilar behaviour; Stat lines (MGStat and DHStat) clustered along with their ancestor lines (MG and DH), denoting few changes between ancestor and evolved strains or diverse changes between the evolved strains in both metabolite and gene expression data (**Figure 8.9c, f**).

### 8.4 *Direction of the observed extent of changes*

For examining the level of observed changes among the strains, we calculated the pairwise Pearson correlation coefficient ( $r$ ) (*PCC*) for all the metabolites and significantly correlating genes. All genes having a threshold of  $r \leq -0.9$  or  $\geq 0.9$  and all metabolites were plotted on both axes of a matrix containing all pairwise metabolite or gene expression profile correlation. When these correlation ( $r$ -values) are colour coded, it allows for a visual inspection to determine the degree of positive and negative correlation among the samples in question. The correlation map of Adp, AdpGal and Stat line comparisons showed various degrees of negative correlation (**Figure 8.9g-l**). Among them, Stat line comparisons (MG/MGStat versus DH/DHStat) displayed a high degree of negative correlation when compared to AdpGal and Adp line comparisons in both metabolite and gene expression correlation maps (**Figure 8.9i, l**) suggesting elevated levels of variabil-



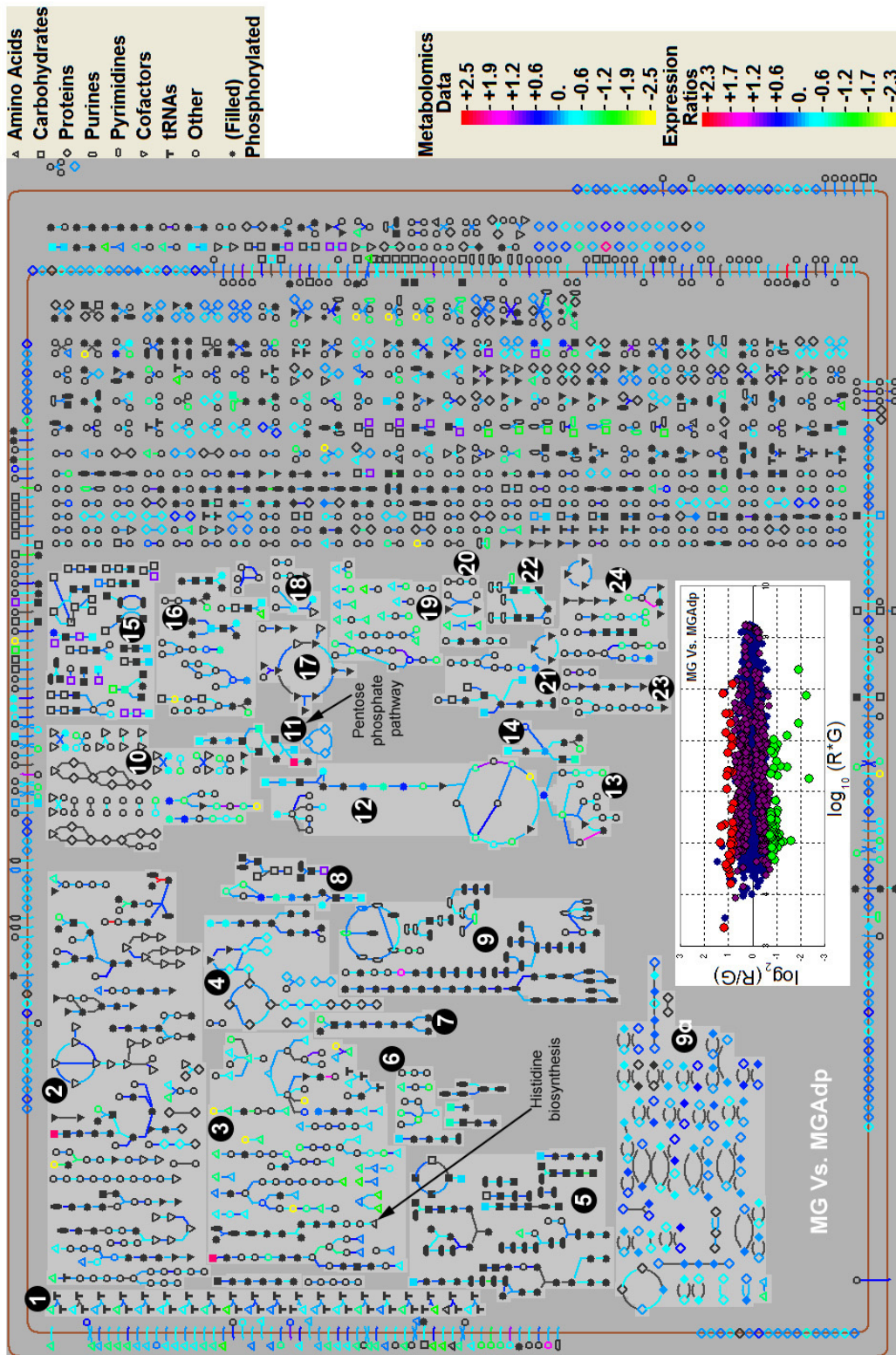


Fig. 8.7: Integration of transcriptome and metabolome data during the comparison of MG and MGAdp strains in excess-nutrient adaptive evolution.

Metabolic overview for *E. coli* illustrating the results of high-throughput transcriptome and metabolome experiments in a global metabolic pathway context. Gene expression levels are mapped to reaction steps involved in a metabolism, and the range of data values levels in a given experimental dataset is mapped to a spectrum of colors. Similarly, for metabolomics experiments, compound nodes are colored according to the data value for the corresponding compound.

- (1- Biosynthesis - Aminoacyl-tRNAs, 2- Biosynthesis - Cofactors, prosthetic group and electron carriers, 3- Biosynthesis - Amino acids, 4- Biosynthesis - Fatty acids and lipids, 5- Biosynthesis - Cell structure, 6- Biosynthesis - Amines and polyamines, 7- Biosynthesis - Secondary metabolites, 8- Biosynthesis - Sugars and polysaccharides, 9- Biosynthesis - Nucleosides and nucleotides, 9a- Biosynthesis -Signal transduction pathways, 10- Respiration, 11- Pentose phosphate pathway, 12- Glycolysis, 13- Fermentation, 14- Superpathway of glycolysis and Entner-Doudorff, 15- Degradation - Sugars and polysaccharides, 16- Degradation - Sugar derivatives, 17- Degradation - Fatty acids and lipids, 18- Degradation - Alcohols, 19- Degradation - Aminoacids, 20- Degradation - Inorganic nutrients, 21- Degradation - Amines and polyamines, 22- Degradation - Nucleosides and nucleotides, 23- Degradation - Aromatic compounds, 24- Degradation - Carboxylates.)



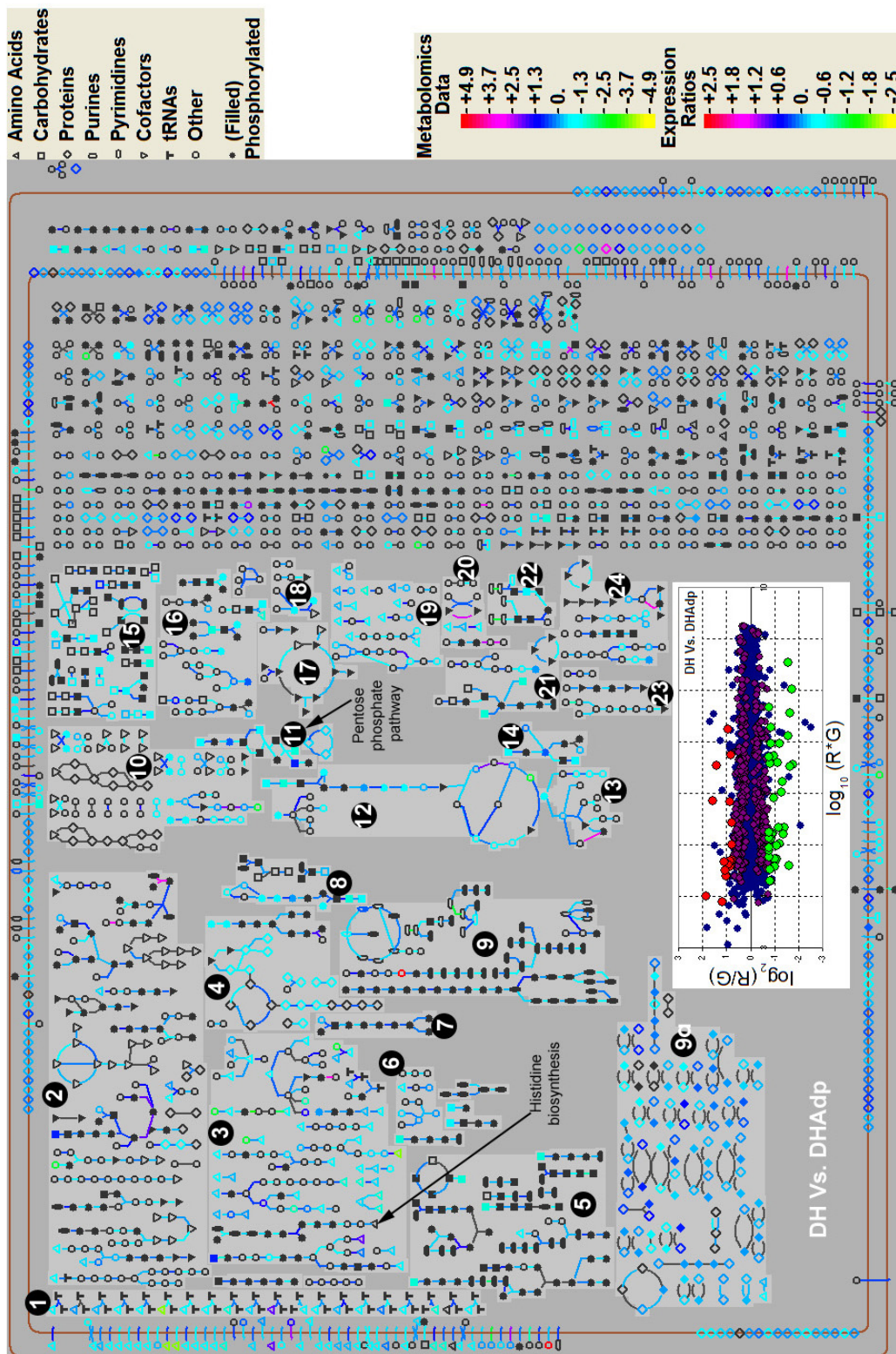


Fig. 8.8: Integration of transcriptome and metabolome data during the comparison of DH and DHAdp strains in excess-nutrient adaptive evolution.

Metabolic overview for *E. coli* illustrating the results of high-throughput transcriptome and metabolome experiments in a global metabolic pathway context. Gene expression levels are mapped to reaction steps involved in a metabolism, and the range of data values levels in a given experimental dataset is mapped to a spectrum of colors. Similarly, for metabolomics experiments, compound nodes are colored according to the data value for the corresponding compound. Various metabolism related pathways are labeled as Figure 8.7

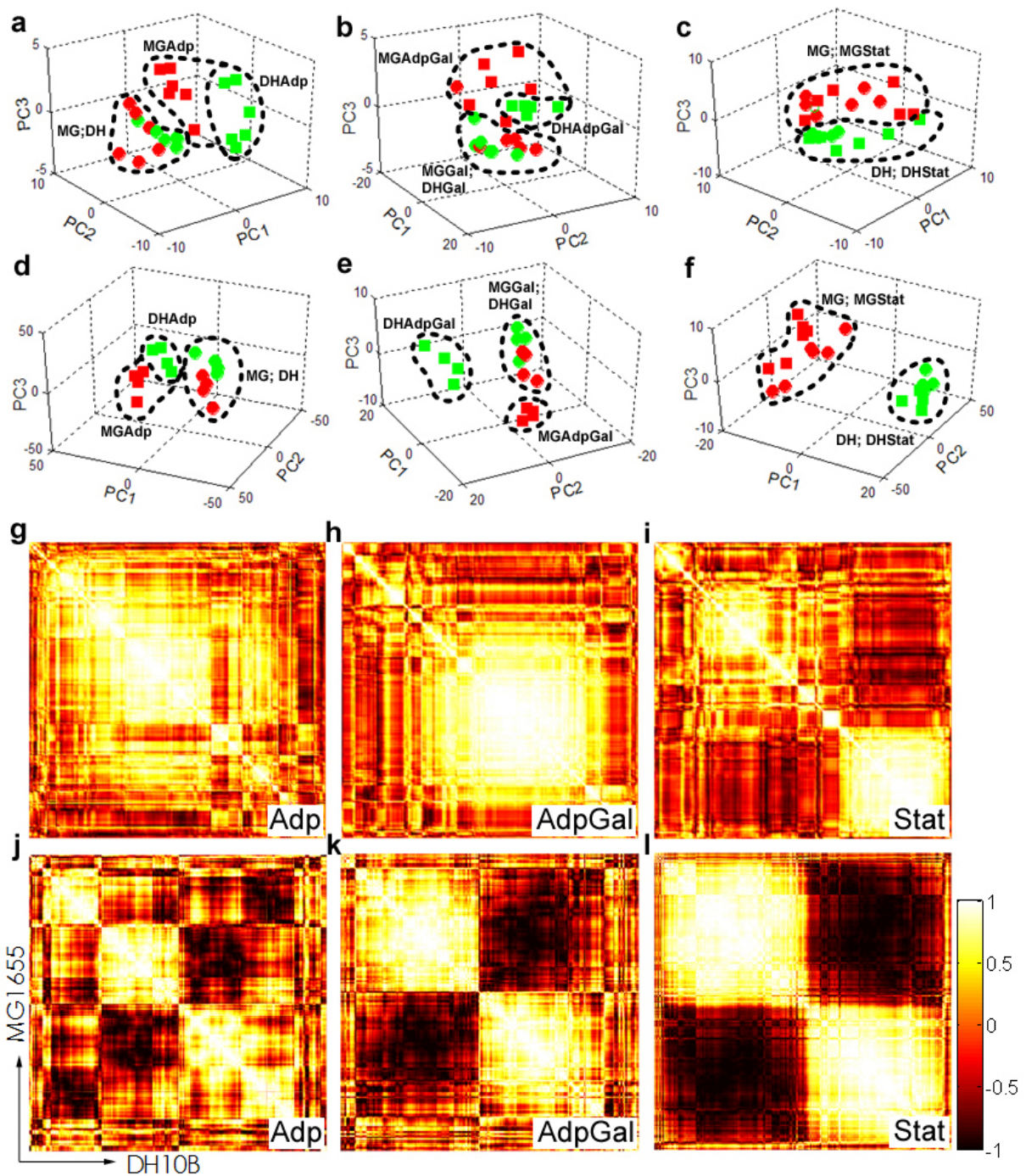


Fig. 8.9: **The extent of changes in experimental evolution among the strains.** (a-f) PCA analysis of the metabolome (a-c) and transcriptome (d-f) data, each data-point represents an experimental sample plotted using the first three principal components. PCA was carried out on the log transformed mean-centered data matrix using all identified metabolites and the genes with  $P \leq 0.05$  (Student t-test) in at least one strain. (g-l) Pairwise correlation maps of the metabolome (g-i) and transcriptome (j-l) data among the strains, using Pearson correlation coefficient ( $r$ ). All the metabolites and the genes having a threshold value of  $r \leq -0.9$  or  $\geq 0.9$  were plotted colour coded on both axes of a matrix containing all pairwise metabolite or gene expression profile correlation. Darker spot indicate higher degree of negative correlation among the strains. Both the analyses were carried out using Matlab 6.5 (The MathWorks, Inc.).



ity in the Stat lines. The correlation map of Adp line comparison (MG/MGAdp versus DH/DHAdp) exhibited a lower degree of negative correlation than the other line comparisons in both metabolite and gene expression correlation maps (**Figure 8.9g, j**) denoting a reduced level of variability in the Adp lines.

### 8.5 Metabolite to gene correlation network analysis

It has been demonstrated that functionally related genes are preferentially linked in co-expression networks [90]. By integrating and comparing the gene expression and metabolite profile patterns, we were able to explore the connections between the gene-to-gene and gene-to-metabolite links and associated functions (**Figure 8.10a**) by assuming that the more similar the expression pattern is, the shorter the distance between genes and/or metabolites in the coexpression network. Relative transcript amounts of all genes and

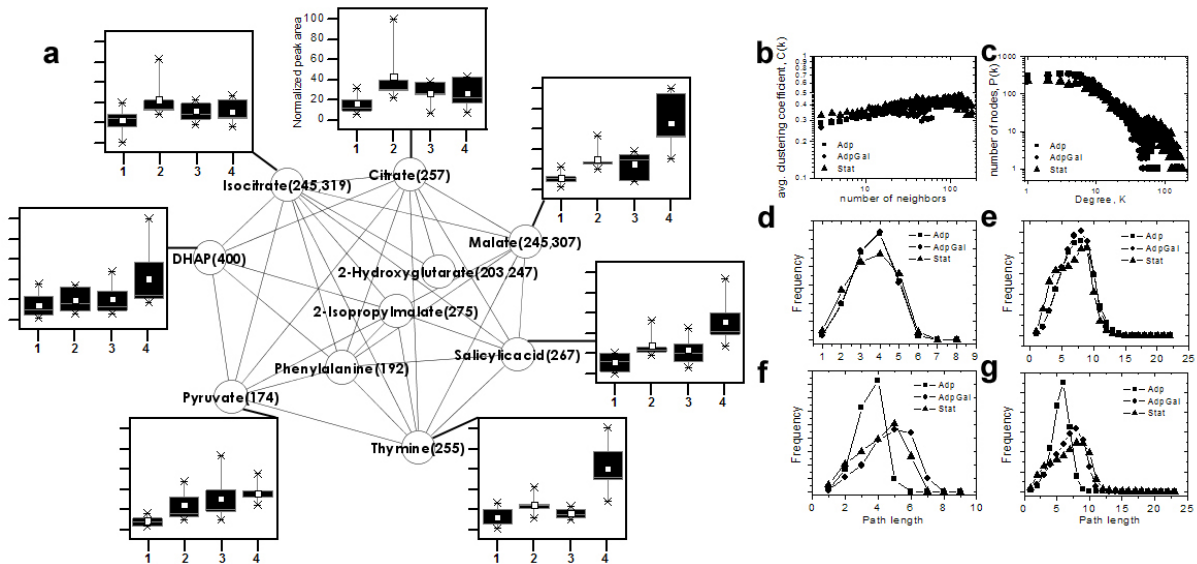


Fig. 8.10: **Metabolite-to-gene correlation network analyses.**

(a) Substructure extracted from Adp correlation network with MCODE algorithm, showing preferentially linked functionally related metabolites. In the Box and Whisker plots of the metabolites 1 and 3 represents MG and DH lines (ancestors), 2 and 4 represents MGAdp and DHAdp lines (evolved). (b-g) Topological properties of all evolution specific coexpression networks. (b) Degree distribution of the clustering coefficients of all the evolution specific network entities. The average clustering coefficient of all the nodes was plotted against the number of neighbours. (c) Degree distribution of the networks, the number of nodes with a given degree ( $k$ ) in the networks approximates power-law ( $P(k) \approx k^{-\gamma}$ ; Adp  $\gamma=1.70$ , AdpGal  $\gamma=1.76$ , Stat  $\gamma=1.32$ ). Distribution of the shortest path between pairs of nodes in the evolution specific (d, e) and intersection (f, g) networks; constructed with  $PCC$  threshold of 0.8 (d, f) and 0.9 (e, g).

relative concentrations of all non-redundant metabolites were combined to form distance

matrices which were calculated by using the Pearson correlation coefficient  $r$  to build coexpression networks. In many cases there were striking relationships between network substructure, gene or metabolite function and coexpression (**Figure 8.10a**). The coexpression network analysis provides a possibility to use it as a quantifiable and analytical tool to unravel the relationships among cellular entities that govern the cellular functions [91].

All-against-all metabolite and gene expression profile comparison for Adp, AdpGal and Stat matrices were used to generate evolution specific coexpression networks constructed using (*PCC*). There was a significantly strong dependence between coexpression and functional relevance of the networks, stating a strong potential of coexpression network analysis (**Figure 8.10a**). In coexpression networks, nodes correspond to genes or metabolites, and edges link two genes or metabolites if they have a threshold correlation coefficient ( $r$ ) at, or above which, genes or metabolites are considered to be changed differentially, exhibiting similar behaviour. Correlation networks as such inherently contain corresponding large noise components which were largely eliminated by setting the threshold of  $r$  at 0.9. The correlation networks based on the high threshold  $r$  of 0.9 reported here are less likely to contain putative noise and simultaneously they are sufficiently dense for analyses of topological properties.

## 8.6 *Evaluation of evolution specific networks*

With respect to a number of parameters describing their common topological properties, all evolution specific coexpression networks (Adp 4170 nodes, 23086 edges; AdpGal 4136 nodes, 20501 edges; and Stat 4166 nodes, 54028 edges) were found to be similar except for the average degree. The average degree ( $\langle k \rangle$ ) is the average number of edges per node [91]. The Stat coexpression network shows higher  $\langle k \rangle$  than the Adp and AdpGal networks, which is consistent with their greater numbers of edges.  $\langle k \rangle$  gives only a rough approximation of how dense the network is. The average clustering coefficient,  $\langle C \rangle$ , determines the network density and characterizes the overall tendency of nodes to form clusters [91]. For all the evolution specific coexpression networks, the  $\langle C \rangle$  was approximately equal and high ( $\approx 0.05$ ), when compared to randomly generated networks of similar size, the observed  $\langle C \rangle$  was quite low ( $\approx 0.0008$ ). The average path length ( $\langle l \rangle$ ) is the average shortest path between all pairs of nodes [91]. For all the evolution specific coexpression networks, the  $\langle l \rangle$  was approximately equal and low ( $\approx 6.97$ ) (**Figure 8.10**). When analysing the networks' generic features, the clustering coefficient  $C(k)$  of all the networks were more or less constant, implying that they do not exhibit a hierarchical structure (**Figure 8.10**). Node degree ( $k$ ) distribution of all the networks

appear to have an exponential drop-off in the tail, following a power-law (**Figure 8.10**). Overall, these evaluations suggest that the global properties of these evolution specific coexpression networks are indistinguishable.

### 8.7 *Evolution specific intersection networks*

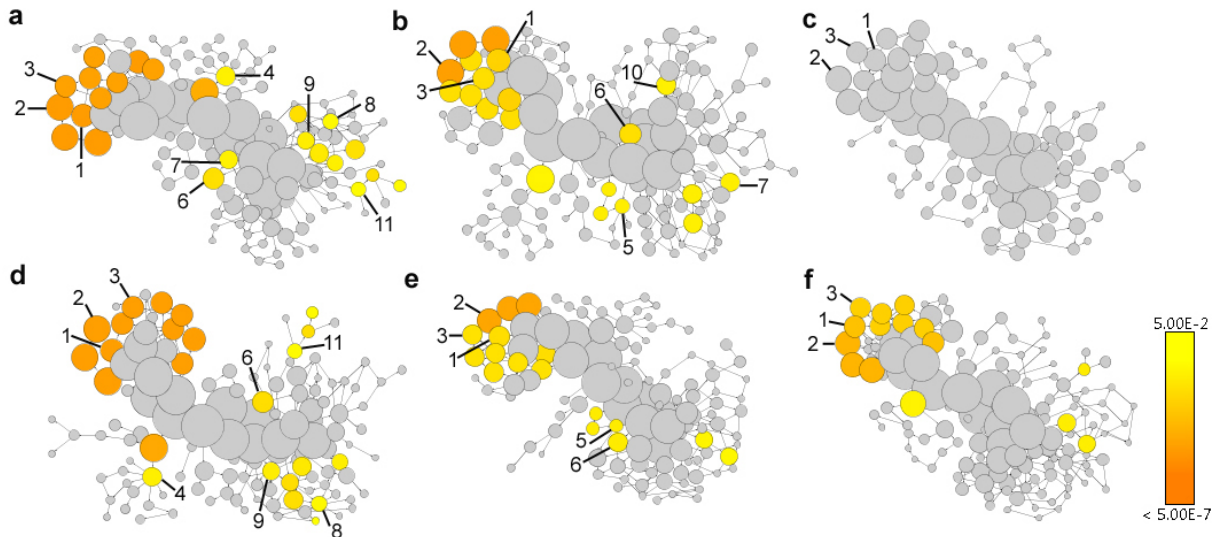
Strain and evolution specific networks were screened for the set of nodes  $N$  where there is a link ( $r \geq 0.9$ ) between two nodes  $a$  and  $b$  in both the strains in the particular evolution type in context to build evolution specific intersection networks. By examining the intersection networks of both the strains, we found that the path length distribution varied among the networks. All intersection networks differed in  $\langle k \rangle$  consistent with their varying numbers of edges. The  $\langle C \rangle$  was slightly higher in the Adp intersection network ( $\langle C \rangle$  Adp intersection= 0.113; AdpGal intersection= 0.07; Stat intersection= 0.089) demonstrating high network density and tendency of nodes to form clusters in the Adp intersection network. The average path length ( $\langle l \rangle$ ) was almost equal in all networks but its distribution in the Adp intersection network differed, indicating high network navigability (**Figure 8.10**). From the observations of the global properties of the evolutionary specific intersection networks, Adp intersection network can be distinguished from other intersection networks demonstrating its unique characteristics.

### 8.8 *Parallelism and functional relevance of molecular evolution*

The generated networks were examined for functional coherence by assigning Gene ontology (GO) functional annotations to the networks' entities, and the level of parallelism in the representation of these functional categories was elucidated. Parallel evolution is the independent development of similar traits in distinct but evolutionarily related lineages through similar selective factors on both lines [92]. Parallel evolution of similar traits across both lines are used as an indicator that the change is adaptive [93]. Previous studies in *E.coli* and *S.cerevisiae* have shown parallel changes in independently adapted lines of replicate populations by utilizing gene expression profiling [5, 7].

Here we examined the parallelism of metabolite and gene expression levels among the evolved lines of different populations which exhibited similar growth behavior. For examining the functional coherence and parallelism among the evolution processes, we mapped the GO functional annotations to the corresponding evolution specific coexpression networks and we attempted to address the extent to which these coexpressed

entities represent functionally related categories. By mapping GO functional categories to the coexpression networks, statistically and significantly overrepresented functional categories were color coded according to the hypergeometric test  $P$ -value which was corrected by Benjamini and Hochberg False Discovery Rate (**Figure 8.11**). To examine



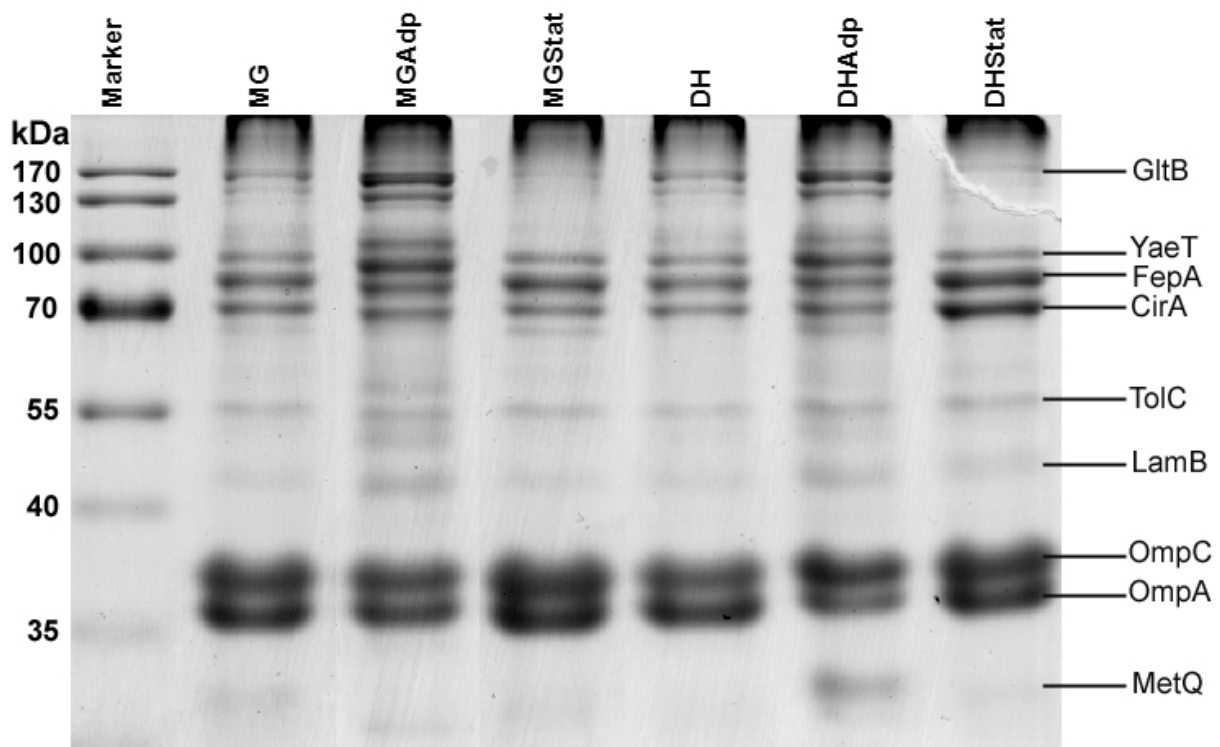
**Fig. 8.11: Parallelism and functional relevance of molecular evolution.** GO functional annotations were mapped to the corresponding evolution specific coexpression networks and examined for commonalities in the coexpressed entities representing functional related categories. Each node represents a GO functional category and the area of a node is proportional to the number of genes in the network matrix to the corresponding GO category. Statistically and significantly overrepresented categories are colour coded based on the hypergeometric test  $P$ -value which was corrected by Benjamini and Hochberg False Discovery Rate. Grey nodes are not significantly overrepresented. **(a-c)** GO annotations were mapped to the evolution specific coexpression networks namely (a) Adp, (b) AdpGal and (c) Stat. **(d-f)** GO annotations mapped evolution specific intersection coexpression networks namely (d) Adp intersection, (e) AdpGal intersection and (f) Stat intersection. (1- membrane, 2- cell wall (sensu Bacteria), 3- inner membrane, 4- transporter activity, 5- transport, 6- catabolism, 7- cellular catabolism, 8- amino acid metabolism, 9- nitrogen compound metabolism, 10- carbohydrate metabolism, 11- energy derivation by oxidation of organic compounds). Not all overrepresented categories are labelled due to the interdependency of functional categories in the GO hierarchy.

the parallelism of evolutionary processes in both the strains in context of GO functional categories, we mapped the GO functional annotations to the coexpression networks ( $r \geq 0.9$ ) generated by merging the data matrix of both the strains forming three evolution specific coexpression networks namely, Adp, AdpGal and Stat networks (**Figure 8.11a, b, c**). The level of parallelism differed among these networks. In the Adp network, for example, membrane, cell wall (sensu Bacteria), inner-membrane, transporter activity, catabolism, and cellular catabolism functional categories were significantly overrepresented ( $P \leq 0.05$ ) (**Figure 8.11a**). In the AdpGal network membrane, cell wall

(sensu Bacteria), inner-membrane, transport, catabolism, and cellular catabolism functional categories were overrepresented ( $P \leq 0.05$ ) (**Figure 8.11b**). However, in the Stat network, none of the GO functional categories were significantly overrepresented, denoting decreased level of parallelism among both the strains (**Figure 8.11c**). Further examination of parallelism of evolutionary processes was extended to intersection coexpression networks, which were created by selecting the nodes that are connected ( $r \geq 0.9$ ) in both the strains in the particular evolutionary process in question. By examining the parallelism in these intersection coexpression networks, apart from other functional categories, we found that the commonly observed distribution of statistically overrepresented GO categories in all the coexpression networks belonged to membrane-associated GO functional categories (**Figure 8.11d, e, f**).

### *8.9 Parallelism in outer-membrane protein expression*

To further examine the extent of parallel evolutionary changes, we determined the expression levels of proteins associated with the outer-membrane (OM) of the ancestor and evolved strains, whose membrane-related GO functional categories were overrepresented in the evolution specific coexpression networks (**Figure 8.11**). OM protein levels revealed substantial differential expression among the ancestor and evolved strains (**Figure 8.12**). In Adp lines, GltB (glutamate synthase (NADPH) large chain precursor), LamB (maltose high-affinity receptor) and YaeT (polypeptide involved in outer-membrane protein biogenesis) proteins were over-expressed, whereas in Stat lines, FepA (outer receptor for ferric enterobactin), CirA (outer-membrane receptor for iron-regulated colicin I receptor), OmpC (outer-membrane porin) and OmpA (outer-membrane porin) proteins were differentially over-expressed (**Figure 8.12**). Significantly, parallelism in the level of protein expression patterns in these evolved strains and involvement of the outer-membrane proteins in these evolutionary processes was observed.



**Fig. 8.12: Parallelism and functional significance in the outer-membrane protein expression.**

SDS gel electrophoresis of the protein samples obtained from the outer-membrane of the ancestor and evolved lines showing the identified proteins by peptide mass fingerprinting.

---

## CHAPTER 9 DISCUSSION (II)

---

In this study, we have examined the metabolome and transcriptome profiles of excess nutrient adaptive evolution, immediate environmental shift changes and prolonged stationary phase evolution in two strains of *Escherichia coli* K-12. We found significant influence of genes involved in transport and membrane related functional categories in all evolutionary conditions determined in this study. In earlier studies, during prolonged nutrient-limited chemostat culture of bacterial populations, it has been reported that the populations tend to mutational adaptation in transport systems to increase the efficiency of utilizing the limiting nutrient [94, 95, 96, 97]. For example, glucose-limited chemostat evolved strains attained diverse mutations at several loci in LamB porin which increased glucose permeability [96, 97, 98]. Earlier study of adaptation of *Ralstonia* in selective environments resulted in the morphological changes in the outer cell envelope in all the lineages examined [99].

In adaptation to excess-nutrient resources, the Adp lines exhibited higher levels of metabolites involved in nucleotide pathway and TCA cycle and its intermediates (**Figure 8.1, 8.4**). In line with these observations, the expression levels of genes involved in these pathways were also over-expressed in the Adp lines (**Figure 8.7, 8.8**). Specifically, the pentose phosphate pathway (produces pentose phosphates for nucleic acid synthesis) was differentially regulated, along with the histidine biosynthesis pathway which shares metabolites with the purine and nucleotide biosynthesis pathways (**Figure 8.6, 9.1**).

For example, glutamate, which is involved in the *de novo* biosynthesis of purine nucleotides and various other pathways as a reactant, was accumulated in higher amounts in Adp lines. In line with this observation, the genes involved in the glutamate biosynthesis and the protein coding for glutamate synthase (GltB) were up-regulated in the Adp lines (**Figure 8.12**). Taken together, the increased growth fitness in Adp lines when compared to their ancestor lines can be assumed to be due to the differential levels of, 1) TCA cycle components which are the first step in generating precursors for several biosyntheses, 2) components involved in pentose phosphate pathway which are the main source of precursor metabolites for the biosynthesis and the main producer of NADPH which is utilized in several biosynthesis pathways; however, the involvement of these pathways in the growth fitness has to be confirmed by additional studies. Our finding that the central metabolism is altered in excess nutrient and famine conditions is consistent with a previously published study focusing on adaptive evolution in yeast in glucose-limited chemostat

## Histidine biosynthesis

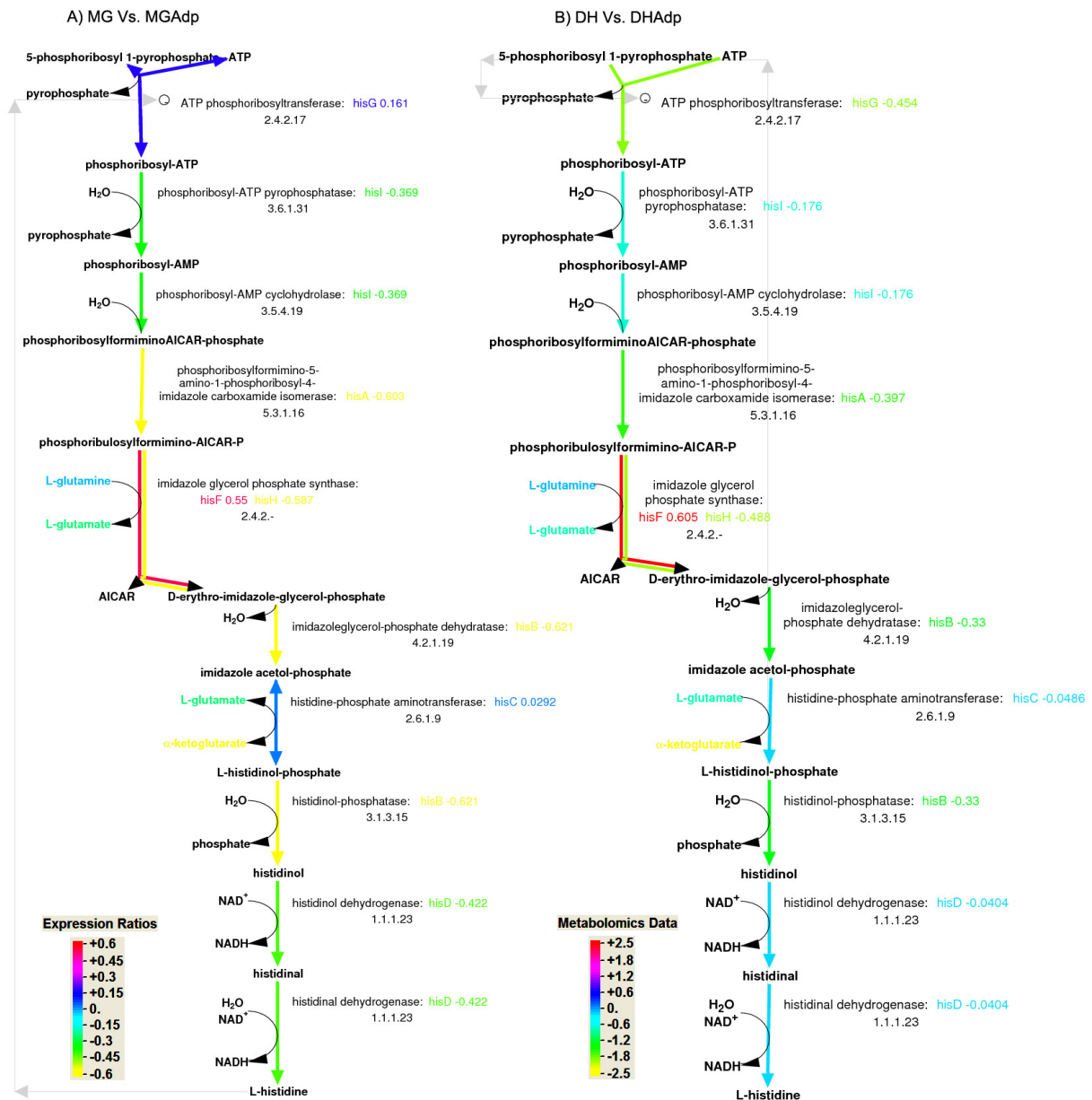


Fig. 9.1: Histidine biosynthesis pathway in excess-nutrient adapted strains. The gene expression level and the metabolite abundance level in the evolved strains compared to their ancestor lines are shown according to the log<sub>2</sub> ratio scale shown in the inset. The reaction lines are color-coded according to the relative values of the expression level of the gene that codes for the enzyme that catalyzes that reaction step. Metabolite names are color-coded according to the relative abundance of the metabolite.



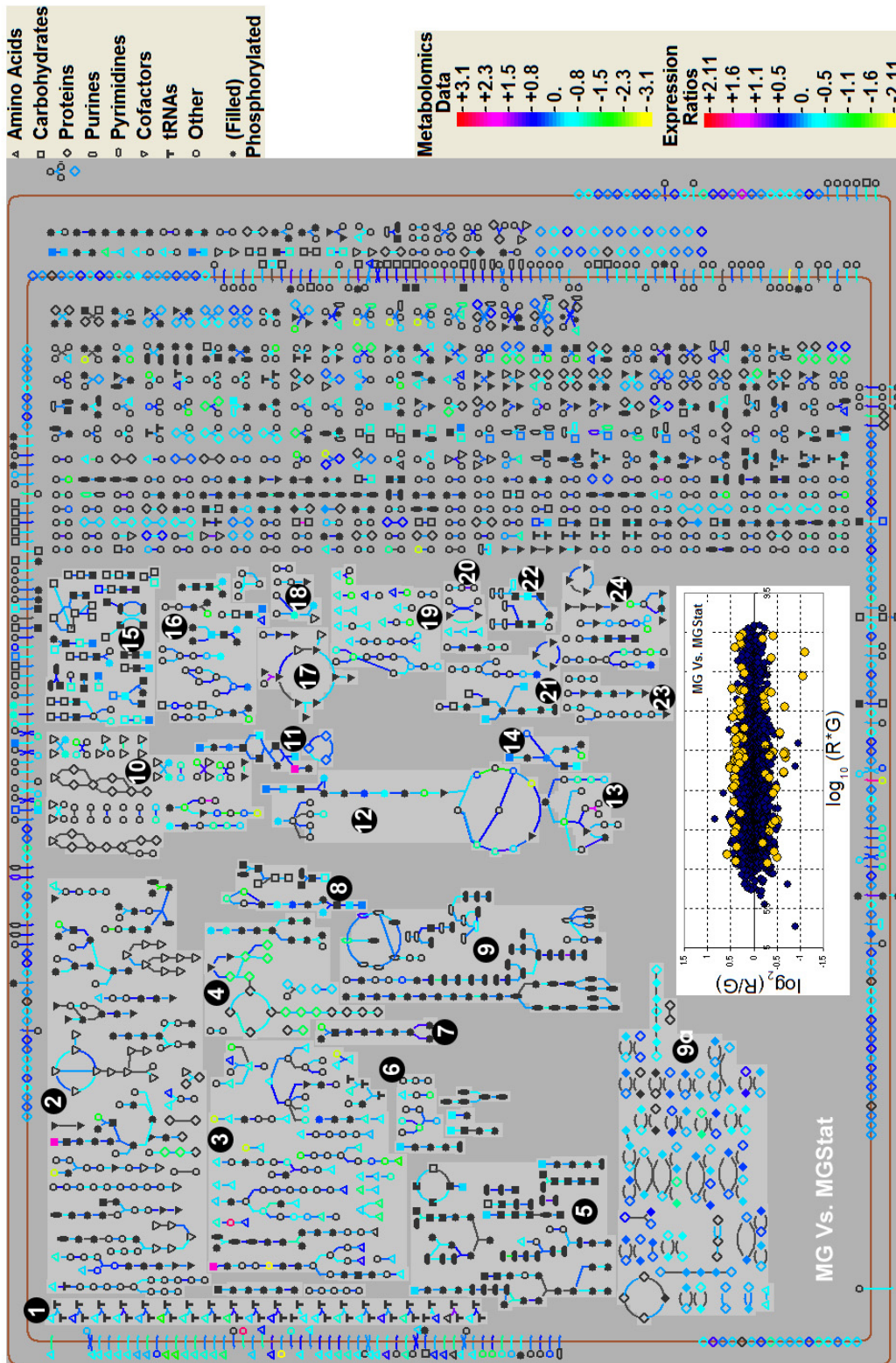


Fig. 9.2: Integration of transcriptome and metabolome data during the comparison of MG and MGStat strains in prolonged stationary phase evolution.

Metabolic overview for *E. coli* illustrating the results of high-throughput transcriptome and metabolome experiments in a global metabolic pathway context. Gene expression levels are mapped to reaction steps involved in a metabolism, and the range of data values levels in a given experimental dataset is mapped to a spectrum of colors. Similarly, for metabolomics experiments, compound nodes are colored according to the data value for the corresponding compound. Various metabolism related pathways are labeled as Figure 8.7

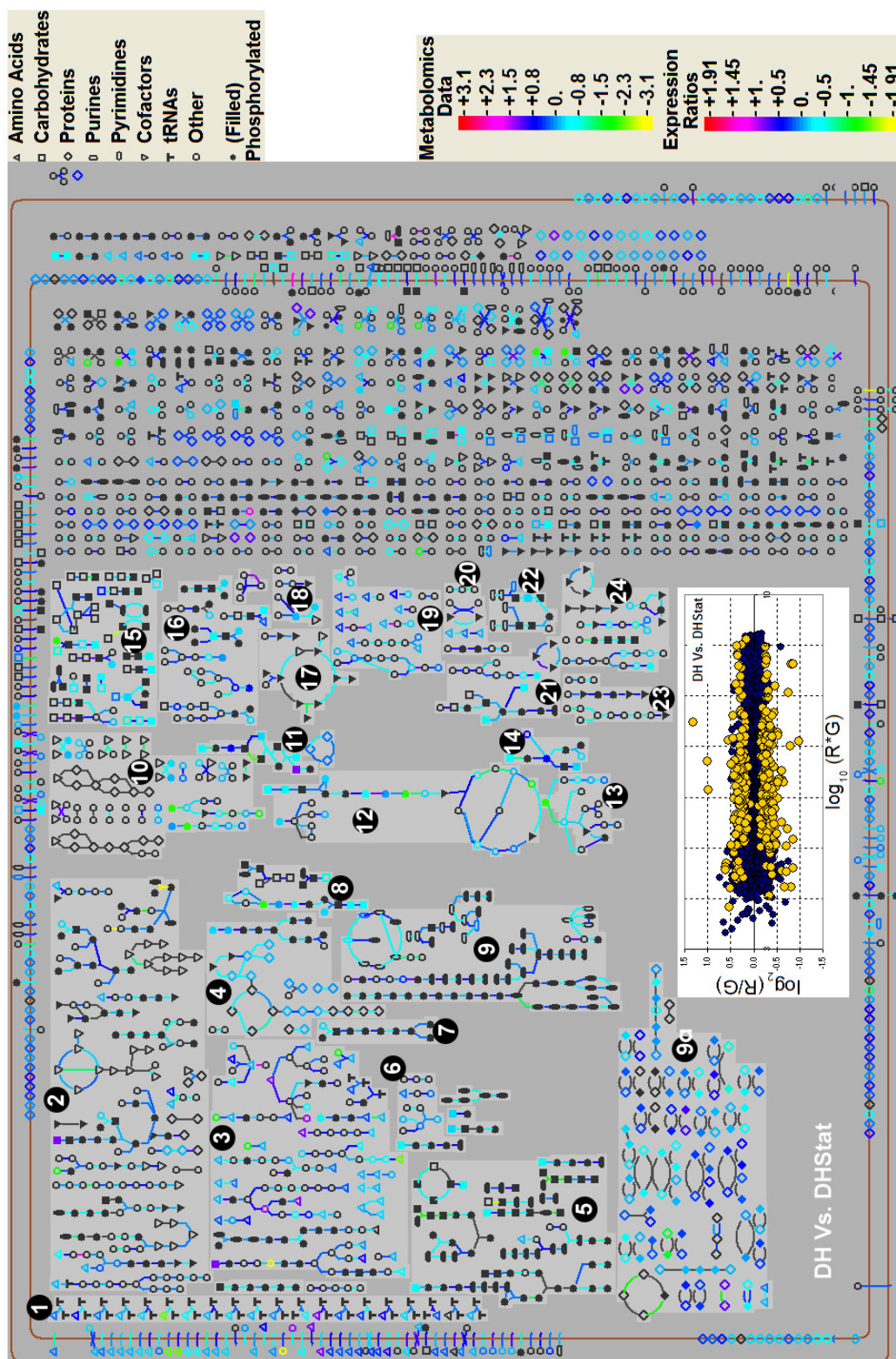


Fig. 9.3: Integration of transcriptome and metabolome data during the comparison of DH and DHStat strains in prolonged stationary phase evolution.

Metabolic overview for *E. coli* illustrating the results of high-throughput transcriptome and metabolome experiments in a global metabolic pathway context. Gene expression levels are mapped to reaction steps involved in a metabolism, and the range of data values levels in a given experimental dataset is mapped to a spectrum of colors. Similarly, for metabolomics experiments, compound nodes are colored according to the data value for the corresponding compound. Various metabolism related pathways are labeled as Figure 8.7



experiments, which showed gene expression variation in glycolysis, the tricarboxylic acid cycle and metabolite transport [5].

In long-term stationary-phase cultures, cells lose their integrity and release their components into the medium as cells enter death phase [70]. For cell maintenance and growth, the surviving cells scavenge nutrient sources from the cellular debris (amino acids from proteins, carbohydrates from the cell wall, lipids from cell membrane material and DNA) of their dead siblings [70]. This nutrient-scavenging process due to nutrient limitation enhances the availability of carbon sources by reconstruction of the outer-membrane (OM) composition and by improving its permeability [100]. The OM of *E. coli* consists of a lipid bilayer structure composed of an outer layer consisting of lipopolysaccharide (LPS) and an inner layer consisting of phospholipids [101]. The genes involved in the biosynthesis pathways of fatty acids (key building blocks for the phospholipid components of cell membranes) and lipids were over-expressed in Stat lines (**Figure 9.2, 9.3**). Other major components of the outer-membrane are proteins - largely consisting of porins which coexist with LPS [102]. OM of the cell is the first point of contact with the external environment and therefore its cellular constituents may be the most sensitive to the external environment. Consistent with this hypothesis, OM proteins FepA, CirA, OmpC, and OmpA were differentially over-expressed in Stat lines (**Figure 8.12**), and the genes belonging to the membrane-associated GO functional categories were significantly over-represented in the corresponding evolutionary networks as well (**Figure 8.11f**). This demonstrates the reliability of the correlation network analysis being robust enough to examine the significant changes in the integrated metabolite and gene profiling dataset.

Mutation rates in stationary phase are known to be influenced by the genetic background of the strain [78]. Initial isogenic long-term stationary-phase cultures are highly dynamic and are known to yield different GASP mutations due to significant genotypic diversity in these cultures [70]. Consistent with this hypothesis, when we applied PCA (**Figure 8.9c, f**) and correlation plot analysis (**Figure 8.9i, l**), the metabolite and gene expression levels of Stat lines showed low degree of parallelism when compared with their ancestor lines. Likewise, when GO functional annotations were mapped onto the Stat coexpression network, we found that none of the GO functional categories were significantly overrepresented, denoting a low level of parallelism (**Figure 8.11c**). However, when applied to the Stat intersection coexpression network, membrane-associated GO functional categories were significantly overrepresented (**Figure 8.11f**). These observations demonstrate the parallelism in membrane-associated categories in Stat intersection coexpression network but not in Stat coexpression network propose the existence of parallelism in membrane-associated categories but not in similar membrane-associated genes in Stat lines. From this, we can conclude the involvement of distinct but functionally-related

genes in the parallelism in the Stat intersection coexpression network.

---

## CHAPTER 10 CONCLUSION (II)

---

We analyzed two different strains in three different evolutionary conditions. Integration of metabolome and gene expression data in the context of evolution facilitated the investigation of the path of evolution and its degree of parallelism. Classifying microarray data according to significantly overrepresented Gene ontology (GO) functional categories, showed that the transport related categories comprised the highest overall representation. Similarly, by mapping the GO annotation to the correlation networks, we found that the membrane-associated functional categories were significantly overrepresented. The outer-membrane (OM) of the cell is the first point of contact with the external environment, which acts as a barrier that is quite resistant to unwanted factors as well as acting as a channel for the diffusion of required nutrients. Components of the OM may therefore be the most sensitive cellular constituents to the external environment. Analyses of the outer-membrane proteins of the ancestor and evolved strains showed clear differential regulation of the outer-membrane proteins. In summary, all the evolutionary experiments reported in this study demonstrate the vital role of the involvement of the membrane-associated components in the evolutionary process. These studies show that adaptive evolution in excess-nutrient conditions are appropriate for examining the extent of parallelism in the evolutionary process of the evolved populations, whereas the prolonged stationary phase conditions are useful to understand the evolution of microbial diversity among the evolved populations and the dynamic state of the evolved condition. Such studies would certainly advance understanding of the process of evolution immensely and along with constructed models [103] would certainly be an ideal initial source of data for systems biology of microbial evolution.

# Part III

Integration of proteome with transcriptome and metabolome profile during molecular evolution processes in *Escherichia coli*.

## Abstract

A comprehensive analysis of molecular evolutionary processes can be accomplished by analysing the evolved strain in all three functional levels of a gene (protein, transcript and metabolite). To examine the effect of different molecular evolution processes, we have analysed the variations in protein abundance in evolved and ancestor strains under three different evolutionary conditions (excess-nutrient adaptation, prolonged stationary phase adaptation and adaptation due to environmental shift) and integrated the gene expression and metabolite levels in two different strains of *Escherichia coli* K-12 (MG1655 and DH10B). We found 173 proteomic changes among the 525 identified proteins. A majority of these proteins belonged to membrane related and energy metabolism functional categories. The direction and degree of these proteomic changes varied among the evolutionary conditions analysed and the pattern was consistent with the transcript and metabolite changes. The integration and comparison of data generated from all the three functional levels showed that they were in complete agreement with each other. All the proteome data discussed in this study is accessible in the form of a relational database at <http://2dbase.techfak.uni-bielefeld.de>.

**Authors:** Chandran Vijayendran, Karl Friehs, Karsten Niehaus, and Erwin Flaschel.

**Author's contribution:** CV conducted all the experiments cited in this study, analysed the results and wrote this report. KF was involved in experimental guidance. KN was involved in experimental design. EF is the scientist in whose laboratory all the experiments were conducted and was involved in experimental design.

---

# CHAPTER 11

## INTRODUCTION (III)

---

In the field of molecular evolution, microorganisms are widely employed for carrying out evolution related experiments. Microbial evolution experiments are performed under various types of evolutionary conditions in the laboratory, as their short life span facilitates such studies within a limited time scale, also allowing for easy manipulation of environmental variables. Among these evolutionary experiments, nutrient-limited adaptive evolution [73, 75, 74] and famine-induced prolonged stationary phase evolution [78, 76, 77] are extensively studied. Micro-organisms in these conditions exhibit various specific mutations resulting in phenotypic attributes responsible for fitness in that particular selective environment, demonstrating adaptive evolution [104]. During prolonged carbon starvation, the evolved mutants harbouring selective advantage out-compete their siblings and takeover the culture by their progeny [79, 81, 80]. Laboratory studies under these evolutionary conditions have addressed fundamental questions regarding adaptation processes and selection pressures, thereby explaining modes of evolution [69, 76, 70, 73].

A complete understanding of any molecular process requires quantitative data generated in parallel for transcript, protein and metabolite levels. Existing technological advancements have considerably extended the ability to analyse and generate large data sets, enhancing the possibility of understanding complex biological systems. Integration and interpretation of large data sets generated from cutting-edge as well as conventional technologies in these three functional levels, however, remains a major scientific challenge [105]. Global changes in gene expression are usually analysed using microarrays, providing quantitative information about the gene expression levels whereas, global quantitative profiling methods applied to analyse protein and metabolite levels are still in their infancy. Proteomics by conventional 2-dimensional polyacrylamide gel electrophoresis is still the major method for global proteome analysis [106, 107]. Commonly used methods for metabolite profiling are gas chromatography - mass spectroscopy (GC-MS), liquid chromatography - mass spectroscopy (LC-MS) and nuclear magnetic resonance (NMR). Experiments focusing on the three functional levels would provide important information concerning the hierarchical control of biological processes [108, 24, 109]. Relatively few studies have integrated the data sets generated from the experiments concerning three functional levels for better understanding the biological processes.

In this study, we have applied a global proteomic profiling method and integrated the data sets generated from global transcript and metabolite profiling methods to understand



excess-nutrient adaptive evolution, prolonged stationary phase evolution and pleiotropic effects due to environmental shift in two strains harbouring different genotypes. For this, we have used *Escherichia coli* K-12 strains (MG1655 and DH10B) from (1) serial passage system for excess-nutrient adaptive evolution studies, (2) constant batch culture for prolonged stationary phase evolution and (3) culture with nutrient alteration following adaptation to a particular nutrient for examining pleiotropic effects due to environmental shift. A detailed proteome analysis of these evolved populations would guide towards a better understanding of the evolution processes involved when combined with the data generated from transcriptomics and metabolomics approaches. These data sets obtained from three functional levels would be of vital importance for viewing a global picture of the experimental sample in question. To eliminate the possibility of the strain-dependent phenomenon of evolution and to examine the parallelism of the laboratory evolution process, we examined all the evolutionary processes in two strains. As a result we were able to compare the protein profiles under various conditions for these strains, namely,

<b>Strain abbreviations</b>	<b>Evolved condition</b>
MG and DH	MG1655 and DH10B grown in glucose
MGGal and DHGal	MG1655 and DH10B grown in galactose
MGAdp and DHAdp	MG1655 and DH10B adapted about 1000 generations in glucose
MGAdpGal and DHAdpGal	MGAdp and DHAdp (glucose evolved strains) grown in galactose
MGStat and DHStat	MG1655 and DH10B grown in prolonged stationary phase (37 days)

Tab. 11.1: **Strains and their evolved conditions.**

Several studies have used gene expression [7, 5, 8, 6] and proteome profiling methods [9, 10, 11, 12] to study molecular evolution, but these studies were confined to a single type of evolution process. In this study, we depicted a complete picture of laboratory molecular evolution processes among two different strains under three different evolutionary conditions, namely, excess-nutrient adaptive evolution, prolonged stationary phase evolution and pleiotropic effects due to environmental shift. By global proteome profiling and integrating multidimensional metabolome and gene expression data, we aim to find vital proteins, metabolites and genes involved in the evolutionary processes in these three conditions. To our knowledge, there have been no previous reports on integrating proteome, transcriptome and metabolome data sets involving microbial molecular evolutionary studies.

---

# CHAPTER 12

## MATERIALS AND METHODS (III)

---

### *12.1 Strain and culture conditions*

Both the bacterial strains, MG1655 and DH10B used in this study are derivatives of *E.coli* K-12. All the experiments were conducted in 250 mL of M9 minimal medium supplemented with 4 gL<sup>-1</sup> of glucose or galactose in covered 1 L Erlenmeyer flasks at 37 °C. Adaptation to excess nutrient experiments was carried out in the presence of 4 gL<sup>-1</sup> glucose through serial passage at exponential phase for about 1000 generations. The cells were grown overnight and were diluted by passage into fresh medium. Passage of each culture into fresh medium was conducted in a laminar flow station using standard sterile technique practices. Serial passage were conducted for 37 days at exponential phase for about 1000 generations. For adaptation due to environmental shift experiments, the strains which were adapted to excess nutrient (glucose) condition for about 1000 generations were grown in 4 gL<sup>-1</sup> galactose. For prolonged stationary phase adaptation experiments, both the strains were incubated for 37 days in M9 minimal medium with 4 gL<sup>-1</sup> glucose as initial source of carbon. The evolved populations were frozen using liquid nitrogen and stored in a freezer at -80 °C.

### *12.2 Two-dimensional SDS-PAGE gel electrophoresis*

Approximately 4.1x10<sup>10</sup> cells were harvested from the exponential phase in all experiments. The cells were pelleted down at 6000 rpm at 4 °C for 10 min. Harvested cells were washed with low salt solution (3 mM KCl, 1.5 mM KH<sub>2</sub>PO<sub>4</sub>, 68 mM NaCl, 9 mM NaH<sub>2</sub>PO<sub>4</sub>) trice. The pellet was then resuspended in rehydration buffer (9 M urea, 4% 3-([3-chloramidopropyl]dimethylammonio)-1-propane-sulfonate (CHAPS), 85 mM dithiothreitol (DTT), 0.5 mM pefabloc SC and stored at -20 °C until next use. The cells were lysed by ultrasonication for 10 s, 5 times at 10% of maximum output (Branson sonifier 450). After 30 minutes of incubation at 37 °C with DNase and RNase, eventually the debris was pelleted down and the proteins present in the supernatant were precipitated with acetone at -20 °C overnight. The precipitated proteins were then resuspended in rehydration buffer and 300 µg of protein sample was loaded to 24 cm, pH 4-7 Immobiline dry strips (Amersham Biosciences) along with 1.5 µl (IPG)-buffer pH 4-7 dissolved in it for each strip. The strips were focused on an IPG-phor (Amersham Biosciences) for 1 h

at 0 V, 12 h at 30 V, 2 h at 60 V, 1 h at 1000 V, and at 8000 V until approximately 75,000 Vh was reached. The strips were equilibrated in 5 ml of a solution containing 6 M urea, 50 mM Tris (pH 8.8), 30% (v/v) glycerol, 20 gL<sup>-1</sup> SDS and 20 gL<sup>-1</sup> DTT on a tilt table for 15 min. The solution was discarded and 5 ml of a second solution was added for 15 min containing 6 M urea, 50 mM Tris (pH 8.8), 30% (v/v) glycerol, 20 gL<sup>-1</sup> SDS and 25 gL<sup>-1</sup> iodoacetamide. The second dimension was performed on an EttanDalt (Amersham Biosciences) electrophoresis unit. The strips were placed on a 1.5 mm thick, 12.5% poly-acrylamide gel and sealed with 0.1% agarose in SDS-electrophoresis buffer containing 0.01% brom-phenol-blue. The gel electrophoresis was performed for 30 min at 3 W per gel followed by a further run at 20 W per gel until the end. For comparative analysis, gels were stained with coomassie blue stain.

### ***12.3 In-gel tryptic digestion and mass spectrometry***

Protein spots were excised from 2-D gels with a spot picker and placed into 96-well microtiter plates, which were washed twice with TFA:acetonitrile:water (0.1:60:40). The tryptic digest was performed as reported previously with slight modifications [27]. The samples containing the tryptic-digested proteins were mixed at a 1:1 ratio with a solution of water:acetonitrile:TFA (67:33:0.1) saturated with  $\alpha$ -cyano-cinnamic acid. The mass spectrum was obtained on a Biflex III MALDI-TOF-MS (Bruker). The annotation of the peptide mass fingerprints was performed by the MASCOT search engine (Matrix Science). The search was done against our local *E.coli* database. The parameters used were, Taxonomy: All entries; Enzyme: Trypsin; Missed cleavages: 1; ppm: 100; Database: *E.coli*.

### ***12.4 Analysis of two-dimensional protein gels***

For a global proteome profile of both the strains and each evolutionary condition, spot densities were determined on three gel images from three independently grown cultures. For comparison of protein spot densities between different strains and evolutionary conditions, gels were scanned and digitized. Image smoothing, spot detection, spot quantification, image alignment, spot matching, spot annotation, molecular weight and pI calculation, and variation analysis of the protein gels was performed using PDQuest software (Bio-Rad). For each protein spot, the annotated information along with the peak area and normalized quantity values were obtained. Along with these exported annotations, the protein spots were analysed by grouping them into various functional categories based on MultiFun and Gene Ontology terms, the classification system for cellular functions of

gene products of *E. coli* [31, 84] consisting of 10 major functional categories. The  $\log_2$  ratios of the mean-centered normalized quantity values were used for unsupervised analysis (hierarchical clustering) using Cluster 3.0 and TreeView [32], where average-linkage clustering was performed using an uncentered correlation metric.

### **12.5 Network analysis**

All the networks reported in this study were constructed based on correlation data characterized by Pearson correlation coefficient (*PCC*)  $r$  value of 0.7 i.e. nodes which correspond to proteins with  $r \geq 0.7$  or  $r \leq -0.7$  were linked by an edge. All-against-all protein expression profile  $r$ -values of evolution-specific matrices were used to generate evolution-specific coexpression networks. Network visualization and analysis was performed using Cytoscape [86].

---

# CHAPTER 13

## RESULTS (III)

---

### 13.1 Proteome profiles of ancestral and evolved lines

The Adp line cultures (MGAdp and DHAdp) were maintained in prolonged exponential growth by daily passage into fresh medium for about 1000 generations (**Figure 13.1A**). The Stat line cultures (MGStat and DHStat) were maintained in constant batch culture

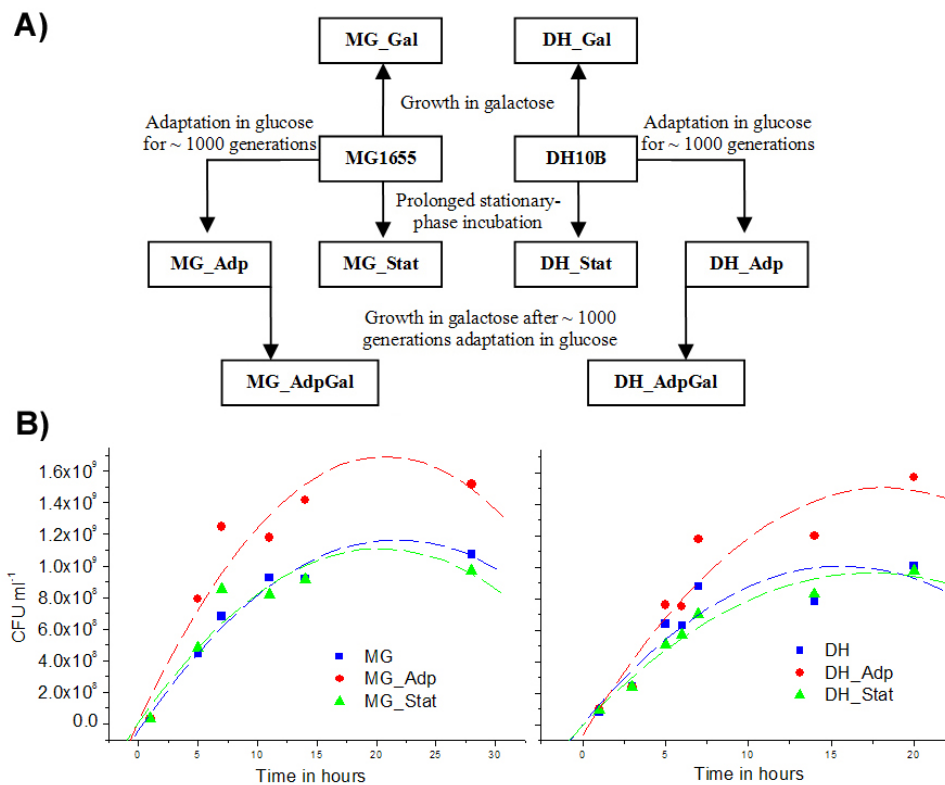


Fig. 13.1: **Flowchart of the experimental scheme.**

(A) The ancestral and evolved strains are shown in labelled rectangular boxes with experimental evolutionary conditions. The comparison between MG1655 or DH10B and MG\_Adp or DH\_Adp was carried out for excess nutrient adaptive evolution experiments. The comparison between MG1655 or DH10B and MG\_Stat or DH\_Stat was carried out for prolonged stationary phase evolution experiments. Similarly, the comparison between MG\_Gal or DH\_Gal and MG\_AdpGal or DH\_AdpGal was carried out for experiments examining the pleiotropic effects due to environmental shift. (B) The growth behaviour (polynomial fit) of the ancestral and evolved strains showing the increased growth fitness of the Adp lines in both the strains and identical growth behaviour of Stat lines when compared with the ancestral strains.

for 37 days where no nutrients were added after initial inoculation and no cells were re-

moved, unlike the former setup (**Figure 13.1A**). For the AdpGal line cultures (MGAdpGal and DHAdpGal), Adp lines (glucose adapted) were grown in medium containing galactose as the carbon source, thus creating an environmental shift for the cells with respect to the nutrient source (**Figure 13.1A**). During this period of adaptation, both Adp lines (evolved) showed increased fitness in their growth, whereas Stat lines (evolved) showed growth behaviour similar to their ancestors (**Figure 13.1B**). The samples of MG, DH, MGGal, DHGal, MGAdp, DHAdp, MGAdpGal, DHAdpGal, MGStat and DHStat lines grown in respective carbon sources were harvested in the mid-exponential phase of growth for proteome analysis, as done for both metabolome and transcriptome analyses. This allowed us to achieve excellent reproducibility by making gel parameters maximally consistent, and therefore enhancing the comparisons of gels. This also provided an opportunity to correlate proteome, transcriptome and metabolome data sets efficiently. Based on the results reported in this study, a proteomics database of *E. coli* (<http://2dbase.techfak.uni-bielefeld.de>) was developed which currently consists of 1185 protein spots information in which 723 proteins were identified and annotated from 12 gels. Among them, 10 gels were generated during microbial evolutionary experiments reported in this study. This database being a relational database, enables re-analysis, comparison, functional classification and extensive searching of protein spots in all the available proteome gels.

### 13.2 *Changes in transcript and metabolite levels*

The whole data sets of transcript and metabolite profiling along with the significant functional categories representation is summarised in **Figure 13.2**. Based on MultiFun and Gene Ontology terms (the classification system for cellular functions of gene products of *Escherichia coli*), in all comparisons, statistically significant functional categories with a  $P \leq 0.05$  (Wilcoxon rank sum test) that showed differences between the ancestral and the evolved strains fell into broad groups of genes involved in transport, biosynthesis and catabolism (**Figure 13.2A**). The expression changes associated with these main and broad functional categories consist of numerous sub-functional categories which varied in number between different evolutionary conditions (**Figure 13.2A**). Commonality observed in the analyses were the overrepresentation of transport and biosynthesis involved functional categories in all the evolutionary conditions studied (excess-nutrient adaptive evolution, prolonged stationary phase evolution and pleiotropic effects due to environmental shift). A similar statistical method for metabolite data sets revealed differences in metabolite levels involved in TCA cycle, nucleotide metabolism, aminoacids and their derivatives and polyamine biosynthesis under the experimental evolutionary conditions

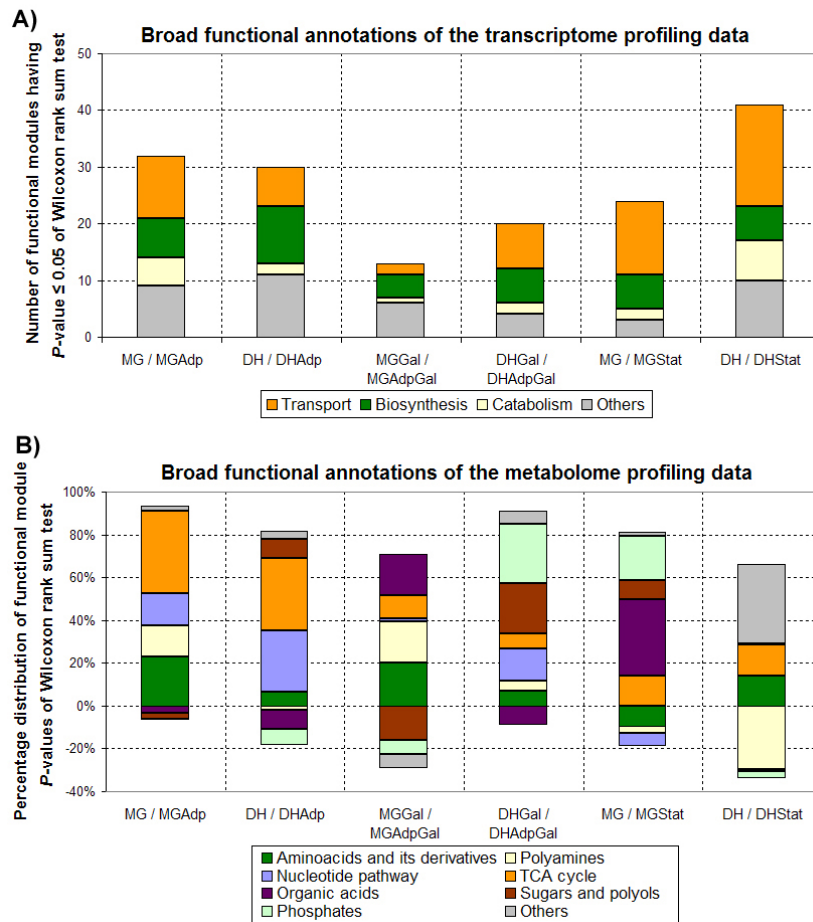


Fig. 13.2: **Broad functional annotations of the transcriptome and metabolome profiling data.**

(A) The stacked column plot comparison of the gene expression data based on broad functional classification. The comparison of the ancestor and evolved strains of individual evolutionary experimental conditions in both the strains showing the distribution of differentially regulated functional modules, having  $P \leq 0.05$  (Wilcoxon rank sum test). (B) The stacked column plot comparison of the metabolite profiling data based on broad functional classification. The comparison of the ancestor and evolved strains of individual evolutionary experimental conditions in both the strains showing the Wilcoxon rank sum test  $P$ -value degree distribution of various functional modules.

analysed. For example, metabolites involved in nucleotide pathway and TCA cycle were significantly over accumulated in the evolved strains adapted to the excess-nutrient condition when compared to the ancestral lines (**Figure 13.2B**). Likewise, significant changes in the levels of metabolites were seen in the evolved lines and their directional behavior is more or less constant in both the strains and in their evolved strains under all the evolutionary conditions studied. As a summary, the genes involved in the membrane associated functional categories and genes involved in the various biosyntheses functional modules were overrepresented in the experimental evolutionary studies. In accordance with these

findings, the metabolites involved in the TCA cycle and the nucleotide pathway were in higher levels in the evolved lines which were adapted to excess-nutrient source, where TCA cycle is the first step in generating precursors for various biosyntheses and one of the main energy producing pathways in a cell.

### *13.3 Changes in protein expression levels in experimental evolutionary conditions*

The normalized quantity values of all the protein spots were utilized to calculate the ratio of their standardized abundance in an evolved line relative to that in the ancestor, using the values obtained in the same data set. The calculated ratios of all the identified protein spots under all the experimental conditions were utilized for unsupervised hierarchical clustering (**Figure 13.4**). The hierarchical cluster shows the possible relevant comparisons among the evolved and the ancestral lines (**Figure 13.4**). In most cases the protein levels were significantly changed in evolved lines and their directional behaviour was more or less constant in both the ancestral and the evolved strains (**Figure 13.3**).

Among 495 and 470 protein spots' information present in the proteome gels of MG1655 and DH10B series of ancestor and evolved strains, 230 and 295 protein spots were identified, respectively. Among the identified proteins which were common between the strains, a comparison of Adp (evolved) and their ancestor lines showed 67% (MG/MGAdp) and 57% (DH/DHAdp) of proteins with significant changes (which corresponds to an estimated fold change of 1 or higher in protein expression) in their expression levels (**Figure 13.4**). Likewise, a comparison of the protein profiles of the MGGal and MGAdpGal showed 59% and DHGal and DHAdpGal showed 78% of proteins with significant changes in expression (**Figure 13.4**). Similarly, the MG/MGStat and DH/DHStat comparison showed a 67% and 48% of proteins with significant changes above the estimated fold change of 1 or higher in protein expression, respectively (**Figure 13.4**). The number of proteins showing differential expression above the estimated fold change of 1 or higher in protein expression varied between the different evolutionary conditions. When the proteome maps of the ancestor lines (MG and DH) were compared with the Adp lines' (MGAdp and DHAdp) proteome maps, the Adp lines shared about 34 proteins over- or under-expressed where, among them 47% (16 proteins) were in common and evolved in the same direction relative to both the ancestral expression levels. The comparison of the MGGal/MGAdpGal and DHGal/DHAdpGal lines showed 41 proteins up or down regulated, of which 34% (14 proteins) were in common. Likewise, the Stat lines (MGStat and DHStat) shared about 36 proteins with their ancestor lines (MG and DH), of which



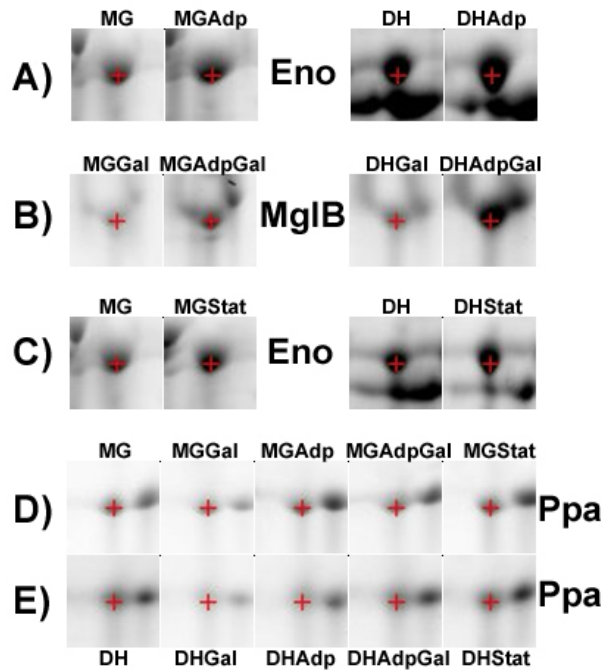


Fig. 13.3: **Typical examples of differential protein expression among the ancestral and evolved lines.**

(A) Sections of proteome gels showing the over expression of the enolase (Eno) protein levels in evolved strains when compared to their ancestors in excess-nutrient adaptive evolution. (B) Over expression of beta-methylgalactoside transporter (MglB) in AdpGal lines during adaptation due to environmental shift experiments. (C) Over expression of the enolase (Eno) protein levels in evolved strains when compared to their ancestors during prolonged stationary phase adaptive evolution. (D, E) Expression levels of inorganic pyrophosphatase (Ppa) which plays an important role in energy metabolism is shown as a control obtained from the same dataset of all the evolutionary experiments analysed.

only 25% (9 proteins) was in the same direction of expression pattern among both the strains.

In adaptation to excess-nutrient condition the evolved lines (MGAdp and DHAdp) exhibited down-regulation of 2 proteins (AceA and MglB) whereas 14 proteins were over-expressed when compared to their ancestors, among them 4 proteins were associated with energy metabolism (Eno, GpmA, Pkg, and Mdh), 4 were membrane associated proteins (MetQ, YbiL, HisJ, and YaeT), 2 were involved in amino acid metabolism (GlyA and ArgT), and the other 4 belonged to other functional categories or hypothetical proteins. In experiments concerning adaptation due to environmental shift, by altering the nutrient resource in the growth medium, the protein spots present in the AdpGal gels (MGAdpGal and DHAdpGal) displayed under-expression of 3 proteins (AceA, OmpF and Udp) and 11 proteins were over-expressed when compared to the proteome map prepared from the ancestral strain grown in galactose (MGGal and DHGal). Among them, 6 proteins were

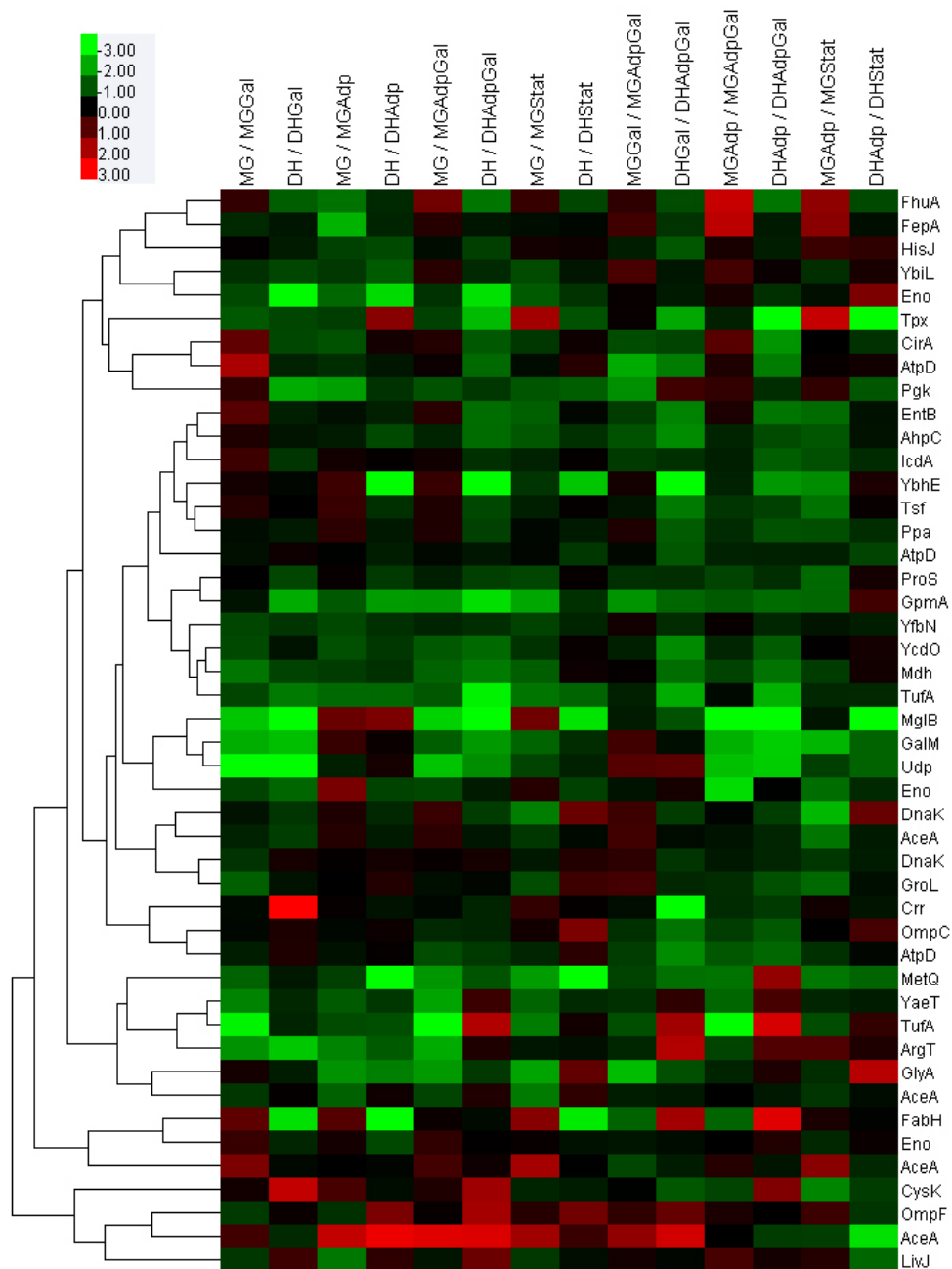


Fig. 13.4: Hierarchical clustering of protein expression levels of ancestral and evolved strains.

Logarithmically transformed (to base 2) response ratios of each comparison are arranged in columns and rows corresponding to single proteins. Dendrograms indicate the degree of similarity between clusters of proteins. Red indicates up-regulation and green indicates down-regulation in protein expression, according to the  $\log_2$  ratio scale on the upper left.

involved in membrane-associated functions (EntB, AtpD, AtpD- isoform, OmpC, MetQ and CirA), 2 proteins belonged to energy metabolism of the cell (GpmA and IcdA), and the remaining 3 proteins corresponded to other functional categories. By adaptation to the prolonged starvation conditions, the evolved strains (MGStat and DHStat) exhibited over-expression of 8 proteins when compared to their ancestor strains (MG and DH). Out of these over-expressed proteins, 4 (Pgc, Eno, GpmA and YbhE) were related to the energy metabolism functional category of the cell, and the remaining 4 (MetQ, GalM, TufA and AhpC) were involved in membrane-associated and other functional categories. The expression patterns of these 39 proteins under the three experimental evolutionary conditions were altered in the same direction in both the strains demonstrating evolutionary parallelism, indicating the effects of adaptive evolution in excess-nutrient, environmental shift and incubation in prolonged stationary phase evolutionary environment.

### *13.4 Direction of the observed protein expression changes*

Observing the high extent of parallel changes in the protein expression levels among both the strains under all experimental evolutionary conditions, we examined the degree and the direction of the observed changes of their expression levels. For examining the level of observed changes among the strains in all the three experimental evolutionary conditions, we calculated the pairwise Pearson correlation coefficient ( $r$ ) for all the normalized spot intensities. All the proteins having a threshold value of  $r \leq -0.7$  or  $\geq 0.7$  were plotted on both axes of a matrix containing all pairwise protein expression profile correlations. When these correlations ( $r$ ) are colour coded, it facilitates a visual inspection to determine the degree of correlation among the samples in question. The correlation map of Adp, AdpGal and Stat lines shows various degrees of negative correlation (**Figure 13.5**). Among them, Stat lines (MG/MGStat versus DH/DHStat) displayed a high degree of negative correlation when compared to AdpGal and Adp lines in the protein correlation maps suggesting elevated level of protein expression level variability in the Stat lines (**Figure 13.5C**). The correlation map of Adp lines (MG/MGAdp versus DH/DHAdp) exhibited a lower degree of negative correlation than the other lines, indicating reduced level of stochastic variability in the protein expression levels of Adp lines (**Figure 13.5A**).

### *13.5 Protein correlation network analysis*

Normalized spot intensities of all the proteins were utilized to form distance matrices which were calculated by using the Pearson correlation coefficient  $r$  to build protein coexpression networks for all the experimental evolutionary conditions. It has been demonstrated pre-

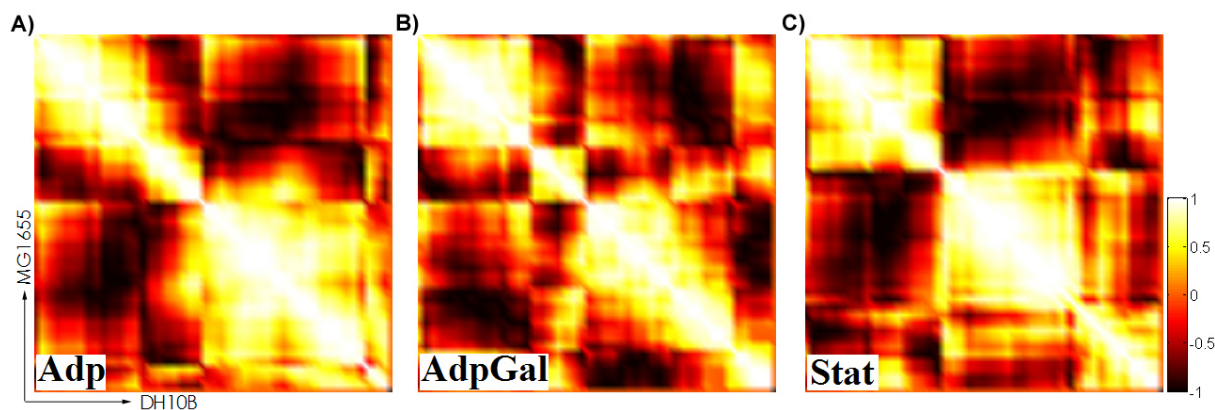


Fig. 13.5: **Comparison of direction of changes in the experimental evolution among the strains.**

Pairwise correlation maps of the proteome data among the strains in all the experimental evolutionary conditions, using the Pearson correlations coefficient ( $r$ ). All the proteins having a threshold value of  $r \leq -0.7$  or  $\geq 0.7$  were plotted colour-coded on both axes of a matrix containing all pairwise protein expression profile correlations. Darker spot indicates higher degree of negative correlation among the strains. The analysis was carried out using Matlab 6.5 (The MathWorks, Inc.).

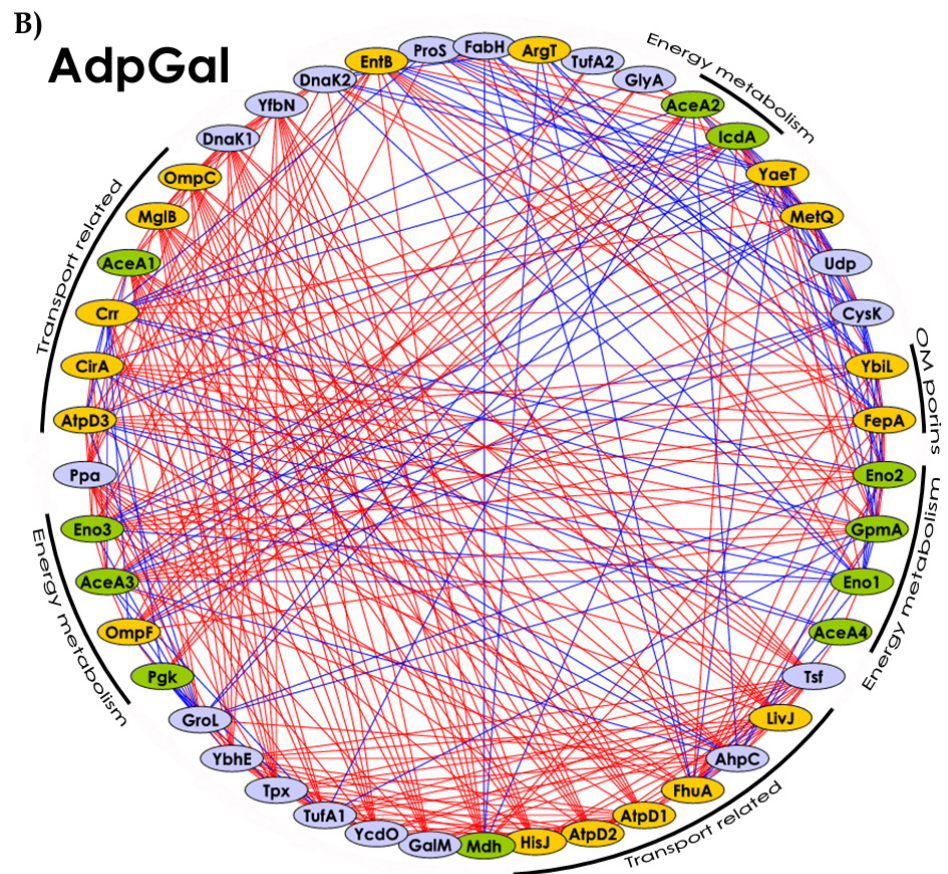
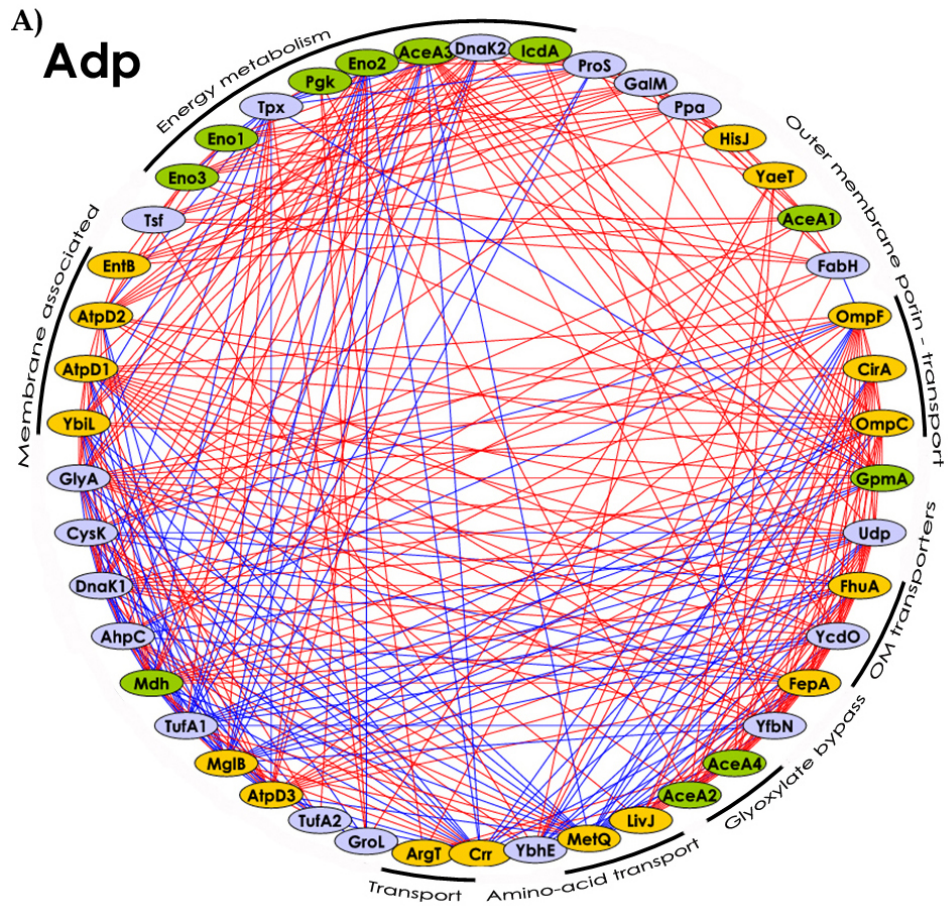
viously that functionally related genes are preferentially linked in coexpression networks [90]. By applying a similar method for building coexpression protein networks from protein expression data, we were able to explore and reproduce the connections between the protein-to-protein links and their corresponding mutual functionality. Functional justification of the protein coexpression network was based on the assumption that the more similar the expression pattern is, the shorter the distance between proteins in the coexpression network. Hence, the closely associated proteins in the coexpression networks will be of special interest in this analysis.

All-against-all protein profile comparison for Adp, AdpGal and Stat matrices were used to generate evolution specific coexpression networks constructed using Pearson correlation coefficient ( $PCC$ ). There was a strong dependence between coexpression and functional relevance of the networks, demonstrating the strong potential of coexpression network analyses (**Figure 13.6**).

In coexpression networks, nodes correspond to proteins, and edges link two proteins if they have a threshold positive or negative correlation coefficient of  $r \leq -0.7$  or  $\geq 0.7$  which are considered to be expressed differentially. The coexpression network analysis provides a possibility to use it as a quantifiable and analytical tool to unravel the relationships among cellular entities that govern the cellular functions [91].

By analysing the protein coexpression networks separately, we found that the details of the connections in the networks were different. Very few coregulated protein pairs are shared among the evolution specific networks. For instance, MetQ and LivJ, involved







c) **Stat**

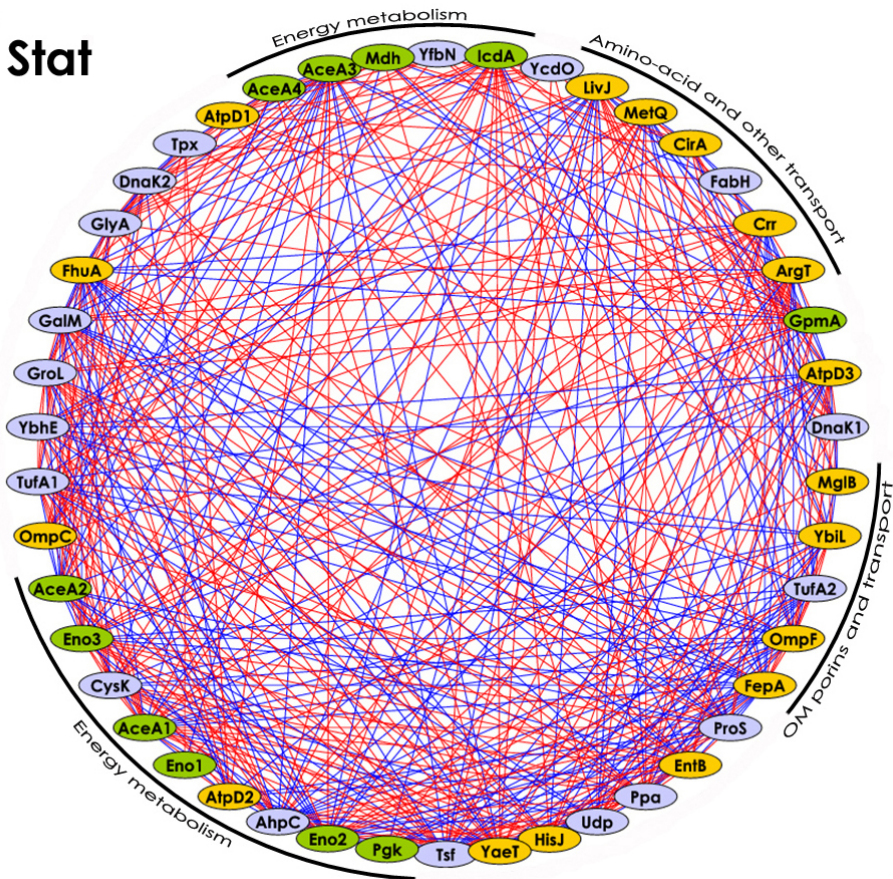


Fig. 13.6: **Protein correlation network analyses.**

Evolution specific coexpression networks were constructed using Pearson correlation coefficient by all-against-all protein profile comparison. Nodes correspond to proteins, and edges link two proteins if they have a threshold correlation coefficient of  $r \leq -0.7$  (blue edge) or  $\geq 0.7$  (red edge). Membrane-associated proteins (nodes) are colour coded with orange and nodes related to energy metabolism are in green. Closely linked proteins in the network with associated function are labelled.

in transport of amino acids, shared a consistent connection in Adp and Stat evolution specific networks, but not in AdpGal coexpression network (**Figure 13.6**). Likewise, in all the evolution specific coexpression networks, Eno and AceA, involved in the energy metabolism of the cell, shared a consistent association (**Figure 13.6**). Major functional categories which were overrepresented in these networks were energy metabolism, transporters, membrane associated proteins and outer-membrane proteins (**Figure 13.6**). Despite the presence of very few commonly coregulating protein pairs in these coexpression networks, the functionally associated protein pairs were consistently connected in all the networks. This suggests that the different experimental evolutionary conditions have changed not only the protein expression levels, but also the coregulatory relationships among the protein pairs. Nevertheless, the global functional category association is

retained in all the evolutionary networks analysed (**Figure 13.6**).

### *13.6 Comparison of transcript, metabolite and protein levels*

The adaptation of both strains to the experimental evolutionary conditions presented in this study affected the expression of those proteins which were associated principally with the membrane related transport functional category and the energy metabolism of the cell. Apart from the over-expressed proteins in the energy metabolism functional category in Adp lines (MGAdp and DHAdp) (**Figure 13.7**), the proteins involved in the amino acid metabolism and transport were differentially over-expressed. For instance, GlyA and ArgT which are involved in amino acid biosynthesis were over-expressed in both the evolved strains adapted to excess-nutrient conditions. GlyA is involved in the biosynthesis of glycine, purines, and methionine. In line with the over-expression of GlyA protein (MG/MGAdp= 3.46 fold; DH/DHAdp= 3.01 fold) in the evolved strains, the levels of glycine, adenine (purine) and methionine metabolites were in higher amount in both the strains evolved in excess-nutrient condition (**Figure 13.7**). ArgT is involved in biosynthesis of arginine and functions as lysine/arginine/ornithine ABC transporter. In parallel with the upregulation of ArgT (MG/MGAdp= 3.11 fold; DH/DHAdp= 2.15 fold) in both the excess-nutrient evolved strains, the levels of lysine, arginine and ornithine metabolites were higher. In correlation with this, the polyamines (4-aminobutyrate, 5-methyl-thioadenosine, putrescine and spermidine), the derivatives of arginine were present in higher amount in both the evolved strains (**Figure 13.7**). Similarly, the proteins involved in the transport of amino acids (MetQ and HisJ) were over-expressed in both the excess-nutrient evolved strains. MetQ which is involved in the uptake of methionine was over-expressed 1.5 fold in comparison of MG/MGAdp and 6.43 fold in comparison of DH/DHAdp. The methionine levels were also high in both lines. HisJ is a member of the histidine transport system and it was 1.54 (MG/MGAdp) and 1.77 (DH/DHAdp) fold over-expressed in both the excess-nutrient evolved strains, respectively. In case of the gene expression data, all the genes involved in the histidine biosynthesis were up-regulated in both the evolved strains which were adapted in excess-nutrient evolutionary condition.

In AdpGal lines (MGAdpGal and DHAdpGal), among the few proteins which were significantly differentially regulated, three outer-membrane porin proteins (OmpC, OmpF and CirA) were differentially expressed in both the strains during the environmental shift experiments. OmpC porin which transports ions and other hydrophilic solutes across the outer-membrane [110] was 1.13 fold (MGGal/MGAdpGal) and 2.74 fold (DHGal/DHAdpGal) over-expressed, during the environmental shift experiments. OmpF porin which is involved in the transport of solutes such as sugars, ions, and amino acids

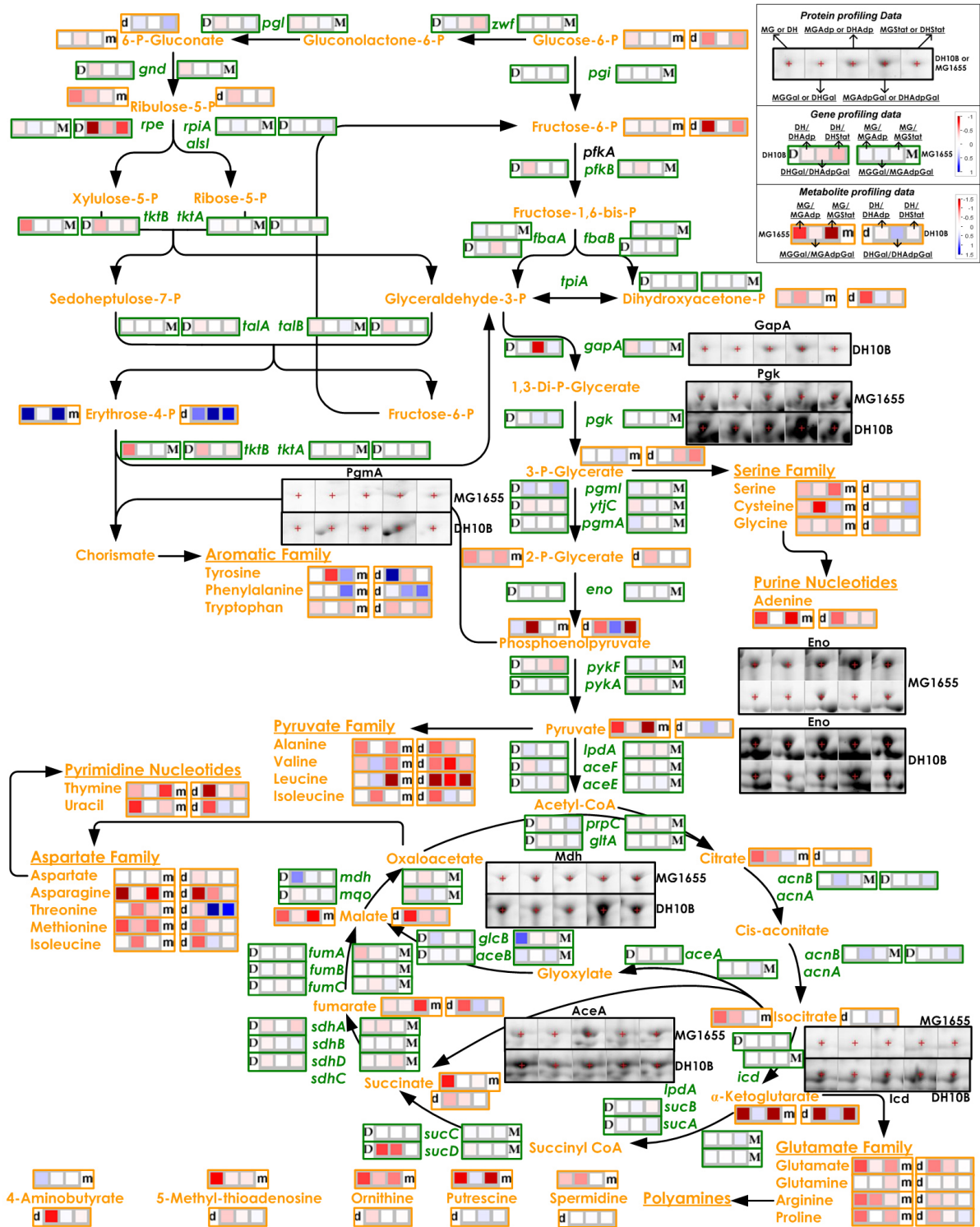


Fig. 13.7: Protein, gene and metabolite levels in the central metabolic routes and the diversion of key intermediates to biosynthetic pathways.

Genes are represented in green text, metabolites in orange text and proteins in black text. Identified protein expression levels in all the experimental conditions are represented as raw protein spots excised from the gel images. Ancestor and evolved strain specific gene expression comparisons are denoted in green boxes (M- MG1655, D-DH10B). Ancestor and evolved strain specific metabolite abundance comparisons are denoted in orange boxes (m- MG1655, d-DH10B). Logarithmically transformed (to base 2) response ratios were utilized for each comparison according to the  $\log_2$  ratio scale on the upper right inset.



[111] was downregulated to a fold range of 1.12 and 2.37 in both the strains. CirA outer-membrane porin was 1.80 (MGGal/MGAdpGal) and 1.63 (DHGal/DHAdpGal) fold over-expressed, which is a TonB dependent iron-siderophore complex uptake receptor.

In Stat lines (MGStat and DHStat), among the proteins which were differentially regulated in both the evolved strains in prolonged stationary phase incubation two membrane bound proteins MetQ (MG/MGStat=3.69, DH/DHStat= 7.77) and GalM (MG/MGStat= 2.37, DH/DHStat= 1.06) were over-expressed. Other major differentially expressed proteins belonged to the energy metabolism of the cell (**Figure 13.7**). In all the evolutionary experimental processes reported in this study, the proteins involved in energy metabolism were significantly differentially regulated. By integration of the metabolite and gene expression levels along with the proteome data generated in this study, the correlation of the observed results were in complete agreement with each other (**Figure 13.7**).

### *13.7 Comparison of parallelism in the three functional levels*

Parallel evolution is the independent progression of similar traits in separate but evolutionarily related lineages through similar selective factors on both lines by adaptation [92]. Here we examined the parallelism among different lines by examining the direction of changes in the protein, gene and metabolite levels adapted to three evolutionary experimental conditions. For examining the parallelism by adaptation in the trend of protein, transcript and metabolite levels under all the evolutionary conditions, the matrix of both the strains having all the normalized values of the protein and gene expression levels and metabolite abundance levels were plotted on both axes. The linear relationship in all the functional level data sets and in all the experimental conditions was measured by means of correlation coefficient between both strains. In all the three functional levels of data sets, Adp lines displayed high linear relationship between both the strains by having a higher correlation coefficient when compared to the other evolutionary conditions (**Figure 13.8**). This indicates that among the evolutionary conditions examined in this study, adaptation to excess-nutrient evolutionary condition exhibits high parallelism in the gene, protein and metabolite levels. AdpGal and Stat lines had slightly lower correlation coefficient than the Adp lines demonstrating variability in the gene, protein and metabolite levels and lower linear relationship between the strains in environmental shift and prolonged stationary phase induced evolutionary conditions (**Figure 13.8**).

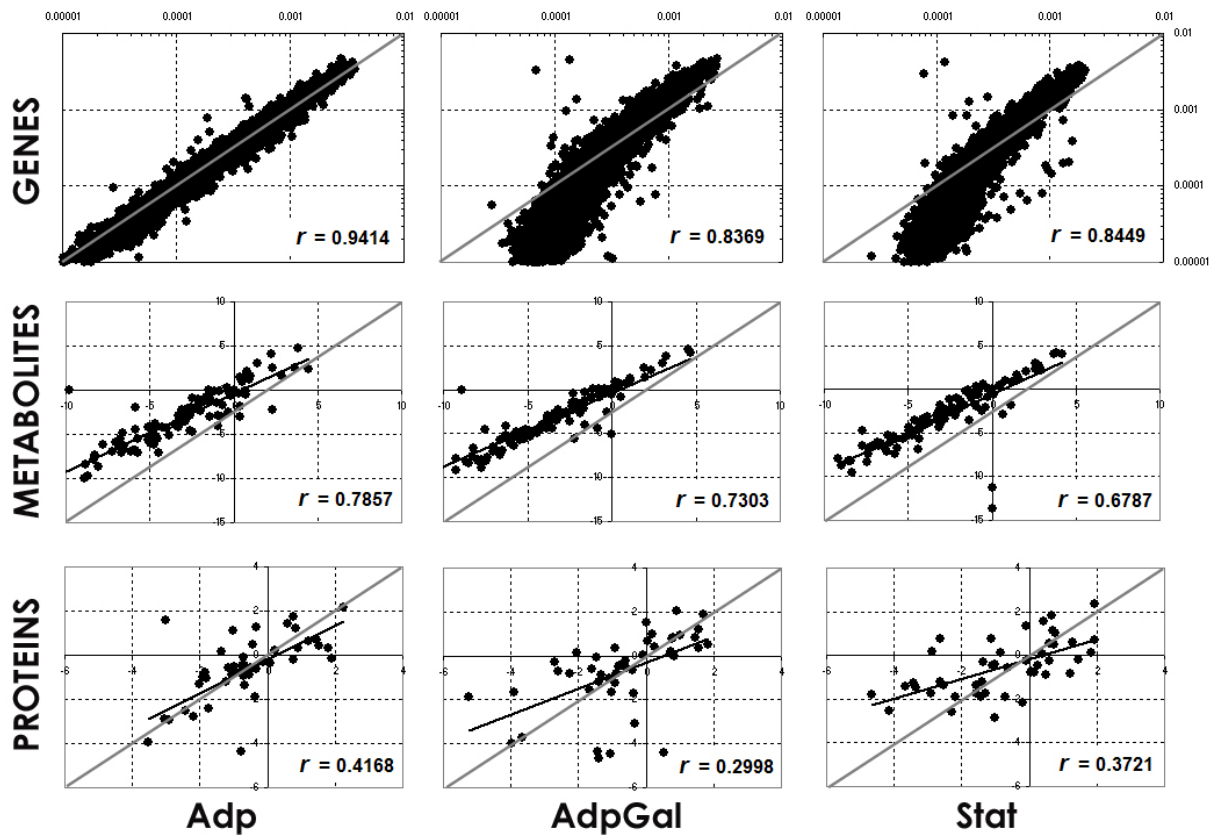


Fig. 13.8: **Evolutionary trends of protein, transcript and metabolite levels.** For all the evolutionary conditions of two strains, the matrix of both the strains having all the normalized values of protein and gene expression and metabolite abundance levels were plotted on both x (MG1655) and y (DH10B) axes. The linear relationship between both the strains in every comparison is denoted by the correlation coefficient  $r$ .

---

## CHAPTER 14 DISCUSSION (III)

---

We have studied the proteomic changes of two strains harbouring different genotypes adapted to three different experimental evolutionary conditions, namely, excess-nutrient adaptive evolution, prolonged starvation adaptive evolution and examining pleiotropic effects due to an environmental shift by nutrient alteration. Along with the proteome data generated in this study, we integrated the gene expression and metabolite profiling data for a better understanding of the experimental evolutionary conditions. We found that among the proteins identified in all the evolutionary conditions studied, 48% to 78% of proteins were differentially regulated when compared to the proteins in their respective ancestors (MG/MGAdp= 67%; DH/DHAdp= 57%; MGGal/MGAdpGal= 59%; DHGal/DHAdpGal= 78%; MG/MGStat= 67%; DH/DHStat= 48%). By identifying the proteins which are common in a particular evolutionary condition between both the strains which are differentially regulated, we could assume the involvement of these intersecting proteins in the experimental evolutionary condition studied. Obtained proteomics data integrated along with gene expression and metabolite levels would certainly ensure the involvement of these proteins in that particular evolutionary condition. In excess-nutrient adaptive evolutionary conditions of both the strains, 47% of differential regulated proteins were common in both the strains. In environmental shift by nutrient alteration condition on both the strains, 34% of differential regulated proteins were common in both the strains. In prolonged starvation adaptive evolutionary condition of both the strains, 25% of differential regulated proteins were intersecting in both the strains. As a result, the protein expression patterns in excess-nutrient adaptive evolution apparently have the highest level of parallelism in the expression patterns between the strains, whilst protein expression patterns in prolonged starvation adaptive evolution have the lowest. A similar conclusion was also derived by pairwise protein expression profile correlation analysis which revealed a higher degree of negative correlation in the prolonged starvation adaptive evolution condition than the other evolutionary conditions studied (**Figure 13.5**). This parallelism was not only seen in the proteomics data, but also in transcriptomics and metabolomics data sets (**Figure 13.8**). Microorganisms in the stationary phase are known to have varied mutation rates which are known to be influenced by the genetic background of the strain [78]. The condition of initial isogenic long-term stationary-phase cultures is known to be highly dynamic in nature and is known to produce diverse mutations due to significant genotypic diversity in these cultures [70]. In complete agreement

with this hypothesis, our results demonstrated the lowest level of parallelism seen in the protein, gene and metabolite levels between the two strains evolved in the prolonged starvation adaptive evolutionary condition when compared to other conditions reported in this study (**Figure 13.4, 13.5, 13.8**).

The functional categories which were differentially regulated among all the evolutionary conditions in their proteome level belonged to membrane-bound functional categories and the energy metabolism of the cell. Similar results were obtained when the gene expression of these strains under all the evolutionary conditions were analysed, where the membrane-bound Gene ontology functional modules were significantly overrepresented. In this study, both the evolved strains adapted to excess-nutrient evolutionary conditions exhibited higher growth fitness than their ancestor strains. One of the possible reasons for increased growth fitness can be assumed to be due to the higher levels of the proteins, genes and metabolites and their intermediates involved in the energy metabolism in the evolved strains (**Figure 13.7**). The energy metabolism of the cell consists of pathways which are the first step in generating precursor metabolites for biosyntheses involved in various vital cellular processes. Altered levels of proteins, genes and metabolites involved in central metabolism in evolved strains of excess-nutrient and starvation conditions reported in this study are consistent with a previous study related to variation of central metabolism gene expression levels in glucose-limited adaptive evolution in yeast [5].

Previous studies of the evolved strains harbouring mutations for increased growth fitness under limited-nutrient conditions have shown that the strains re-structured their membrane for increased permeability [96, 97, 98]. In long-term stationary-phase cultures, for cellular maintenance the nutrient-scavenging process on the cellular debris enhances the availability of carbon sources by restructuring the outer-membrane composition and by improving its permeability [100]. The outer-membrane of the cell being the initial point of contact with the external environment, its corresponding cellular components are susceptible to changes to adjust to the external diverse environment. In complete agreement with this hypothesis, the membrane bound functional categories were significantly overrepresented in the evolved strains in their gene expression levels (**Figure 13.2**). Consistent with this, the proteomic results presented reveal that the outer-membrane porin proteins (OmpC, OmpF, CirA and YbiL) were differentially regulated under the evolutionary conditions reported (**Figure 13.4, 13.6** and Supplementary Material; **Table 19.8**).

---

# CHAPTER 15

## CONCLUSION (III)

---

We report the evolutionary changes in protein expression profiles in two strains of three different experimental evolutionary conditions along with the integrated metabolite and gene expression data. As a result of the analysis of the evolved and ancestor strains in three functional levels, we could confirm the involvement of the membrane related functional categories and energy metabolism of the cell in all the evolutionary conditions studied. The membrane-bound components are the first point of contact with the evolved external environment and thus their involvement in the evolutionary processes examined in this study is understandable. The pathways of energy metabolism of the cell are known to generate energy under various conditions of growth for all cellular biosyntheses and produce various intermediates which are the precursor metabolites of several biosynthetic pathways. The evolved strains which were adapted in different environmental conditions exhibited growth in stressful conditions or increased growth fitness; these evolved strains would certainly require active involvement of the functional category energy metabolism of the cell for their cellular metabolism. Apart from the already known mutants which are known to play a vital role in the evolved conditions [69, 70], further experiments will be required for screening the mutations in these evolved strains in the membrane-related and energy metabolism functional categories or regulators affecting these functional modules' candidates. Analyses of the established isogenic constructs of these mutants along with combined changes in protein, transcript and metabolite levels would certainly demonstrate the involvement of the mutant in the evolutionary processes, thus enhancing the understanding of the same. Even though comprehensive analyses at all three levels provide no direct answer for the evolutionary processes analysed, analysis of the responses of individual parameters at all three levels would certainly provide valuable functional information about the evolutionary processes.

# Part IV

2DBase: 2D-PAGE Database of *Escherichia coli*.

## Abstract

We present a web-based integrated proteome database, termed 2DBase of *Escherichia coli* which was designed to store, compare, analyse and retrieve various information obtained by 2D polyacrylamide gel electrophoresis (2D PAGE) and mass spectrometry (MS). The main objectives of this database are (1) to provide the features for query and data-mining applications to access the stored proteomics data (2) to efficiently compare the specific protein spots present in the comparable proteome maps and (3) to analyse the data with the integrated classification for cellular functions of gene products of *E. coli*. This database currently contains 12 gels consisting of 1185 protein spots information in which 723 proteins were identified and annotated. Individual protein spots in the existing gels can be displayed, queried, analysed and compared in a tabular format based on various functional categories enabling quick and subsequent analyses. Our database satisfies the requirement to be a federated 2-DE database by accomplishing various tasks through a web interface providing access to a relational database system. The 2DBase of *Escherichia coli* database can be accessed at <http://2dbase.techfak.uni-bielefeld.de/>

**Authors:** Chandran Vijayendran, Sebastian Burgemeister, Karl Friehs, Karsten Niehaus, and Erwin Flaschel

**Author's contribution:** CV conducted all the experiments cited in this study, analysed the results and wrote this report. SB was involved in the database construction and maintenance as a thesis to fulfill his masters. KF and KN was involved in experimental guidance. EF is the scientist in whose laboratory all the experiments were conducted and was involved in experimental design.

# CHAPTER 16

## INTRODUCTION (IV)

Ever since the term "proteome" was introduced [112], conventional 2-D gel electrophoresis has remained the major method for proteome analysis [106, 107]. High throughput proteomic data generated by 2-D gel experiments require elaborate data handling to ensure comprehensive analyses. Increasing amount of data increases the complexity of comparing maps present in any existing database. Several 2-D gel electrophoresis databases have been published in recent times which contain large amounts of experimental proteomics data generated by various high-throughput methodologies (<http://expasy.org/world-2dpage/>). With the rapid increase in the raw proteomics data within or between laboratories, it has become increasingly challenging to meaningfully compare the results from such large datasets containing numerous 2-D maps. Database management systems, along with proficient methods of map comparison, would enhance the analyses.

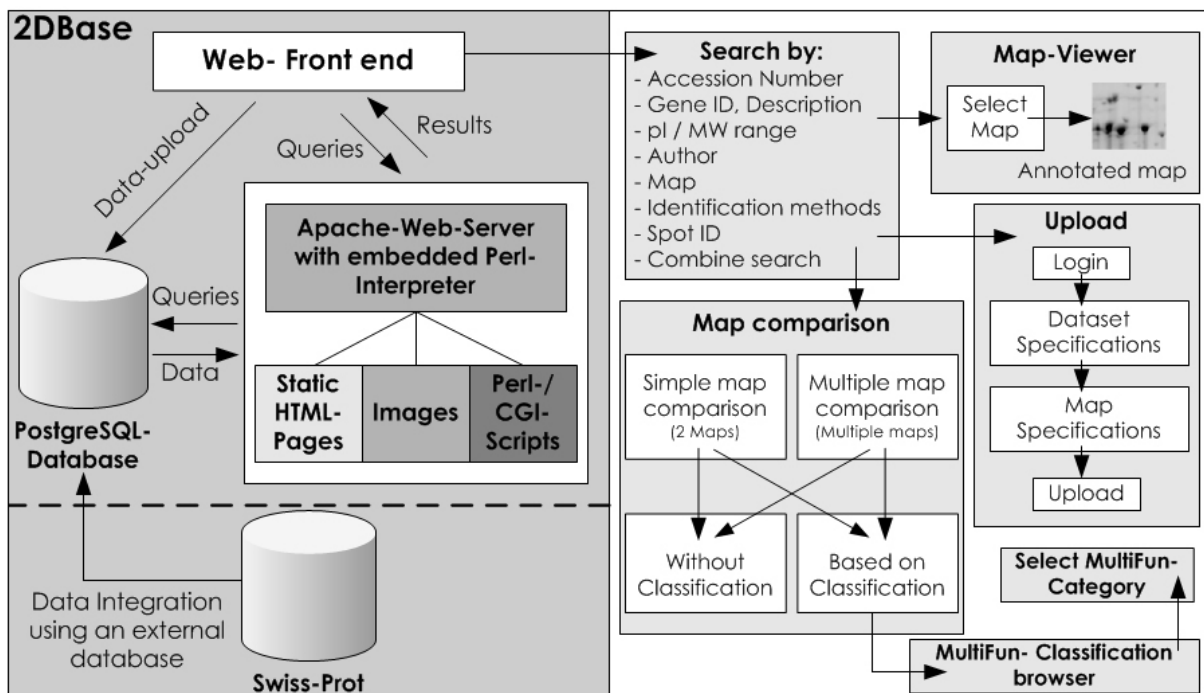


Fig. 16.1: Data flow diagram of 2DBase of *Escherichia coli*.

The database is a relational database system, showing its extensive search, comparison and classification options.

Here we report a proteomics database of *Escherichia coli* which currently consists of 1185 protein spots information in which 723 proteins were identified and annotated from



12 gels. Among them, 10 gels were generated during microbial evolutionary experiments (unpublished) and the remaining two gels are discussed later on in this report. The database is a relational database system (**Figure 16.1**) and supports extensive search functions according to several fields (accession number, gene id, description, author, spot id, pI/MW) (**Figure 16.1**). We have applied an extensive, quick, efficient and easy approach to compare various gels by classifying each protein spot utilizing a previously published classification system for cellular functions of gene products of *Escherichia coli* [31, 84]. We used a scoring function generated from the peak value and normalized quantity of the protein spots, and this information was utilized in a sortable table format. We also made a functional classification of the protein spots. All of this enabled us to quickly analyse and compare the gels in an efficient manner. To our knowledge, there have been no previous reports of a proteome database which has the option of analysing and comparing 2-D gels at a single protein comparison level across all the gels aided by functional category classification.

---

## CHAPTER 17

### DESCRIPTION (IV)

---

The database was designed to systematically input, store, compare, analyse and output all the information related to an experiment. Proteome maps generated from 2-D gel electrophoreses were scanned and the protein spots were subjected to tryptic digestion and identified by comparing the peptide masses, which were obtained on a Biflex III MALDI-TOF-MS (Bruker). The digitized gel images were normalized, and annotated using the PDQUEST program (version 6.2; Bio-Rad). For each protein spot, the annotated information along with the obtained X and Y coordinates, the peak area and normalized quantity values were stored in the database. Internally these data, along with the gel images (**Figure 17.1**) were stored by an upload function option in the database.

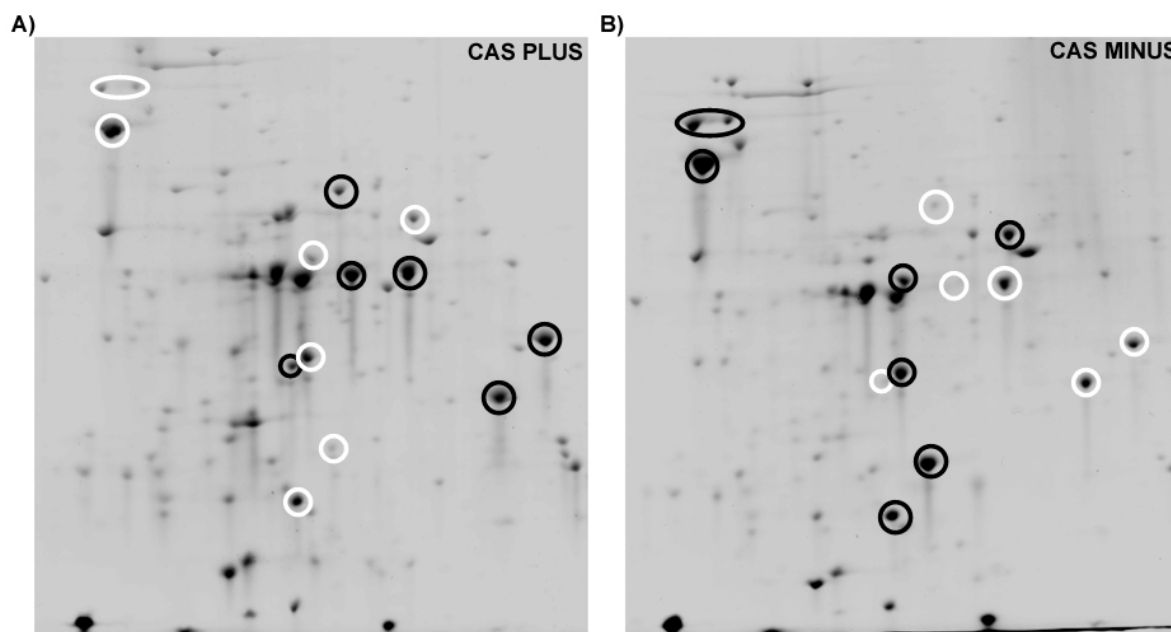


Fig. 17.1: **CAS-PLUS and CAS-MINUS 2Dgels.**

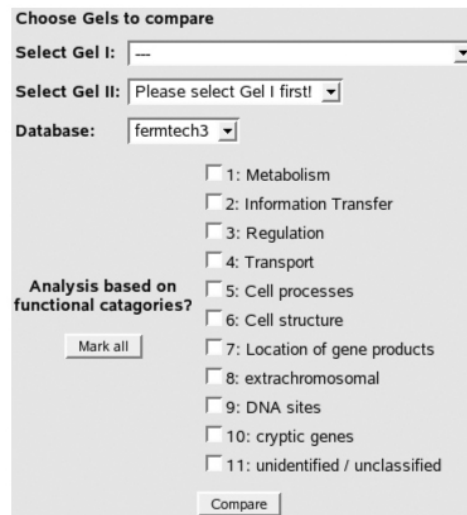
The proteome gels obtained from a strain (MG1655) which was grown in the presence (CAS-PLUS) and absence (CAS-MINUS) of Casamino-acids respectively. Encircled black and white symbols represent protein over- and under-expression in the corresponding gel, respectively.

This database was created using the Make2D-DB II Package/ version: 2.00.1 (<http://expasy.org/ch2d/make2ddb/>) [113] with additional characteristics based on our modifications. The identified protein spots with the SWISS-PROT accession number were stored

in a relational database that was made accessible online via a common gateway interface (cgi) script on a linux web server (**Figure 16.1**). Information pertaining to each protein spot can be accessed via a clickable gel image (**Figure 16.1**).

Expression levels of proteins can be compared one-by-one by viewing the corresponding protein spot image in a tabular format from the selected gels based on a scoring function, calculated based on the peak and normalized quantity values. For each identified protein spot, the spot ID, SWISSPROT accession number, B number, name, function, peak value, normalized quantity, protein spot image, and score are available in a tabular format. Moreover, additional information can be accessed via links to the SWISS-PROT database.

For further analysis of the proteome data, we utilized MultiFun- the classification system for cellular functions of gene products of *E. coli* [31, 84] consisting of 10 major functional categories (**Figure 17.2**). These major categories are further sub-divided into



The image shows a web form titled "Choose Gels to compare". It contains the following elements:

- "Select Gel I:" dropdown menu with "--" selected.
- "Select Gel II:" dropdown menu with "Please select Gel I first!" selected.
- "Database:" dropdown menu with "fermtech3" selected.
- A section titled "Analysis based on functional categories?" with a "Mark all" button.
- A list of 11 functional categories, each with an unchecked checkbox:
  - 1: Metabolism
  - 2: Information Transfer
  - 3: Regulation
  - 4: Transport
  - 5: Cell processes
  - 6: Cell structure
  - 7: Location of gene products
  - 8: extrachromosomal
  - 9: DNA sites
  - 10: cryptic genes
  - 11: unidentified / unclassified
- A "Compare" button at the bottom.

Fig. 17.2: MultiFun- the classification system for cellular functions of gene products of *Escherichia coli*.

This classification system consist of 10 major functional categories. These major categories are further sub-divided into a hierarchical scheme.

a hierarchical scheme. The complete hierarchical structure of the MultiFun classification can be screened for the expression of various proteins involved in a particular functional category in question from the selected gels. As a result, all the individual protein spots are summarized in a table consisting of vital information with the spot image along with the classification based on functional category which enhances the analysis.

# CHAPTER 18

## IMPLEMENTATION & CONCLUSION (IV)

To demonstrate the performance and the utility of the method used for comparing the proteome maps in this database, we applied this method to the proteome gels derived from a strain (MG1655) which was grown in the presence and absence of Casamino-acids respectively (**Figure 17.1A, B**). For this experiment approximately  $4.1 \times 10^{10}$  cells were harvested from M9 minimal medium culture supplemented with or without Casamino-acids from the exponential phase of growth. A protein sample of 300 g was loaded on an Immobiline dry strip (Amersham Biosciences) with a length of 24 cm, pI-range 4-7. The first dimension was developed on an electrophoresis apparatus IPG-phor (Amersham Biosciences) until 75,000 Vh was reached. The second dimension was developed by means of a 12.5% poly-acrylamide gel. For comparative analysis the gels were stained with coomassie blue. The excised protein spots were tryptic digested and the mass spectra were obtained on a Biflex III MALDI-TOF-MS (Bruker). Annotations of the peptide mass fingerprints were carried out by the MASCOT search engine (Matrix Science). The parameters used were: Taxonomy: all entries; Enzyme: trypsin; Missed cleavages: 1; p.p.m.: 100; Database: *E. coli*. 110 protein spots were analysed in each gel, as a result 99 proteins could be annotated in both the gels. Many proteins were differentially expressed (**Figure 17.1A, B**). Comprehensive screening of the functional category metabolism (1- MultiFun class) revealed substantial over-expression of the tryptophan amino-acid biosynthesis (1.5.1.15- MultiFun class) enzymes in the CAS-MINUS proteome map (**Figure 18.1**). Similarly, browsing the table of transport functional category (4-

1.5.1.15 Tryptophan (4)

Spot	Swissprot ID	B-Number	Gene	Function	Peak Value	Norm. Quantity	Gel 1	Peak Value	Norm. Quantity	Gel 2	Score <sup>ii</sup>
6201	P0A879	b1261	trpB	tryptophan synthase, beta subunit	1.647	61093.8		0.121	3647.65		58
8303	P0A879	b1261	trpB	tryptophan synthase, beta subunit	1.263	38329.6		0.007	95.8		42
4003	P0A8G6	b1004	wrbA	predicted flavoprotein in Trp regulation	0.079	2233.05		0.406	12723.4		12
6101	P0A877	b1260	trpA	tryptophan synthase, alpha subunit	0.12	15474.8		0.016	1056.01		10

**Fig. 18.1: MultiFun- metabolism- functional category.**

Comprehensive screening of the functional category metabolism (1- MultiFun class) showing substantial over-expression of the tryptophan amino-acid biosynthesis (1.5.1.15- MultiFun class) enzymes in the CAS-MINUS (Gel 1) proteome map.

MultiFun class) revealed that proteins involved in the category of amino-acid transport were over-expressed in CAS-PLUS proteome map (**Figure 18.2**). Micro-organisms are

MG1655_CAS_MINUS : MG1655_CAS_PLUS												
4: Transport												
(Browse functional category Transport based on MultiFun classification)												
Spot	Swissprot ID	B-Number	Gene	Function	Peak Value	Norm. Quantity	Gel 1	Peak Value	Norm. Quantity	Gel 2	Score	†
6603	P0AD96	b3460	livJ	leucine/isoleucine/valine transporter subunit -/- periplasmic-binding component of ABC superfamily	0.013	1089.63	+	1.231	31312.3		36	
6601	P23843	b1243	oppA	oligopeptide transporter subunit -/- periplasmic-binding component of ABC superfamily	0.707	39313		0.257	8765.98		28	
8309	P39172	b1857	znuA	zinc transporter subunit -/- periplasmic-binding component of ABC superfamily	0	3.72	+	0.692	23989.5		25	
5102	P0AEM9	b1920	fliY	cystine transporter subunit -/- periplasmic-binding component of ABC superfamily	1.136	37768.4		0.797	32574.4		17	
1	P69783	b2417	crr	glucose-specific enzyme IIA component of PTS	1.373	38663.2		1.256	33229.3		16	

Fig. 18.2: **MultiFun- transport- functional category.**

The transport functional category (4- MultiFun class) showing the proteins involved in the amino-acid transport being over-expressed in CAS-PLUS (Gel 2) proteome map

known to utilize the available nutrient resources present in the environment rather than to synthesise the nutrients by themselves. Amino-acids present in the medium have to be transported into the cytoplasm for utilization. In line with these known phenomena, the proteins involved in the transport function were over-expressed in the sample grown in the presence of Casamino-acids (**Figure 18.2**) and on the other hand the protein involved in the amino-acid biosynthesis were over-expressed in the sample grown in the absence of Casamino-acids (**Figure 18.1**). Examining the 2-D gel protein spots one-by-one along with the functional classification enhanced the efficiency of the analyses enormously. This approach, when applied for multiple gels emerges as a valuable approach to analyse the available data simultaneously (**Figure 18.3**).

Commonly, the proteome maps are compared by aligning and overlapping the maps. These methods are not applicable for proteome maps which are prepared on a different scale of iso-electric focusing strips and the gels which are prepared from a different percentage of acrylamide in the second dimension separation. In our database, by examining the proteins in the proteome maps spot-by-spot individually, aided with the alignment obtained from 2-D gels analysis software, we were able to combine and utilize the functional classification to enhance the analysis. Similar approaches would certainly improve the efficiency of comparison of the data generated from different conditions and from various laboratories. Furthermore, the approaches outlined here could be applied to the analysis of proteomic databases of other organisms.

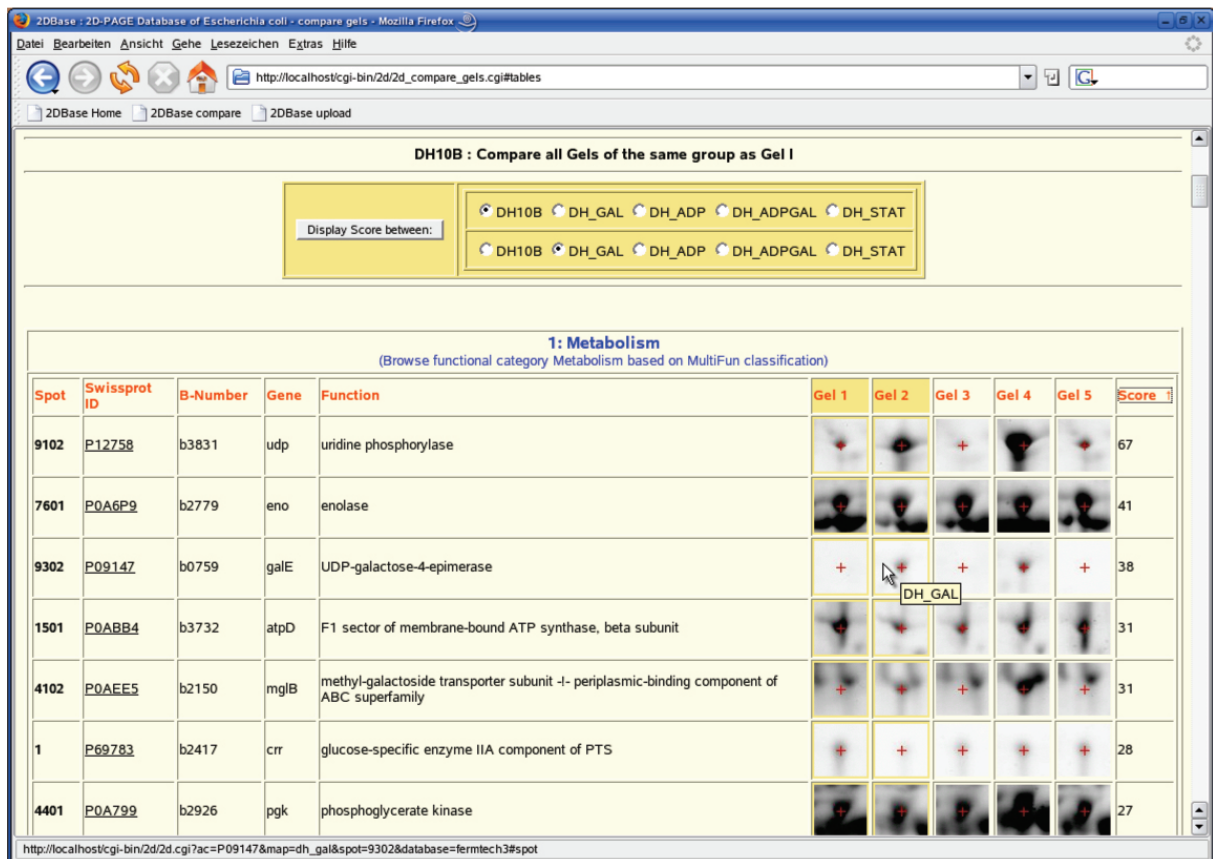


Fig. 18.3: Multiple gels comparison. Simultaneously analysis of five proteome gels spot-by-spot with functional classification.

---

# CHAPTER 19

## EPILOGUE

---

As a result of the analysis of the growth behaviour of the closely related *E.coli* sub-strains W3110 and MG1655 in strictly controlled conditions and the global proteome and transcriptome pools we were able to answer the questions raised in the beginning of the study:

- ***Is the high degree of similarity at the nucleotide level reflected in the metabolic phenotype?***

- The genomes being highly similar at the nucleotide level, was not reflected in metabolic phenotype, showing that complex regulatory patterns can differ even between closely-related sub-strains.

- ***Do sub-strains with almost identical genome structures exhibit similar behaviour in cellular metabolism?***

- The growth behaviour and genome-based data point to the conclusion that both the sub-strains examined may have similar metabolic phenotypes. However, the global transcriptome and proteome data lead to the conclusion that the global proteomes and the genome expression patterns exhibit high plasticity in these closely related sub-strains of *E. coli* - K-12, notwithstanding the strict and controlled growth conditions used.

- ***How do global aspects of cell metabolism, protein synthesis and gene expression differ among closely related sub-strains of the same species revealing possible complexities of cellular metabolism?***

- Among these closely related strains, in MG1655, genes involved in osmotic stress and enterobactin were effectively up-regulated; in W3110, genes involved in central metabolism were up-regulated and RpoS-dependent genes were significantly down-regulated. Identification of a single base change in *rpoS* gene presented in this study emphasizes that complex regulatory pathways are altered by such modifications at the genome level. Functional differences in central metabolism, which generates precursor metabolites and energy, were also apparent between these two closely-related and commonly-used sub-strains of *E. coli* K-12 in both transcript and protein levels.

As a result of the analysis of gene, protein and metabolite levels of the laboratory evolved strains under three different evolutionary conditions, we were able to answer the following questions raised in the beginning of the study:

- ***What are the transcriptome, proteome and metabolome changes occurring during the excess-nutrient adaptive evolution process?***

- Higher levels of metabolites involved in nucleotide pathway and TCA cycle and its intermediates were seen in the excess-nutrient evolved strains. In line with these observations, the expression levels of genes involved in these pathways were also over-expressed in them. Specifically, the pentose phosphate pathway (produces pentose phosphates for nucleic acid synthesis) was differentially regulated, along with the histidine biosynthesis pathway which shares metabolites with the purine and nucleotide biosynthesis pathways. Genes involved in the membrane-related functional categories and their proteins were significantly over-expressed in both the evolved strains.

- ***What are the transcript, protein and metabolite changes occurring due to the pleiotropic effects due to environmental shift?***

- In environmental shift adapted strains several outer-membrane porin proteins were differentially expressed and in accordance with this the genes involved in the membrane-associated GO functional categories were also significantly overrepresented.

- ***Which genes, proteins and metabolites are vitally involved in the prolonged stationary phase evolution process?***

- The genes involved in the biosynthesis pathways of fatty acids (key building blocks for the phospholipid components of cell membranes) and lipids were over-expressed in prolonged stationary phase evolved strains. Several outer-membrane proteins were differentially over-expressed in them, and the genes belonging to the membrane-associated GO functional categories were significantly overrepresented as well.

- ***To what extent the changes occurring during these evolutionary processes are seen in both strains?***

- Clustering by PCA analysis of the multidimensional datasets revealed various degrees of changes among the strains in the evolutionary conditions studied. Excess-nutrient evolved strains cluster separately from their ancestor lines, implying substantial adaptive changes. Environmental shift adapted strains clustered separately from their ancestor lines, indicating considerable environmental shift due to pleiotropic changes. Prolonged stationary phase evolved strains clustered along with their ancestor lines, denoting few changes between ancestor and evolved strains or diverse changes between the evolved strains.



- *Among both the strains, is the path of evolution similar in these evolutionary processes (parallelism)?*

- To examine the extent of parallelism among the strains in particular evolutionary conditions, we applied pairwise correlation map analysis. Prolonged stationary phase evolved strains displayed a high degree of negative correlation when compared to other evolved strains suggesting elevated levels of variability. This demonstrates a low level of parallelism among the prolonged stationary phase evolved strains. The correlation map of excess-nutrient adapted evolved strains exhibited a lower degree of negative correlation than the other strains denoting a reduced level of variability in them representing high level of parallelism among the excess-nutrient adapted strains.

In summary, all the evolutionary experiments demonstrate the vital role of the involvement of the membrane associated components in the evolutionary process. Even though comprehensive analyses at all three functional levels provide no direct answer for the evolutionary processes analysed, further experiments regarding the screening of mutations in these evolved strains in the membrane-related and energy metabolism functional categories, or regulators affecting these functional modules' candidates would certainly provide valuable functional information about the evolutionary processes.

---

## BIBLIOGRAPHY

---

- [1] F. R. Blattner, 3rd Plunkett, G., C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao. The complete genome sequence of escherichia coli k-12. *Science*, 277(5331):1453–74, 1997.
- [2] Genobase database. [<http://ecoli.aist-nara.ac.jp/>].
- [3] K. Hayashi, N. Morooka, Y. Yamamoto, K. Fujita, K. Isono, S. Choi, E. Ohtsubo, T. Baba, B. L. Wanner, H. Mori, and T. Horiuchi. Highly accurate genome sequences of escherichia coli k-12 strains mg1655 and w3110. *Mol Syst Biol*, 2:2006 0007, 2006.
- [4] T. Itoh, T. Okayama, H. Hashimoto, J. Takeda, R. W. Davis, H. Mori, and T. Gojobori. A low rate of nucleotide changes in escherichia coli k-12 estimated from a comparison of the genome sequences between two different substrains. *FEBS Lett*, 450(1-2):72–6, 1999.
- [5] T. L. Ferea, D. Botstein, P. O. Brown, and R. F. Rosenzweig. Systematic changes in gene expression patterns following adaptive evolution in yeast. *Proc Natl Acad Sci U S A*, 96(17):9721–6, 1999.
- [6] M. M. Riehle, A. F. Bennett, R. E. Lenski, and A. D. Long. Evolutionary changes in heat-inducible gene expression in lines of escherichia coli adapted to high temperature. *Physiol Genomics*, 14(1):47–58, 2003.
- [7] T. F. Cooper, D. E. Rozen, and R. E. Lenski. Parallel changes in gene expression after 20,000 generations of evolution in escherichiacoli. *Proc Natl Acad Sci U S A*, 100(3):1072–7, 2003.
- [8] S. S. Fong, A. R. Joyce, and B. O. Palsson. Parallel adaptive evolution cultures of escherichia coli lead to convergent growth phenotypes with different gene expression states. *Genome Res*, 15(10):1365–72, 2005.
- [9] C. G. Knight, N. Zitzmann, S. Prabhakar, R. Antrobus, R. Dwek, H. Hebestreit, and P. B. Rainey. Unraveling adaptive evolution: how a single point mutation affects the protein coregulation network. *Nat Genet*, 38(9):1015–22, 2006.
- [10] A. Kurlandzka, R. F. Rosenzweig, and J. Adams. Identification of adaptive changes in an evolving population of escherichia coli: the role of changes with regulatory and highly pleiotropic effects. *Mol Biol Evol*, 8(3):261–81, 1991.
- [11] L. Pelosi, L. Kuhn, D. Guetta, J. Garin, J. Geiselmann, R. E. Lenski, and D. Schneider. Parallel changes in global protein profiles during long-term experimental evolution in escherichia coli. *Genetics*, 173(4):1851–69, 2006.

- [12] L. M. Wick, M. Quadroni, and T. Egli. Short- and long-term changes in proteome composition and kinetic properties in a culture of *Escherichia coli* during transition from glucose-excess to glucose-limited growth conditions in continuous culture and vice versa. *Environ Microbiol*, 3(9):588–99, 2001.
- [13] A. R. Fernie, R. N. Trethewey, A. J. Krotzky, and L. Willmitzer. Metabolite profiling: from diagnostics to systems biology. *Nat Rev Mol Cell Biol*, 5(9):763–9, 2004.
- [14] L. M. Raamsdonk, B. Teusink, D. Broadhurst, N. Zhang, A. Hayes, M. C. Walsh, J. A. Berden, K. M. Brindle, D. B. Kell, J. J. Rowland, H. V. Westerhoff, K. van Dam, and S. G. Oliver. A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nat Biotechnol*, 19(1):45–50, 2001.
- [15] A. Bernal, U. Ear, and N. Kyrpides. Genomes online database (gold): a monitor of genome projects world-wide. *Nucleic Acids Res*, 29(1):126–7, 2001.
- [16] K. Liolios, N. Tavernarakis, P. Hugenholtz, and N. C. Kyrpides. The genomes on line database (gold) v.2: a monitor of genome projects worldwide. *Nucleic Acids Res*, 34(Database issue):D332–4, 2006.
- [17] J. B. Barbara. Derivations and genotypes of some mutant derivatives of *Escherichia coli* k-12. In F. C. Neidhardt, editor, *Escherichia coli and Salmonella: cellular and molecular biology*, pages 2460–88. ASM Press, Washington, D.C, 1996.
- [18] R. B. Brem, G. Yvert, R. Clinton, and L. Kruglyak. Genetic dissection of transcriptional regulation in budding yeast. *Science*, 296(5568):752–5, 2002.
- [19] V. G. Cheung, L. K. Conlin, T. M. Weber, M. Arcaro, K. Y. Jen, M. Morley, and R. S. Spielman. Natural variation in human gene expression assessed in lymphoblastoid cells. *Nat Genet*, 33(3):422–5, 2003.
- [20] W. Enard, P. Khaitovich, J. Klose, S. Zollner, F. Heissig, P. Giavalisco, K. Nieselt-Struwe, E. Muchmore, A. Varki, R. Ravid, G. M. Doxiadis, R. E. Bontrop, and S. Paabo. Intra- and interspecific variation in primate gene expression patterns. *Science*, 296(5566):340–3, 2002.
- [21] M. F. Oleksiak, G. A. Churchill, and D. L. Crawford. Variation in gene expression within and among natural populations. *Nat Genet*, 32(2):261–6, 2002.
- [22] E. E. Schadt, S. A. Monks, T. A. Drake, A. J. Lusis, N. Che, V. Colinayo, T. G. Ruff, S. B. Milligan, J. R. Lamb, G. Cavet, P. S. Linsley, M. Mao, R. B. Stoughton, and S. H. Friend. Genetics of gene expression surveyed in maize, mouse and man. *Nature*, 422(6929):297–302, 2003.
- [23] L. M. Steinmetz, H. Sinha, D. R. Richards, J. I. Spiegelman, P. J. Oefner, J. H. McCusker, and R. W. Davis. Dissecting the architecture of a quantitative trait locus in yeast. *Nature*, 416(6878):326–30, 2002.

- [24] S. P. Gygi, Y. Rochon, B. R. Franza, and R. Aebersold. Correlation between protein and mrna abundance in yeast. *Mol Cell Biol*, 19(3):1720–30, 1999.
- [25] T. Ideker, V. Thorsson, J. A. Ranish, R. Christmas, J. Buhler, J. K. Eng, R. Bumgarner, D. R. Goodlett, R. Aebersold, and L. Hood. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science*, 292(5518):929–34, 2001.
- [26] E. Soupene, W. C. van Heeswijk, J. Plumbridge, V. Stewart, D. Bertenthal, H. Lee, G. Prasad, O. Paliy, P. Charernnoppakul, and S. Kustu. Physiological studies of escherichia coli strain mg1655: growth defects and apparent cross-regulation of gene expression. *J Bacteriol*, 185(18):5611–26, 2003.
- [27] W. J. Henzel, T. M. Billeci, J. T. Stults, S. C. Wong, C. Grimley, and C. Watanabe. Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc Natl Acad Sci U S A*, 90(11):5011–5, 1993.
- [28] D. P. Zimmer, E. Soupene, H. L. Lee, V. F. Wendisch, A. B. Khodursky, B. J. Peter, R. A. Bender, and S. Kustu. Nitrogen regulatory protein c-controlled genes of escherichia coli: scavenging as a defense against nitrogen limitation. *Proc Natl Acad Sci U S A*, 97(26):14674–9, 2000.
- [29] T. Polen, D. Rittmann, V. F. Wendisch, and H. Sahm. Dna microarray analyses of the long-term adaptive response of escherichia coli to acetate and propionate. *Appl Environ Microbiol*, 69(3):1759–74, 2003.
- [30] O. Thimm, O. Blasing, Y. Gibon, A. Nagel, S. Meyer, P. Kruger, J. Selbig, L. A. Muller, S. Y. Rhee, and M. Stitt. Mapman: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J*, 37(6):914–39, 2004.
- [31] M. H. Serres and M. Riley. Multifun, a multifunctional classification scheme for escherichia coli k-12 gene products. *Microb Comp Genomics*, 5(4):205–22, 2000.
- [32] M. B. Eisen, P. T. Spellman, P. O. Brown, and D. Botstein. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A*, 95(25):14863–8, 1998.
- [33] L. Earle Arnow. Colorimetric determination of the components of 3,4-dihydroxyphenylalaninetyrosine mixtures. *J Biol Chem*, 118:531–37, 1937.
- [34] B. Schwyn and J. B. Neilands. Universal chemical assay for the detection and determination of siderophores. *Anal Biochem*, 160(1):47–56, 1987.
- [35] S. Altuvia, M. Almiron, G. Huisman, R. Kolter, and G. Storz. The dps promoter is activated by oxyr during growth and by ihf and sigma s in stationary phase. *Mol Microbiol*, 13(2):265–72, 1994.

- [36] A. Arnqvist, A. Olsen, and S. Normark. Sigma s-dependent growth-phase induction of the *csgba* promoter in *escherichia coli* can be achieved in vivo by sigma 70 in the absence of the nucleoid-associated protein h-ns. *Mol Microbiol*, 13(6):1021–32, 1994.
- [37] M. P. Castanie-Cornet, T. A. Penfound, D. Smith, J. F. Elliott, and J. W. Foster. Control of acid resistance in *escherichia coli*. *J Bacteriol*, 181(11):3525–35, 1999.
- [38] A. Conter, C. Menchon, and C. Gutierrez. Role of dna supercoiling and rpos sigma factor in the osmotic and growth phase-dependent induction of the gene *osme* of *escherichia coli* k12. *J Mol Biol*, 273(1):75–83, 1997.
- [39] S. Gordia and C. Gutierrez. Growth-phase-dependent expression of the osmotically inducible gene *osmc* of *escherichia coli* k-12. *Mol Microbiol*, 19(4):729–36, 1996.
- [40] R. Hengge-Aronis, W. Klein, R. Lange, M. Rimmele, and W. Boos. Trehalose synthesis genes are controlled by the putative sigma factor encoded by rpos and are involved in stationary-phase thermotolerance in *escherichia coli*. *J Bacteriol*, 173(24):7918–24, 1991.
- [41] M. Ibanez-Ruiz, V. Robbe-Saule, D. Hermant, S. Labrude, and F. Norel. Identification of rpos (sigma(s))-regulated genes in *salmonella enterica* serovar typhimurium. *J Bacteriol*, 182(20):5749–56, 2000.
- [42] S. Lacour and P. Landini. Sigmas-dependent gene expression at the onset of stationary phase in *escherichia coli*: function of sigmas-dependent genes and identification of their promoter sequences. *J Bacteriol*, 186(21):7186–95, 2004.
- [43] C. L. Patten, M. G. Kirchhof, M. R. Schertzberg, R. A. Morton, and H. E. Schellhorn. Microarray analysis of rpos-mediated gene expression in *escherichia coli* k-12. *Mol Genet Genomics*, 272(5):580–91, 2004.
- [44] M. Rahman, M. R. Hasan, T. Oba, and K. Shimizu. Effect of rpos gene knock-out on the metabolism of *escherichia coli* during exponential growth phase and early stationary phase based on gene expressions, enzyme activities and intracellular metabolite concentrations. *Biotechnol Bioeng*, 2006.
- [45] H. E. Schellhorn, J. P. Audia, L. I. Wei, and L. Chang. Identification of conserved, rpos-dependent stationary-phase genes of *escherichia coli*. *J Bacteriol*, 180(23):6283–91, 1998.
- [46] A. R. Strom and I. Kaasen. Trehalose metabolism in *escherichia coli*: stress protection and stress regulation of gene expression. *Mol Microbiol*, 8(2):205–10, 1993.
- [47] H. Weber, T. Polen, J. Heuveling, V. F. Wendisch, and R. Hengge. Genome-wide analysis of the general stress response network in *escherichia coli*: sigmas-dependent genes, promoters, and sigma factor selectivity. *J Bacteriol*, 187(5):1591–603, 2005.
- [48] W. Yang, L. Ni, and R. L. Somerville. A stationary-phase protein of *escherichia coli* that affects the mode of association between the trp repressor protein and operator-bearing dna. *Proc Natl Acad Sci U S A*, 90(12):5796–800, 1993.

- [49] R. Hengge-Aronis. Regulation of gene expression during entry into stationary phase. In F. C. Neidhardt, editor, *Escherichia coli and Salmonella: cellular and molecular biology*, pages 1497–1512. ASM Press, Washington, D.C, 1996.
- [50] R. Hengge-Aronis. Survival of hunger and stress: the role of rpos in early stationary phase gene regulation in e. coli. *Cell*, 72(2):165–8, 1993.
- [51] P. C. Loewen and R. Hengge-Aronis. The role of the sigma factor sigma s (katf) in bacterial global regulation. *Annu Rev Microbiol*, 48:53–80, 1994.
- [52] P. C. Loewen, B. Hu, J. Strutinsky, and R. Sparling. Regulation in the rpos regulon of escherichia coli. *Can J Microbiol*, 44(8):707–17, 1998.
- [53] R. Hengge-Aronis, R. Lange, N. Henneberg, and D. Fischer. Osmotic regulation of rpos-dependent genes in escherichia coli. *J Bacteriol*, 175(1):259–65, 1993.
- [54] R. Hengge-Aronis. Back to log phase: sigma s as a global regulator in the osmotic control of gene expression in escherichia coli. *Mol Microbiol*, 21(5):887–93, 1996.
- [55] H. Tao, C. Bausch, C. Richmond, F. R. Blattner, and T. Conway. Functional genomics: expression analysis of escherichia coli growing on minimal and rich media. *J Bacteriol*, 181(20):6425–40, 1999.
- [56] C. Petersen, L. B. Moller, and P. Valentin-Hansen. The cryptic adenine deaminase gene of escherichia coli. silencing by the nucleoid-associated dna-binding protein, h-ns, and activation by insertion elements. *J Biol Chem*, 277(35):31373–80, 2002.
- [57] K. Schnetz and B. Rak. Is5: a mobile enhancer of transcription in escherichia coli. *Proc Natl Acad Sci U S A*, 89(4):1244–8, 1992.
- [58] R. Hengge-Aronis. Signal transduction and regulatory mechanisms involved in control of the sigma(s) (rpos) subunit of rna polymerase. *Microbiol Mol Biol Rev*, 66(3):373–95, table of contents, 2002.
- [59] T. Schweder, K. H. Lee, O. Lomovskaya, and A. Martin. Regulation of escherichia coli starvation sigma factor (sigma s) by clpxp protease. *J Bacteriol*, 178(2):470–6, 1996.
- [60] T. Atlung, H. V. Nielsen, and F. G. Hansen. Characterisation of the allelic variation in the rpos gene in thirteen k12 and six other non-pathogenic escherichia coli strains. *Mol Genet Genomics*, 266(5):873–81, 2002.
- [61] P. R. Subbarayan and M. Sarkar. A comparative study of variation in codon 33 of the rpos gene in escherichia coli k12 stocks: implications for the synthesis of sigma(s). *Mol Genet Genomics*, 270(6):533–8, 2004.
- [62] C. N. Peterson, V. J. Carabetta, T. Chowdhury, and T. J. Silhavy. Lrha regulates rpos translation in response to the rcs phosphorelay system in escherichia coli. *J Bacteriol*, 188(9):3175–81, 2006.

- [63] T. J. Barnard, Jr. Watson, M. E., and M. A. McIntosh. Mutations in the escherichia coli receptor fepa reveal residues involved in ligand binding and transport. *Mol Microbiol*, 41(3):527–36, 2001.
- [64] K. G. Held and K. Postle. Exbb and exbd do not function independently in tonb-dependent energy transduction. *J Bacteriol*, 184(18):5170–3, 2002.
- [65] D. L. Stephens, M. D. Choe, and C. F. Earhart. Escherichia coli periplasmic protein fepb binds ferrienterobactin. *Microbiology*, 141 ( Pt 7):1647–54, 1995.
- [66] T. J. Brickman and M. A. McIntosh. Overexpression and purification of ferric enterobactin esterase from escherichia coli. demonstration of enzymatic hydrolysis of enterobactin and its iron complex. *J Biol Chem*, 267(17):12350–5, 1992.
- [67] A. Treffry, Z. Zhao, M. A. Quail, J. R. Guest, and P. M. Harrison. How the presence of three iron binding sites affects the iron storage function of the ferritin (ecftna) of escherichia coli. *FEBS Lett*, 432(3):213–8, 1998.
- [68] F. E. Charles. Uptake and metabolism of iron and molybdenum. In F. C. Neidhardt, editor, *Escherichia coli and Salmonella: cellular and molecular biology*, pages 1075–90. ASM Press, Washington, D.C, 1996.
- [69] S. F. Elena and R. E. Lenski. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat Rev Genet*, 4(6):457–69, 2003.
- [70] S. E. Finkel. Long-term survival during stationary phase: evolution and the gasp phenotype. *Nat Rev Microbiol*, 4(2):113–20, 2006.
- [71] B. E. Wright. Stress-directed adaptive mutations and evolution. *Mol Microbiol*, 52(3):643–50, 2004.
- [72] E. R. Zinser and R. Kolter. Escherichia coli evolution during stationary phase. *Res Microbiol*, 155(5):328–36, 2004.
- [73] R. E. Lenski and M. Travisano. Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. *Proc Natl Acad Sci U S A*, 91(15):6808–14, 1994.
- [74] P. D. Sniegowski, P. J. Gerrish, and R. E. Lenski. Evolution of high mutation rates in experimental populations of e. coli. *Nature*, 387(6634):703–5, 1997.
- [75] D. Papadopoulos, D. Schneider, J. Meier-Eiss, W. Arber, R. E. Lenski, and M. Blot. Genomic evolution during a 10,000-generation experiment with bacteria. *Proc Natl Acad Sci U S A*, 96(7):3807–12, 1999.
- [76] S. E. Finkel and R. Kolter. Evolution of microbial diversity during prolonged starvation. *Proc Natl Acad Sci U S A*, 96(7):4023–7, 1999.
- [77] L. Loewe, V. Textor, and S. Scherer. High deleterious genomic mutation rate in stationary phase of escherichia coli. *Science*, 302(5650):1558–60, 2003.

- [78] I. Bjedov, O. Tenaillon, B. Gerard, V. Souza, E. Denamur, M. Radman, F. Taddei, and I. Matic. Stress-induced mutagenesis in bacteria. *Science*, 300(5624):1404–9, 2003.
- [79] M. J. Lombardo, I. Aponyi, and S. M. Rosenberg. General stress response regulator rpos in adaptive mutation and amplification in escherichia coli. *Genetics*, 166(2):669–80, 2004.
- [80] E. R. Zinser and R. Kolter. Prolonged stationary-phase incubation selects for lrp mutations in escherichia coli k-12. *J Bacteriol*, 182(15):4361–5, 2000.
- [81] E. R. Zinser and R. Kolter. Mutations enhancing amino acid catabolism confer a growth advantage in stationary phase. *J Bacteriol*, 181(18):5800–7, 1999.
- [82] J. Kopka, N. Schauer, S. Krueger, C. Birkemeyer, B. Usadel, E. Bergmuller, P. Dorman, W. Weckwerth, Y. Gibon, M. Stitt, L. Willmitzer, A. R. Fernie, and D. Steinhilber. Gmd@csb.db: the golm metabolome database. *Bioinformatics*, 21(8):1635–8, 2005.
- [83] S. Ruberg, Z. X. Tian, E. Krol, B. Linke, F. Meyer, Y. Wang, A. Puhler, S. Weidner, and A. Becker. Construction and validation of a sinorhizobium meliloti whole genome dna microarray: genome-wide profiling of osmoadaptive gene expression. *J Biotechnol*, 106(2-3):255–68, 2003.
- [84] M. H. Serres, S. Goswami, and M. Riley. Genprotec: an updated and improved analysis of functions of escherichia coli k-12 proteins. *Nucleic Acids Res*, 32(Database issue):D300–2, 2004.
- [85] V. Batagelj and A. Mrvar. Pajekprogram for large network analysis. *Connections*, 21:4757, 1998.
- [86] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, 13(11):2498–504, 2003.
- [87] G. D. Bader and C. W. Hogue. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics*, 4:2, 2003.
- [88] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock. Gene ontology: tool for the unification of biology. the gene ontology consortium. *Nat Genet*, 25(1):25–9, 2000.
- [89] S. Maere, K. Heymans, and M. Kuiper. Bingo: a cytoscape plugin to assess over-representation of gene ontology categories in biological networks. *Bioinformatics*, 21(16):3448–9, 2005.
- [90] C. J. Wolfe, I. S. Kohane, and A. J. Butte. Systematic survey reveals general applicability of "guilt-by-association" within gene coexpression networks. *BMC Bioinformatics*, 6:227, 2005.



- [91] A. L. Barabasi and Z. N. Oltvai. Network biology: understanding the cell's functional organization. *Nat Rev Genet*, 5(2):101–13, 2004.
- [92] D. Schluter. Parallel evolution and inheritance of quantitative traits. *Am Nat*, 163(6):809–22, 2004.
- [93] J. J. Bull, M. R. Badgett, H. A. Wichman, J. P. Huelsenbeck, D. M. Hillis, A. Gulati, C. Ho, and I. J. Molineux. Exceptional convergent evolution in a virus. *Genetics*, 147(4):1497–507, 1997.
- [94] R. B. Helling, C. N. Vargas, and J. Adams. Evolution of escherichia coli during growth in a constant environment. *Genetics*, 116(3):349–58, 1987.
- [95] R. V. Sonti and J. R. Roth. Role of gene duplications in the adaptation of salmonella typhimurium to growth on limiting carbon sources. *Genetics*, 123(1):19–28, 1989.
- [96] L. Notley-McRobb and T. Ferenci. Adaptive mgl-regulatory mutations and genetic diversity evolving in glucose-limited escherichia coli populations. *Environ Microbiol*, 1(1):33–43, 1999.
- [97] L. Notley-McRobb and T. Ferenci. The generation of multiple co-existing mal-regulatory mutations through polygenic evolution in glucose-limited populations of escherichia coli. *Environ Microbiol*, 1(1):45–52, 1999.
- [98] L. Notley-McRobb and T. Ferenci. Experimental analysis of molecular events during mutational periodic selections in bacterial evolution. *Genetics*, 156(4):1493–501, 2000.
- [99] M. S. Riley, V. S. Cooper, R. E. Lenski, L. J. Forney, and T. L. Marsh. Rapid phenotypic change and diversification of a soil bacterium during 1000 generations of experimental evolution. *Microbiology*, 147(Pt 4):995–1006, 2001.
- [100] X. Liu and T. Ferenci. An analysis of multifactorial influences on the transcriptional control of ompf and ompc porin expression under nutrient limitation. *Microbiology*, 147(Pt 11):2981–9, 2001.
- [101] H. Nikaido and T. Nakae. The outer membrane of gram-negative bacteria. *Adv Microb Physiol*, 20:163–250, 1979.
- [102] H. Nikaido. Molecular basis of bacterial outer membrane permeability revisited. *Microbiol Mol Biol Rev*, 67(4):593–656, 2003.
- [103] M. W. Covert, E. M. Knight, J. L. Reed, M. J. Herrgard, and B. O. Palsson. Integrating high-throughput and computational data elucidates bacterial networks. *Nature*, 429(6987):92–6, 2004.
- [104] B. G. Hall. Adaptive mutations in escherichia coli as a model for the multiple mutational origins of tumors. *Proc Natl Acad Sci U S A*, 92(12):5669–73, 1995.
- [105] R. Guimera and L. A. Nunes Amaral. Functional cartography of complex metabolic networks. *Nature*, 433(7028):895–900, 2005.

- [106] J. E. Celis, H. H. Rasmussen, P. Gromov, E. Olsen, P. Madsen, H. Leffers, B. Honore, K. Dejgaard, H. Vorum, D. B. Kristensen, and et al. The human keratinocyte two-dimensional gel protein database (update 1995): mapping components of signal transduction pathways. *Electrophoresis*, 16(12):2177–240, 1995.
- [107] T. Rabilloud. Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains. *Proteomics*, 2(1):3–10, 2002.
- [108] Y. Gibon, B. Usadel, O. E. Blaesing, B. Kamlage, M. Hoehne, R. Trethewey, and M. Stitt. Integration of metabolite with transcript and enzyme activity profiling during diurnal cycles in arabidopsis rosettes. *Genome Biol*, 7(8):R76, 2006.
- [109] K. M. Oksman-Caldentey and K. Saito. Integrating genomics and metabolomics for engineering plant metabolic pathways. *Curr Opin Biotechnol*, 16(2):174–9, 2005.
- [110] H. Nikaido and M. Vaara. Molecular basis of bacterial outer membrane permeability. *Microbiol Rev*, 49(1):1–32, 1985.
- [111] S. W. Cowan, T. Schirmer, G. Rummel, M. Steiert, R. Ghosh, R. A. Paupit, J. N. Jansonius, and J. P. Rosenbusch. Crystal structures explain functional properties of two e. coli porins. *Nature*, 358(6389):727–33, 1992.
- [112] V. C. Wasinger, S. J. Cordwell, A. Cerpa-Poljak, J. X. Yan, A. A. Gooley, M. R. Wilkins, M. W. Duncan, R. Harris, K. L. Williams, and I. Humphery-Smith. Progress with gene-product mapping of the mollicutes: *Mycoplasma genitalium*. *Electrophoresis*, 16(7):1090–4, 1995.
- [113] K. Mostaguir, C. Hoogland, P. A. Binz, and R. D. Appel. The make 2d-db ii package: conversion of federated two-dimensional gel electrophoresis databases into a relational format and interconnection of distributed databases. *Proteomics*, 3(8):1441–4, 2003.

---

## SUPPLEMENTARY MATERIALS

---

Gene name	Primer name	Primers
<i>nuoA</i>	nuoA-1	5' GAAAAACGTGCCGTTTGAAT 3'
	nuoA-2	5' GCAGCTTCCACAAAGCCTAC 3'
<i>nuoH</i>	nuoH-1	5' TCGGGATTTTGTTCCTTCCCTG 3'
	nuoH-2	5' TTGACGATGTTCGGTCATGTT 3'
<i>lrhA</i>	lrhA-1	5' ATGGCAGAAATGCTGGAATC 3'
	lrhA-2	5' CGCCATGGAATATCTGCTTT 3'
<i>entD</i>	entD-1	5' TCAACCGATTGGCATTGATA 3'
	entD-2	5' TTGCCTTAAATGCGCTCTCT 3'
<i>fepB</i>	fepB-1	5' TGCGCAGTTTGATAAGCAAC 3'
	fepB-2	5' CATGGCGTTTACCCTGACTT 3'
<i>fepE</i>	fepE-1	5' CGAAAATCGACGAACTGGAT 3'
	fepE-2	5' CGGACGTTTTCTAGCGACTC 3'
<i>feoA</i>	feoA-1	5' GCATATCGCCAAAAACTGCT 3'
	feoA-2	5' TAACAGGAAACCGCTTCCAC 3'
<i>feoB</i>	feoB-1	5' ATTTTCCTGAGCGCTTTCAA 3'
	feoB-2	5' TTCTGCCGGATTGAACTCTT 3'
<i>rpoS</i>	rpoS-1	5' GCACGTGAGTTGTCCATAA 3'
	rpoS-2	5' TAAGACGAAGCATAACGGCTG 3'
<i>dnaA</i>	dnaA-1	5' ATCATTCTCACCTCGGATCG 3'
	dnaA-2	5' AGACGCTTGGCGATAAAGAA 3'
1- Forward primer and 2- Reverse primer		

Tab. 19.1: Primers used for amplification.

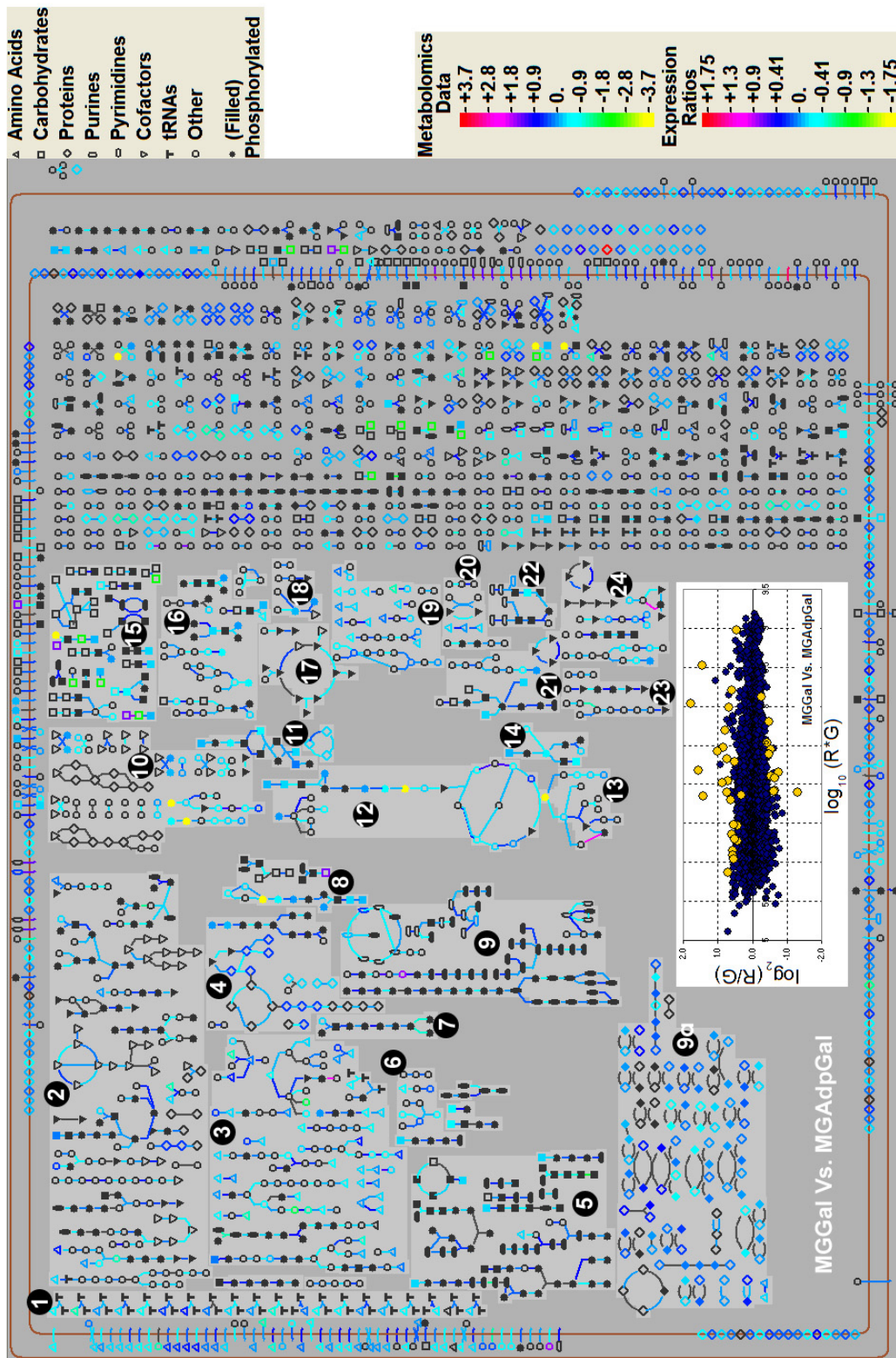


Fig. 19.1: Integration of transcriptome and metabolome data during the comparison of MGGal and MGAdpGal strains in environmental shift evolutionary condition.

Metabolic overview for *E. coli* illustrating the results of high-throughput transcriptome and metabolome experiments in a global metabolic pathway context. Gene expression levels are mapped to reaction steps involved in a metabolism, and the range of data values levels in a given experimental dataset is mapped to a spectrum of colors. Similarly, for metabolomics experiments, compound nodes are colored according to the data value for the corresponding compound. Various metabolism related pathways are labeled as Figure 8.7



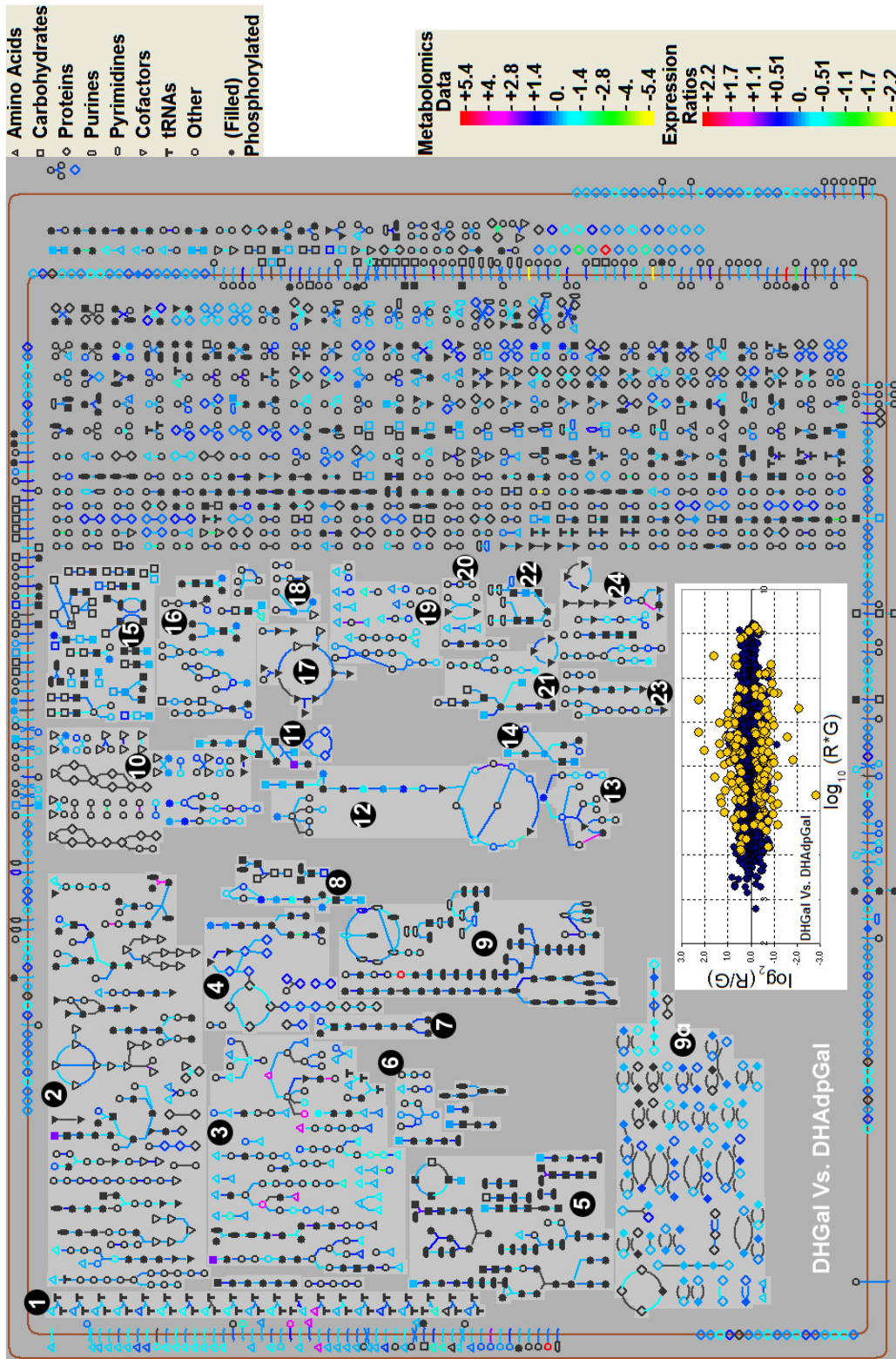


Fig. 19.2: Integration of transcriptome and metabolome data during the comparison of DHGal and DHAdpGal strains in environmental shift evolutionary condition.

Metabolic overview for *E. coli* illustrating the results of high-throughput transcriptome and metabolome experiments in a global metabolic pathway context. Gene expression levels are mapped to reaction steps involved in a metabolism, and the range of data values levels in a given experimental dataset is mapped to a spectrum of colors. Similarly, for metabolomics experiments, compound nodes are colored according to the data value for the corresponding compound. Various metabolism related pathways are labeled as Figure 8.7

Ratio	B No	Gene	Function	Reference
X (-3.92)	b4376	osmY	Hyperosmotically inducible periplasmic protein	[43]
X (-3.65)	b1732	katE	Catalase; hydroperoxidase HPII (III), RpoS dependent	[47]
X (-3.59)	b0871	poxB	Pyruvate dehydrogenase/oxidase:FAD -binding	[47]
X (-3.01)	b2465	tktB	Transketolase 2, thiamin-binding, isozyme	[43]
X (-2.68)	b2464	talA	Transaldolase A	[43]
X (-2.63)	b1897	otsB	Trehalose-6-phosphate phosphatase,biosynthetic	[46]
X (-2.53)	b3073	ygjG	Putrescine:2-oxoglutaric acid aminotransferase	[47]
X (-2.52)	b1896	otsA	Trehalose-6-phosphate synthase	[46]
X (-2.44)	b2080	yegP	Conserved hypothetical protein	[43]
X (-2.42)	b1050	yceK	Unknown CDS	[47]
X (-2.42)	b1783	yeaG	Nucleotide triphosphate hydrolase domain	[43]
X (-2.25)	b0812	dps	Stress response DNA-binding protein	[35]
X (-2.01)	b1681	sufD	Required for stability of Fe-S component of FhuF	[43]
A (-2.00)	b4127	yjdJ	Putative acyl-CoA N-acyltransferase	[43]
A (-1.99)	b1927	amyA	Cytoplasmic alpha-amylase	[47]
A (-1.96)	b3517	gadA	Glutamate decarboxylase A, isozyme	[37]
A (-1.91)	b1493	gadB	Glutamate decarboxylase, PLP-dependent	[37]
A (-1.83)	b1967	yedU	Molecular chaperone Hsp 31	[43]
A (-1.68)	b2137	yohF	Putative oxidoreductase with NAD(P)-binding domains	[47]
A (-1.62)	b3012	dkgV	2,5-Diketo-D-gluconate reductase A	[47]
A (-1.59)	b3453	ugpB	Sn-Glycerol 3-phosphate transport protein	[47]
B (-1.50)	b1283	osmB	Lipoprotein, osmotically inducible	[40]
B (-1.50)	b0435	bolA	Activator of morphogenic pathway (BolA family)	[43]
B (-1.39)	b0865	yhjP	Conserved hypothetical protein	[47]
B (-1.38)	b0186	IdcC	Lysine decarboxylase 2, constitutive	[47]
B (-1.32)	b3099	yqjE	Conserved protein	[43]
B (-1.32)	b3098	yqjD	Conserved hypothetical protein	[43]
B (-1.31)	b2660	ygaF	Putative enzyme	[47]
B (-1.30)	b0113	pdhR	Transcriptional repressor for dehydrogenase complex	[47]

**Table 19.2 to be continued...**

Ratio	B No	Gene	Function	Reference
B (-1.26)	b1197	treA	Trehalase, periplasmic	[40]
B (-1.21)	b1004	wrbA	Flavodoxin-like, trp repressor-binding protein	[48]
B (-1.19)	b0643	ybeL	Conserved hypothetical protein	[47]
B (-1.18)	b3519	treF	Trehalase, cytoplasmic	[47]
B (-1.18)	b1724	ydiZ	Conserved hypothetical protein	[43]
B (-1.16)	b0453	ybaY	glycoprotein/polysaccharide metabolism	[43]
B (-1.15)	b0808	ybiO	Putative transport protein, integral membrane location	[47]
B (-1.13)	b1480	rpsV	30S ribosomal subunit protein S22	[43]
C (-1.07)	b3003	yghA	Putative oxidoreductase; NAD(P)-binding domain	[47]
C (-1.07)	b4135	yjdC	Putative regulator with DNA-binding domain	[47]
C (-1.05)	b0329	yahO	Conserved hypothetical protein	[41]
C (-1.04)	b2661	gabD	Succinate-semialdehyde dehydrogenase	[47]
C (-1.04)	b1614	ydgA	Conserved protein	[47]
C (-0.99)	b3509	hdeB	Conserved hypothetical protein	[36]
C (-0.99)	b1482	osmC	Resistance protein, osmotically inducible	[39]
C (-0.90)	b2922	yggE	Conserved protein	[47]
C (-0.89)	b2732	ygbA	Conserved hypothetical protein	[47]
C (-0.84)	b3510	hdeA	Acid-resistance protein, possible chaperone	[36]
C (-0.83)	b4263	yjgR	Putative enzyme contains hydrolase domain	[47]
C (-0.79)	b2659	csiD	Conserved protein with clavaminatase synthase domain	[47]
C (-0.79)	b1235	rssB	Response regulator involved in protein turnover	[47]
C (-0.76)	b2452	eutH	Putative transport protein, ethanolamine utilization	[47]
C (-0.76)	b0419	yajO	Putative oxidoreductase, NAD(P) dependent	[47]
C (-0.76)	b3830	ysgA	Putative diene lactone hydrolase	[47]
C (-0.76)	b1619	hdhA	Alpha-hydroxysteroid dehydrogenase, NAD dependent	[47]
D (-0.61)	b1646	sodC	Superoxide dismutase precursor (Cu-Zn)	[43]
D (-0.60)	b3516	gadX	Transcriptional regulator of (gadA, gad BC)	[43]
D (-0.57)	b1440	ydcS	Putative transport protein (ABC superfamily)	[45]
D (-0.56)	b1217	chaB	Cation transport regulator	[47]

**Table 19.2 to be continued...**

Ratio	B No	Gene	Function	Reference
D (-0.54)	b0864	artP	Arginine transport protein	[47]
D (-0.51)	b1519	tam	Trans-aconitate methyltransferase	[47]
D (-0.51)	b0854	potF	Putrescine transport protein	[47]
D (-0.48)	b3291	mscL	Mechanosensitive channel	[47]
D (-0.47)	b4187	aidB	Putative acyl-CoA dehydrogenase (flavoprotein)	[43]
D (-0.45)	b3024	ygiW	Conserved hypothetical protein	[43]
D (-0.45)	b3450	ugpC	Sn-Glycerol 3-phosphate transport protein	[47]
D (-0.43)	b3449	ugp	Glycerophosphodiester phosphodiesterase, cytosolic	[47]
D (-0.39)	b1739	osmE	Transcriptional activator of ntrL gene	[38]
D (-0.38)	b3515	gadW	Activator of (GadA,BC), and repressor of gadX	[43]
E (-0.21)	b0389	yaiA	Unknown CDS	[47]
E (-0.19)	b2717	hycI	Protease involved in processing C-terminal end of HycE	[47]
E (-0.14)	b0861	artM	Arginine transport protein	[47]
E (-0.13)	b2301	yfcF	Putative glutathione S-transferase enzyme	[47]
E (-0.01)	b1100	ycfH	Putative metallo-dependent hydrolase	[43]
E (0.01)	b1957	yodC	Unknown CDS	[43]
E (0.12)	b2543	yphA	Putative transmembrane protein	[47]
E (0.22)	b4051	qor	Quinone oxidoreductase, NADPH dependent	[47]

Tab. 19.2: Expression data of RpoS dependent and osmotic genes during exponential phase.



<b>Ratio</b>	<b>B No</b>	<b>Gene</b>	<b>Function</b>
C (-0.78)	b1891	flhC	Transcriptional regulator of flagellar class II biosynthesis, anaerobic respiration
F (0.55)	b2711	norW	NADH:flavorubredoxin (FIRd) oxidoreductase
F (0.56)	b2279	nuoK	NADH dehydrogenase I chain K
F (0.57)	b2202	napC	Periplasmic nitrate reductase, cytochrome c-type protein
F (0.58)	b2265	menF	Isochorismate synthase (isochorismate hydroxymutase 2)
F (0.58)	b0729	sucD	Succinyl-CoA synthetase, alpha subunit, NAD(P)-binding, TCA
F (0.58)	b2283	nuoG	NADH dehydrogenase I chain G
F (0.60)	b3843	ubiD	3-octaprenyl-4-hydroxybenzoate decarboxylase
F (0.61)	b2729	hypD	Hydrogenase expression/formation protein
F (0.62)	b3236	mdh	Malate dehydrogenase, NAD(P)-binding, TCA.
F (0.62)	b0730	farR	Transcriptional repressor of the mngAB operon
G (0.67)	b3604	lldR	Putative transcriptional repressor for L-lactate utilization (GntR family)
G (0.69)	b2280	nuoJ	NADH dehydrogenase I chain J
G (0.70)	b0733	cydA	Cytochrome d terminal oxidase, polypeptide subunit I, Aerobic Growth.
G (0.71)	b0118	acnB	Bifunctional: aconitate hydratase 2; 2-methylisocitrate dehydratase, TCA
G (0.73)	b2240	glpT	Glycerol-3-phosphate transport protein (MFS family)
G (0.74)	b2994	hybC	Hydrogenase-2, large subunit
G (0.79)	b3366	nirD	Nitrite reductase, small subunit
G (0.82)	b2210	mqo	Malate dehydrogenase, TCA.
G (0.82)	b1800	yeaU	Putative tartrate dehydrogenase
G (0.82)	b2281	nuoI	NADH dehydrogenase I chain I, 2Fe-2S ferredoxin-related
G (0.86)	b2486	hyfF	Hydrogenase 4, F membrane subunit
G (0.87)	b3891	fdhE	Formate dehydrogenase formation protein
G (0.87)	b2287	nuoB	NADH dehydrogenase I chain B, binds FeS cluster N2
G (0.87)	b2726	hypA	Guanine-nucleotide-binding protein in formate-hydrogenlyase system
G (0.88)	b2488	hyfH	Hydrogenase 4, Fe-S subunit
G (1.04)	b0723	sdhA	Succinate dehydrogenase, catalytic and NAD/flavoprotein subunit, TCA
G (1.04)	b3028	mdaB	Flavoprotein modulator of drug activity B

**Table 19.3 to be continued...**

Ratio	B No	Gene	Function
G (1.10)	b2278	nuoL	NADH dehydrogenase I chain L, membrane subunit
G (1.10)	b2487	hyfG	Hydrogenase 4 subunit
H (1.11)	b2288	nuoA	NADH dehydrogenase I chain A
H (1.11)	b2286	nuoC	NADH dehydrogenase I chain C, D
H (1.14)	b0724	sdhB	Succinate dehydrogenase, Fe-S protein, Aerobic Growth;TCA
H (1.30)	b2285	nuoE	NADH dehydrogenase I chain E
I (1.56)	b2483	hyfC	Hydrogenase 4, C (membrane) subunit
I (1.57)	b2208	napF	Fe-S ferredoxin-type protein, electron transfer
I (1.61)	b1468	narZ	Nitrate reductase 2, alpha subunit, Anaerobic Growth.
I (1.63)	b4152	frdC	Fumarate reductase, anaerobic, membrane anchor polypeptide
I (1.93)	b2282	nuoH	NADH dehydrogenase I chain H
I (1.94)	b2730	hypE	Hydrogenase 3 maturation protein

Tab. 19.3: Expression data of energy metabolism genes during exponential phase.

<b>Ratio</b>	<b>B No</b>	<b>Gene</b>	<b>Function</b>
X (-2.31)	b0583	entD	Phosphopantetheinyltransferase, component of the enterobactin synthase
X (-2.25)	b0812	dps	Stress response DNA-binding protein, Iron storage Protein.
X (-2.01)	b1681	sufD	Required for stability of Fe-S component of FhuF
A (-1.91)	b0584	fepA	Outer membrane porin, receptor for ferric enterobactin and colicins B and D
B (-1.53)	b1102	fhuE	Outer membrane pore protein, receptor for Fe(III)-coprogen
B (-1.53)	b0595	entB	Aryl carrier protein (C-terminal) (enterobactin synthase multienzyme complex )
B (-1.38)	b0594	entE	2,3-dihydroxybenzoate-AMP ligase, component of the enterobactin synthase
B (-1.31)	b0805	ybiL	Outer membrane porin protein, putative ferrisiderophore receptor
B (-1.20)	b0585	fes	Ferric enterobactin esterase
B (-1.15)	b0808	ybiO	Putative transport protein, integral membrane location
C (-1.06)	b0593	entC	Isochorismate synthetase, enterobactin-specific
C (-0.88)	b2526	hscA	Chaperone (Hsp70 family), believed to be involved in assembly of Fe-S clusters
C (-0.87)	b0592	fepB	Ferric enterobactin transport protein (ABC superfamily)
C (-0.79)	b0586	entF	Serine activating enzyme, component of the enterobactin synthase complex
C (-0.75)	b1252	tonB	Energy transducer; uptake of iron, cyanocobalamin;
D (-0.60)	b3005	exbD	Uptake of enterobactin; tonB-dependent uptake of B colicins
D (-0.57)	b0591	ybdA	Putative transport protein (POT family)
D (-0.47)	b3006	exbB	Uptake of enterobactin; tonB-dependent uptake of B colicins
D (-0.40)	b0588	fepC	Ferric enterobactin transport protein (ABC superfamily)
D (-0.34)	b0589	fepG	Ferric enterobactin transport protein (ABC superfamily, membrane)
D (-0.24)	b0596	entA	2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase
G (1.03)	b1905	ftnA	Cytoplasmic ferritin (iron storage protein)
H (1.53)	b1902	ftnB	Putative ferritin-like protein
H (1.44)	b3408	feoA	Ferrous iron transport protein A
H (1.13)	b3409	feoB	Ferrous iron transport protein B (FeoB family)

Tab. 19.4: **Expression data of iron related metabolism genes during exponential phase.**

<b>W3110</b>	<b>MG1655</b>	<b>B No</b>	<b>Gene</b>	<b>Function</b>	<b>Reference</b>
Y (3.48)	F (0.62)	b3830	ysgA	Putative diene lactone hydrolase	[47]
Y (3.15)	F (0.42)	b3515	gadW	Activator of (GadA,BC), and repressor of gadX	[43]
Y (3.07)	Y (2.30)	b0113	pdhR	Transcriptional repressor for pyruvate dehydrogenase	[47]
Y (2.59)	E (-0.07)	b4135	yjdB	Putative regulator with homeodomain domain	[47]
Y (2.57)	F (0.45)	b3516	gadX	transcriptional regulator of (gadA, gad BC)	[43]
Y (2.53)	B (-1.55)	b3517	gadA	glutamate decarboxylase A, isozyme	[37]
Y (2.37)	G (0.93)	b1004	wrbA	flavodoxin-like, trp repressor-binding protein	[48]
I (2.00)	F (0.42)	b0854	potF	Putrescine transport protein	[47]
I (1.93)	E (-0.21)	b4263	yjgR	Putative enzyme contains nucleoside domain	[47]
I (1.86)	A (-1.83)	b1283	osmB	lipoprotein, osmotically inducible	[40]
I (1.82)	G (0.86)	b2465	tkdB	transketolase 2, thiamin-binding, isozyme	[43]
I (1.74)	E (0.14)	b0389	yaiA	Unknown CDS	[47]
I (1.73)	H (1.54)	b4376	osmY	hyperosmotically inducible periplasmic protein	[43]
I (1.73)	B (-1.22)	b1493	gadB	glutamate decarboxylase, PLP-dependent	[37]
H (1.28)	G (0.95)	b3099	yqjE	conserved protein	[43]
H (1.18)	C (-1.02)	b1482	osmC	resistance protein, osmotically inducible	[39]
H (1.15)	E (0.00)	b1619	hdhA	alpha-hydroxysteroid dehydrogenase	[47]
G (1.01)	D (-0.47)	b1050	yceK	Unknown CDS	[47]
G (0.95)	F (0.35)	b1681	sufD	required for stability of Fe-S component of FhuF	[43]
G (0.94)	F (0.45)	b3024	ygiW	conserved hypothetical protein	[43]
G (0.81)	E (-0.01)	b1480	rpsV	30S ribosomal subunit protein S22	[43]
G (0.79)	D (-0.39)	b1897	otsB	trehalose-6-phosphate phosphatase, biosynthetic	[46]
G (0.70)	C (-0.82)	b3003	yghA	Putative oxidoreductase; NAD(P)-binding domain	[47]
G (0.67)	B (-1.26)	b2080	yegP	conserved hypothetical protein	[43]
F (0.55)	B (-1.16)	b2137	yohF	Putative oxidoreductase	[47]
F (0.44)	D (-0.46)	b3450	ugpC	Sn-Glycerol 3-phosphate transport protein	[47]

**Table 19.5 to be continued...**

<b>W3110</b>	<b>MG1655</b>	<b>B No</b>	<b>Gene</b>	<b>Function</b>	<b>Reference</b>
F (0.41)	I (1.60)	b1732	katE	Catalase; hydroperoxidase HPII (III)	[47]
F (0.39)	E (-0.03)	b1927	amyA	Cytoplasmic alpha-amylase	[47]
F (0.35)	F (0.46)	b0871	poxB	Pyruvate dehydrogenase/oxidase	[47]
F (0.34)	A (-1.91)	b0435	bolA	activator of morphogenic pathway (Bola family)	[43]
F (0.32)	B (-1.26)	b3073	ygjG	Putrescine:2-oxoglutaric acid aminotransferase	[47]
F (0.28)	B (-1.24)	b1197	treA	trehalase, periplasmic	[40]
F (0.25)	B (-1.50)	b0329	yahO	conserved hypothetical protein	[41]
F (0.22)	A (-1.79)	b0812	dps	stress response DNA-binding protein	[35]
E (0.01)	C (-0.81)	b3012	dkgV	2,5-Diketo-D-gluconate reductase A	[47]
E (-0.16)	B (-1.12)	b1967	yedU	molecular chaperone Hsp 31	[43]
C (-0.85)	X (-2.37)	b3453	ugpB	Sn-Glycerol 3-phosphate transport protein	[47]
C (-0.93)	X (-2.08)	b2659	csiD	Conserved protein with clavamate synthase domain	[47]

Tab. 19.5: Expression data of RpoS dependent genes during exponential phase versus early stationary phase.

Tab. 19.6: Gene expression changes in experimental evolutionary conditions.

Bin	Functional annotation	No of Genes	MG/MGadp	No of Genes	DH/ DHadp
1.2	Macromolecule degradation > GO:macromolecule catabolism ; GO:0009057	105	<b>0.045</b>	105	0.312
1.3	Energy metabolism (carbon) > GO:energy derivation by oxidation of organic compounds ; GO:0015980	299	<b>0.012</b>	299	0.920
2.1	DNA related > GO:DNA metabolism ; GO:0006259	163	<b>0.006</b>	163	<b>0.004</b>
4	Transport > GO:transporter activity ; GO:0005215 > GO:transport ; GO:0006810	1215	0.783	1215	<b>0.028</b>
5.1	Cell division > GO:cytokinesis ; GO:0000910	95	<b>0.035</b>	95	0.230
6.6	Ribosome > GO:cytosolic ribosome (sensu Bacteria) ; GO:0009281	68	0.098	68	<b>0.018</b>
8	extrachromosomal > GO:.	300	0.300	300	<b>0.027</b>
8.4	Colicin related > GO:.	9	0.165	9	<b>0.034</b>
1.1.2.2	3-phenylpropionate and 3-(3-hydroxyphenyl)propionate degradation > GO:3-phenylpropionate catabolism ; GO:0019380 > GO:3-(3-hydroxy)phenylpropionate catabolism ; GO:0019622	11	<b>0.049</b>	11	<b>0.000</b>
1.1.3	Amino acids > GO:amino acid catabolism ; GO:0009063	74	0.927	74	<b>0.034</b>
1.1.3.5	Glycine cleavage > GO:glycine decarboxylation via glycine cleavage system ; GO:0019464	7	0.611	7	<b>0.010</b>
1.1.5.2	Ethanol degradation > GO:ethanol catabolism ; GO:0006068	12	<b>0.025</b>	12	0.271
1.2.2	DNA > GO:DNA catabolism ; GO:0006308	26	<b>0.043</b>	26	0.068
1.3.9	Entner-Doudoroff pathway > GO:Entner-Doudoroff pathway ; GO:0009255	4	<b>0.041</b>	4	0.323
1.5.1.11	Serine > GO:serine biosynthesis ; GO:0006564	6	<b>0.025</b>	6	0.080
1.5.1.12	Cysteine > GO:cysteine biosynthesis ; GO:0019344	6	<b>0.049</b>	6	0.927
1.5.1.15	Tryptophan > GO:tryptophan biosynthesis ; GO:0000162	13	<b>0.012</b>	13	<b>0.020</b>
1.5.1.17	Alanine > GO:alanine biosynthesis ; GO:0006523	4	<b>0.038</b>	4	0.770
1.5.1.21	Homoserine > GO:homoserine biosynthesis ; GO:0009090	6	0.752	6	<b>0.010</b>
1.5.1.4	Proline > GO:proline biosynthesis ; GO:0006561	8	0.752	8	<b>0.036</b>
1.5.1.7	Lysine, diaminopimelate > GO:lysine biosynthesis via diaminopimelate ; GO:0009089	16	0.153	16	<b>0.038</b>
1.5.3.14	Enterochelin (enterobactin) > GO:enterobactin biosynthesis ; GO:0009239	6	0.589	6	<b>0.011</b>
1.5.3.4	Molybdenum (molybdopterin) > GO:Mo-molybdopterin cofactor biosynthesis ; GO:0006777	18	<b>0.021</b>	18	0.108
1.5.3.6	Pyridoxine (vitamin B6) > GO:pyridoxine biosynthesis ; GO:0008615	9	0.677	9	<b>0.036</b>
1.5.4	Fatty acid and phosphatidic acid > GO:fatty acid biosynthesis ; GO:0006633 ; GO:phosphatidic acid biosynthesis ; GO:0006654	26	<b>0.005</b>	26	0.215
1.6.12	Flagella > GO:flagella biogenesis ; GO:0009296	37	0.258	37	<b>0.043</b>

1.6.3.2	Core region > GO:lipopolysaccharide core region biosynthesis ; GO:0009244	17	0.952	17	<b>0.029</b>
1.6.6	Osmoregulated periplasmic glucan > GO:glucan biosynthesis ; GO:0009250	2	<b>0.049</b>	2	0.218
1.7.1.18	Betaine biosynthesis > GO:betaine biosynthesis ; GO:0006578	4	0.302	4	<b>0.011</b>
1.7.3	Pentose phosphate shunt, non-oxidative branch > GO:pentose-phosphate shunt, non-oxidative branch ; GO:0009052	8	<b>0.036</b>	8	<b>0.019</b>
2.1.5	DNA degradation > GO:DNA catabolism ; GO:0006308	27	<b>0.049</b>	27	0.069
2.2.3	RNA modification > GO:RNA modification ; GO:0009451	53	<b>0.028</b>	53	<b>0.029</b>
2.3.2	Translation > GO:protein biosynthesis ; GO:0006412	74	0.121	74	<b>0.024</b>
2.3.6	Turnover, degradation > GO:proteolysis and peptidolysis ; GO:0006508	10	0.527	10	<b>0.002</b>
2.3.8	Ribosomal proteins > GO:structural constituent of ribosome ; GO:0003735	57	0.190	57	<b>0.017</b>
3.1.1	DNA structure level > GO:	14	<b>0.010</b>	14	<b>0.002</b>
3.1.1.2	Methylation > GO:DNA methylation ; GO:0006306	5	<b>0.022</b>	5	0.083
3.1.3.1	Translation attenuation and efficiency > GO:translational attenuation ; GO:0009386	10	<b>0.024</b>	10	0.101
3.1.3.4	Proteases, cleavage of compounds > GO:peptidase activity ; GO:0008233	9	0.268	9	<b>0.036</b>
4.10.119	methylgalactoside/galactose > GO:galactose transporter activity ; GO:0005354 > GO:galactose transporter activity ; GO:0005354 > GO:methylgalactoside transporter activity ; GO:0015592 > GO:methylgalactoside transporter ; GO:0015765	4	0.660	4	<b>0.049</b>
4.10.155	phosphate > GO:phosphate transport ; GO:0006817 > GO:phosphate transporter activity ; GO:0015114	7	<b>0.030</b>	7	0.643
4.10.185	uracil > GO:uracil transport ; GO:0015857 > GO:uracil transporter activity ; GO:0015210	2	<b>0.032</b>	2	0.058
4.10.31	chloramphenicol > GO:chloramphenicol transport ; GO:0042892	2	<b>0.023</b>	2	<b>0.022</b>
4.10.5	alkanesulfonate > GO:alkanesulphonate transport ; GO:0042918 > GO:alkanesulphonate transporter activity ; GO:0042959	3	0.159	3	<b>0.036</b>
4.10.78	glycerol-3-P > GO:glycerol transport ; GO:0015793 > GO:glycerol-3-phosphate transporter activity ; GO:0015169	5	<b>0.020</b>	5	0.103
4.10.79	glycine betaine choline transport > GO:betaine transport ; GO:0015838 > choline transport ; GO:0015871	5	<b>0.046</b>	5	0.325
4.2.1	Porters (Uni-, Sym- and Antiporters) > GO:symporter activity ; GO:0015293 > GO:antiporter activity ; GO:0015297 > GO:uniporter activity ; GO:0015292	203	<b>0.050</b>	203	0.178
4.2.1.20	The Inorganic Phosphate Transporter (PiT) Family > GO:inorganic phosphate transporter ; GO:0005315	3	<b>0.018</b>	3	0.454
4.2.1.21	The Solute:Sodium Symporter (SSS) Family > GO:solute:sodium symporter ; GO:0015370	4	<b>0.015</b>	4	<b>0.008</b>
4.2.1.7	The Small Multidrug Resistance (SMR) Family > GO:multidrug transporter ; GO:0015239	6	<b>0.050</b>	6	0.190
4.3.1	Pyrophosphate Bond (ATP; GTP; P2) Hydrolysis-driven Active Transporters > GO:ATPase activity ; GO:0016887	262	<b>0.046</b>	262	0.603
4.3.1.1	The ATP-binding Cassette (ABC) Superfamily + ABC-type Uptake Permeases > GO:ATP-binding cassette (ABC) transporter activity ; GO:0004009	235	<b>0.029</b>	235	0.417
4.8.1.3	MPA1 Family auxiliary transport protein> GO:	2	0.083	2	<b>0.035</b>
4.9.1	Recognized transporters of unknown biochemical mechanism > GO:transporter activity ; GO:0005215 > GO:transport ; GO:0006810	4	0.243	4	<b>0.034</b>

Tab. 19.6: Significant (Wilcoxon rank sum test) functional categories involved in environmental shift evolutionary condition.

	Functional annotation	No of Genes	MGal/ MGAdpGal	No of Genes	DHGal/ DHAdpGal
<b>B3</b>					
1.2	Macromolecule degradation > GO:macromolecule catabolism ; GO:0009057	104	0.801	104	0.022
6.3	Surface antigens (ECA, O antigen of LPS) > GO:cell surface antigen activity, host-interacting ; GO:0042280	63	0.184	63	0.002
8	extrachromosomal > GO:	292	0.233	292	0.002
8.1	Prophage genes and phage related functions > GO:	258	0.459	258	0.002
8.4	Colicin related > GO:	9	0.028	9	0.660
1.2.2	DNA > GO:DNA catabolism ; GO:0006308	25	0.854	25	0.023
1.2.4	Polysaccharides > GO:polysaccharide catabolism ; GO:000272	7	0.034	7	0.978
1.5.1.11	Serine > GO:serine biosynthesis ; GO:0006564	6	0.046	6	0.600
1.5.1.16	Histidine > GO:histidine biosynthesis ; GO:0000105	12	0.003	12	0.845
1.5.3.8	Thiamine (Vitamin B1) > GO:thiamin biosynthesis ; GO:0009228	15	0.666	15	0.013
1.6.3	Lipopolysaccharide > GO:lipopolysaccharide biosynthesis ; GO:0009103	47	0.081	47	0.001
1.6.3.1	O antigen > GO:O antigen biosynthesis ; GO:0009243	15	0.014	15	0.000
1.6.3.2	Core region > GO:lipopolysaccharide core region biosynthesis ; GO:0009244	17	0.829	17	0.004
1.6.4	Enterobacterial common antigen (surface glycolipid) > GO:enterobacterial common antigen biosynthesis ; GO:0009246	12	0.053	12	0.016
1.7.10	Sugar nucleotide biosynthesis, conversions > GO:nucleotide-sugar biosynthesis ; GO:0009226	15	0.341	15	0.008
1.7.13	Amino acid conversion > GO:amino acid metabolism ; GO:0006520	3	0.011	3	0.770
1.7.14	Polyamine biosynthesis > GO:polyamine biosynthesis ; GO:0006596	19	0.004	19	0.710
1.7.7	Galactose metabolism > GO:galactose metabolism ; GO:0006012	5	0.036	5	0.286
2.2.5	tRNA > GO:tRNA metabolism ; GO:0006399	2	0.036	2	0.401
3.1.2.3	Repressor > GO:transcriptional repressor activity ; GO:0016564	138	0.037	138	0.509
3.1.4	Regulation level unknown > GO:	73	0.019	73	0.746
3.3.1	Operon (regulation of one operon) > GO:	127	0.205	127	0.027
4.10.118	methionine > GO:methionine transport ; GO:0015821 > GO:methionine transporter activity ; GO:0015191	4	0.904	4	0.045
4.10.12	amino acid > GO:amino acid transporter activity ; GO:0015171 > GO:amino acid transport ; GO:0006865	12	0.628	12	0.047
4.10.173	sugar > GO:carbohydrate transport ; GO:0008643 > GO:carbohydrate transporter activity ; GO:0015144	33	0.006	33	0.674
4.10.177	thiamine > GO:thiamin transport ; GO:0015888 > GO:thiamin transporter activity ; GO:0015234	4	0.920	4	0.001
4.10.191	Zn > GO:zinc ion transport ; GO:0006829 > GO:zinc ion transporter activity ; GO:0005385	3	0.028	3	0.491
4.10.71	glucose > GO:glucose transport ; GO:0015758 > GO:glucose transporter activity ; GO:0005355	2	0.099	2	0.047
4.10.82	H+ > GO:proton transport ; GO:0015992 > GO:hydrogen ion transporter activity ; GO:0015078	14	0.530	14	0.007



4.2.1.38	The K+ Transporter (Trk) Family > GO:potassium:hydrogen antiporter ; GO:0015386 > GO:potassium:hydrogen symporter ; GO:0015387	3	0.776	3	<b>0.039</b>
4.3.1.1.1	ATP binding component > GO:	89	0.159	89	<b>0.018</b>
4.3.1.2	The H+/Na+-translocating F-, V- and A-type ATPase (F-ATPase) Superfamily > GO:hydrogen-translocating F-type ATPase complex (sensu Bacteria) ; GO:0045256 > GO:hydrogen-translocating V-type ATPase complex ; GO:0016471	8	0.640	8	<b>0.012</b>

Tab. 19.6: Significant (Wilcoxon rank sum test) functional categories involved in prolonged stationary phase evolutionary condition.

Bin	Functional Annotation	No of Genes	MG/ MGStat	No of Genes	DH/ DHStat
1.6	Macromolecules (cellular constituent) biosynthesis > GO:macromolecule biosynthesis ; GO:0009059	272	<b>0.001</b>	272	0.335
5.2	Cell cycle physiology > GO:cell cycle ; GO:0007049	5	0.786	5	<b>0.009</b>
5.3	Motility (incl. chemotaxis, energytaxis, aerotaxis, redotaxis) > GO:taxis ; GO:0042330	59	<b>0.043</b>	59	0.946
6.3	Surface antigens (ECA, O antigen of LPS) > GO:cell surface antigen activity, host-interacting ; GO:0042280	65	<b>0.048</b>	65	0.280
6.6	Ribosome > GO:cytosolic ribosome (sensu Bacteria) ; GO:0009281	66	0.514	66	<b>0.031</b>
8	extrachromosomal > GO:	298	0.214	298	<b>0.001</b>
8.1	Prophage genes and phage related functions > GO:	264	0.107	264	<b>0.000</b>
1.1.1.18	Trehalose degradation, low osmolarity > GO:trehalose catabolism ; GO:0005993	3	0.898	3	<b>0.047</b>
1.1.1.25	L-ascorbate degradation > GO:L-ascorbic acid catabolism ; GO:0019854	11	0.981	11	<b>0.001</b>
1.1.1.3	D-arabinose catabolism > GO:D-arabinose catabolism ; GO:0019571	3	0.077	3	<b>0.011</b>
1.1.2.2	3-phenylpropionate and 3-(3-hydroxyphenyl)propionate degradation > GO:3-phenylpropionate catabolism ; GO:0019380 > GO:3-(3-hydroxy)phenylpropionate catabolism ; GO:0019622	11	0.527	11	<b>0.018</b>
1.1.2.3	Propionate degradation > GO:propionate catabolism ; GO:0019543	9	0.968	9	<b>0.010</b>
1.1.3.7	Threonine catabolism > GO:threonine catabolism ; GO:0006567	12	<b>0.050</b>	12	<b>0.023</b>
1.1.4.3	Ornithine degradation > GO:ornithine catabolism ; GO:0006593	2	<b>0.026</b>	2	0.104
1.1.5	Others > GO:	29	0.171	29	<b>0.033</b>
1.1.5.2	Ethanol degradation > GO:ethanol catabolism ; GO:0006068	10	0.347	10	<b>0.018</b>
1.5.1.17	Alanine > GO:alanine biosynthesis ; GO:0006523	4	0.962	4	<b>0.004</b>
1.5.1.21	Homoserine > GO:homoserine biosynthesis ; GO:0009090	6	0.271	6	<b>0.024</b>
1.5.1.6	Asparagine > GO:asparagine biosynthesis ; GO:0006529	4	<b>0.038</b>	4	0.191
1.5.3	Cofactor, small molecule carrier > GO:coenzymes and prosthetic group biosynthesis ; GO:0046138	168	<b>0.009</b>	168	0.300
1.5.3.12	Heme, porphyrine > GO:heme biosynthesis ; GO:0006783 > GO:porphyrin biosynthesis ; GO:0006779	16	0.971	16	<b>0.041</b>

1.5.3.2	Folic acid > GO:folic acid biosynthesis ; GO:0046656	10	<b>0.035</b>	10	<b>0.061</b>
1.5.3.5	Coenzyme A > GO:coenzyme A biosynthesis ; GO:0015937	11	<b>0.038</b>	11	<b>0.467</b>
1.5.3.7	Nicotinamide adenine dinucleotide (NAD) > GO:nicotinamide adenine dinucleotide biosynthesis ; GO:0009435	15	0.142	15	<b>0.025</b>
1.5.4	Fatty acid and phosphatidic acid > GO:fatty acid biosynthesis ; GO:0006633 ; GO:phosphatidic acid biosynthesis ; GO:0006654	27	0.319	27	<b>0.020</b>
1.6.1	Phospholipid > GO:phospholipid biosynthesis ; GO:0008654	21	<b>0.000</b>	21	0.142
1.7.26	Allantoin assimilation > GO:allantoin assimilation ; GO:0009442	8	0.722	8	<b>0.018</b>
1.7.8	Gluconeogenesis > GO:gluconeogenesis ; GO:0006094	17	0.056	17	<b>0.050</b>
2.3.2	Translation > GO:protein biosynthesis ; GO:0006412	73	0.543	73	<b>0.021</b>
2.3.8	Ribosomal proteins > GO:structural constituent of ribosome ; GO:0003735	56	0.241	56	<b>0.028</b>
3.1.2.1	Sigma factors, anti-sigmafactors > GO:sigma factor activity ; GO:0016987 > GO:sigma factor antagonist activity ; GO:0016989	17	<b>0.004</b>	17	0.132
4.1.2	Beta barrel porins (The Outer Membrane Porin (OMP) Functional Superfamily) > GO:porin ; GO:0015288	18	<b>0.045</b>	18	0.570
4.10.10	allose/ribose > GO:allose transport ; GO:0015754 > GO:ribose transport ; GO:0015752 > GO:allose transporter activity ; GO:0015593 > GO:ribose transporter activity ; GO:0015591	4	<b>0.039</b>	4	<b>0.003</b>
4.10.118	methionine > GO:methionine transport ; GO:0015821 > GO:methionine transporter activity ; GO:0015191	4	<b>0.010</b>	4	0.118
4.10.119	methylgalactoside/galactose > GO:galactose transporter activity ; GO:0005354 > GO:galactose transporter activity ; GO:0005354 > GO:methylgalactoside transport ; GO:0015765	4	0.462	4	<b>0.016</b>
4.10.143	Ni++ > GO:nickel ion transport ; GO:0015675 > GO:nickel ion transporter activity ; GO:0015099	5	0.343	5	<b>0.032</b>
4.10.148	oligopeptide > GO:oligopeptide transport ; GO:0006857 > GO:oligopeptide transporter activity ; GO:0015198	12	0.557	12	<b>0.001</b>
4.10.152	peptide > GO:peptide transport ; GO:0015833 > GO:peptide transporter activity ; GO:0015197	12	<b>0.025</b>	12	0.060
4.10.155	phosphate > GO:phosphate transport ; GO:0006817 > GO:phosphate transporter activity ; GO:0015114	7	0.277	7	<b>0.006</b>
4.10.185	uracil > GO:uracil transport ; GO:0015857 > GO:uracil transporter activity ; GO:0015210	2	0.832	2	<b>0.034</b>
4.10.29	cellobiose/arbutin/salicin > GO:salicin transport ; GO:0042948 > GO:salicin transporter activity ; GO:0042950 > GO:cellobiose transport ; GO:0019533 > GO:cellobiose transporter activity ; GO:0019191 > GO:salicin transporter activity ; GO:0042950	4	<b>0.004</b>	4	<b>0.027</b>
4.10.48	D-ribose > GO:ribose transport ; GO:0015752 > GO:D-ribose transporter activity ; GO:0015591	6	<b>0.042</b>	6	0.469
4.10.54	Fe++ > GO:ferric iron transport ; GO:0015682 > GO:ferric iron transporter activity ; GO:0015091	2	<b>0.025</b>	2	0.955
4.10.58	ferrichrome > GO:ferrichrome transport ; GO:0042928 > GO:ferrichrome transporter activity ; GO:0042929	5	0.581	5	<b>0.017</b>
4.10.75	glutamate/aspartate > GO:glutamate transport ; GO:0015813 > GO:aspartate transport ; GO:0015810 > GO:glutamate transporter activity ; GO:0005313 > GO:aspartate transporter activity ; GO:0015183	4	0.789	4	<b>0.008</b>
4.10.78	glycerol-3-P > GO:glycerol transport ; GO:0015793 > GO:glycerol-3-phosphate transporter activity ; GO:0015169	5	0.256	5	<b>0.033</b>
4.10.82	H+ > GO:proton transport ; GO:0015992 > GO:hydrogen ion transporter activity ; GO:0015078	14	0.971	14	<b>0.022</b>
4.2.1.13	The C4-Dicarboxylate Uptake (Dcu) Family > GO:C4-dicarboxylate transporter ; GO:0015556	2	<b>0.045</b>	2	0.209
4.2.1.21	The Solute:Sodium Symporter (SSS) Family > GO:solute:sodium symporter ; GO:0015370	4	0.848	4	<b>0.020</b>
4.2.1.40	The Nucleobase:Cation Symporter-2 (NCS2) Family > GO:nucleobase:cation symporter ; GO:0015391	5	0.879	5	<b>0.006</b>

4.2.1.42	The Aromatic Amino Acid Permease (AraAP) Family > GO:hydroxy/aromatic amino-acid permease ; GO:0015507	7	<b>0.001</b>	7	<b>0.946</b>
4.2.1.47	The Divalent Anion:Na+ Symporter (DASS) Family > GO:.	3	<b>0.048</b>	3	<b>0.153</b>
4.3.1.1.1	ATP binding component > GO:.	92	<b>0.005</b>	92	<b>0.068</b>
4.3.1.1.3	periplasmic binding component> GO:.	51	0.428	51	<b>0.041</b>
4.4.1.3	PTS Lactose-N'-Diacetylchitobiose-betaucoside (Lac) Family > GO:.	3	<b>0.018</b>	3	<b>0.016</b>
4.4.1.5	The PTS Galactitol (Gat) Family > GO:galactitol permease ; GO:0015588	5	<b>0.021</b>	5	<b>0.710</b>
4.8.1.3	MPA1 Family auxiliary transport protein> GO:.	2	0.151	2	<b>0.035</b>
4.8.1.7	The Phosphotransferase System Enzyme I (EI) Family> GO:.	7	0.214	7	<b>0.030</b>
4.9.2.22	The Putative Permease (PerM) Family> GO:.	2	0.680	2	<b>0.034</b>
5.5.4	pH response > GO:response to pH ; GO:0009268	13	0.394	13	<b>0.033</b>
5.5.6	Other stresses (mechanical, nutritional, oxidative) > GO:response to mechanical stimulus ; GO:0009612 > GO:response to nutrients ; GO:0007584 > GO:response to oxidative stress ; GO:0006979	22	0.353	22	<b>0.025</b>

Tab.19.7: Identified metabolites of the ancestral and evolved strains by gas chromatography-mass spectrometry. Response ratios of metabolites having  $P$ -Value  $\leq 0.05$  are shown in bold. Metabolites having correlation coefficient  $\geq 0.5$  are underlined. Response ratios of metabolites having  $P$ -Value  $\leq 0.05$  and correlation coefficient  $\geq 0.5$  are shown in bold underline.

Metabolites	MG / MG_Gal	MG / MG_Adp	MG / MG_Adpgal	MG / MG_stat	MG_Gal / MG_Adpgal	MG_Adp / MG_Adpgal	MG_Adp / MG_stat	DH / DH_Gal	DH / DH_Adp	DH / DH_Adpgal	DH / DH_stat	DH_Gal / DH_Adpgal	DH_Adp / DH_Adpgal	DH_Adp / DH_stat
2-Aminoadipate (260)	1.26	1.08	0.50	1.03	0.40	0.46	0.96	0.78	1.56	0.99	0.59	1.27	0.64	0.38
2-Hydroxyglutarate (203,247)	1.38	0.69	0.91	1.24	<b>0.66</b>	1.33	1.81	1.34	<b>1.84</b>	1.41	1.53	1.05	0.77	0.83
2-Isopropylmalate (275)	<u>0.99</u>	0.63	0.59	1.33	0.60	0.94	2.12	<u>0.95</u>	<u>0.75</u>	<b>0.44</b>	0.75	<b>0.47</b>	0.59	1.00
2-Ketocaproate (216)	1.00	0.45	0.52	<u>0.32</u>	<u>0.52</u>	1.16	<u>0.71</u>	0.95	0.87	0.55	<u>0.67</u>	<u>0.59</u>	<u>0.64</u>	<u>0.77</u>
2-Methylcitrate (287)	<u>0.52</u>	0.54	0.53	0.99	1.03	0.99	1.85	<u>0.35</u>	0.72	<b>0.21</b>	0.78	0.59	<b>0.28</b>	1.08
2-Methylisocitrate (259)	<u>3.02</u>	0.75	<u>3.02</u>	1.21	1.00	<b>4.04</b>	1.62	<u>2.64</u>	0.69	<u>1.25</u>	0.53	<u>0.47</u>	1.80	0.77
4-Aminobutyrate (174)	2.29	<u>1.44</u>	<u>2.60</u>	0.96	1.14	<b>1.81</b>	0.66	<b>2.62</b>	<u>0.44</u>	<u>2.94</u>	0.83	1.12	<b>6.61</b>	1.87
5-Methyl-thioadenosine (236)	<u>0.72</u>	<b>0.43</b>	<b>0.57</b>	0.78	0.79	1.33	1.82	<u>1.36</u>	<b>0.68</b>	0.92	1.13	0.85	<b>1.69</b>	<b>1.66</b>
Adenine (264)	0.71	<b>0.47</b>	0.66	0.41	0.92	1.40	0.88	1.22	<b>0.53</b>	0.92	0.76	0.75	<b>1.75</b>	1.44
Adenosine (236)	1.10	1.05	1.50	2.09	1.36	1.42	1.99	0.84	1.35	2.70	1.06	3.20	1.99	0.78
a-Glycerophosphate (357)	0.97	0.66	0.85	0.55	0.88	1.30	0.84	0.99	<b>0.67</b>	0.99	0.91	1.00	<b>1.49</b>	<b>1.36</b>
a-Ketoglutarate (198)	0.54	0.20	0.73	0.15	1.34	3.65	0.78	1.41	<b>0.19</b>	2.03	<b>0.26</b>	1.45	<b>10.94</b>	1.41
Alanine (116)	0.67	0.55	0.73	0.54	1.08	1.33	1.00	1.58	<b>0.51</b>	1.01	0.89	<u>0.64</u>	<b>1.96</b>	<b>1.73</b>
Arabinose (217)	1.42	1.04	0.83	<b>1.57</b>	<b>0.59</b>	0.80	1.51	1.13	<u>1.74</u>	1.56	1.37	1.38	0.90	0.79
Arginine (256)	1.58	0.52	0.92	0.76	0.58	1.75	1.45	<b>1.58</b>	0.69	1.21	1.10	0.77	1.76	1.59
Asparagine (216)	0.65	<b>0.33</b>	0.53	<b>0.41</b>	0.83	1.62	1.25	1.16	<b>0.37</b>	0.66	0.81	0.57	1.79	2.19
b-Alanine (248)	1.80	<u>0.81</u>	1.13	1.39	0.63	1.39	1.71	1.01	<u>1.08</u>	1.23	1.23	1.21	1.13	1.13
Citrate (257)	1.06	<u>0.52</u>	<b>0.62</b>	1.25	<b>0.59</b>	<u>1.20</u>	<b>2.41</b>	1.24	<u>1.01</u>	1.11	0.77	<u>0.89</u>	<u>1.10</u>	0.76
Cystathionine (128)	0.91	<b>0.19</b>	<u>0.28</u>	0.61	<u>0.31</u>	1.44	<b>3.17</b>	0.87	0.35	<u>0.54</u>	1.41	<u>0.62</u>	1.54	4.07
DHAP (400)	0.88	0.75	0.87	0.82	0.98	1.16	1.09	1.70	<u>0.51</u>	1.98	0.83	1.16	3.91	<u>1.65</u>
Diaminopimelate (200,272)	1.33	0.57	1.03	0.47	0.77	1.81	0.83	1.13	1.01	1.62	1.50	1.43	1.60	1.48
Erythrose-4-P (357)	3.04	4.45	3.45	4.83	1.13	0.78	1.09	0.17	1.85	<u>0.80</u>	2.48	4.84	0.43	<u>1.34</u>
Fructose-6-P (315)	0.75	0.86	0.61	0.81	0.82	0.71	0.94	0.91	<b>0.38</b>	0.81	<b>0.56</b>	0.89	<b>2.15</b>	1.49
Fructose (307)	<b>0.66</b>	1.01	<b>0.53</b>	1.31	0.81	<b>0.52</b>	1.30	<b>0.45</b>	0.62	<b>0.46</b>	0.93	1.01	0.73	1.49

Tab.19.7: Identified metabolites of the ancestral and evolved strains by gas chromatography-mass spectrometry.

Fumarate (245)	2.38	0.75	1.98	0.47	0.83	2.64	0.63	2.58	0.51	3.68	0.88	1.43	7.28	1.74
Gluconate-6-P (387)	1.20	1.02	0.91	1.06	0.76	0.89	1.03	0.89	1.28	0.90	1.51	1.01	0.71	1.18
Gluconate (333)	3.22	1.24	2.32	1.21	0.72	1.87	0.98	1.09	1.18	1.97	1.63	1.81	1.67	1.38
Glucose (319)	1.47	2.03	0.35	0.55	0.24	0.17	0.27	0.85	0.70	0.70	0.80	0.82	0.99	1.13
Glucose-6-P (387)	0.77	0.74	0.68	1.13	0.88	0.91	1.52	0.86	0.56	0.74	0.62	0.86	1.30	1.10
Glucuronic acid (333)	1.12	0.17	0.71	1.71	0.63	4.09	9.83	3.07	1.99	3.04	1.43	0.99	1.53	0.72
Glutamate (230,246)	0.90	0.50	0.71	0.58	0.79	1.42	1.17	1.70	0.55	1.18	0.95	0.69	2.12	1.71
Glutamine (155)	0.93	0.91	0.90	0.72	0.98	0.99	0.79	1.34	0.92	1.18	1.07	0.88	1.29	1.16
Glycerate-2-P (299,315,459)	1.00	0.58	0.68	0.62	0.68	1.16	1.06	1.12	0.65	1.13	0.99	1.02	1.74	1.52
Glycerate (189,192)	1.56	1.17	1.20	0.91	0.77	1.03	0.78	1.36	0.66	0.99	0.76	0.73	1.48	1.14
Glycerate-3-P (227,299,459)	1.29	1.04	1.35	1.33	1.05	1.30	1.28	1.17	0.98	0.76	0.56	0.65	0.77	0.57
Glycine (174)	1.30	0.72	1.00	0.68	0.77	1.39	0.95	0.94	0.63	1.10	0.95	1.17	1.74	1.50
Isocitrate (245,319)	1.00	0.53	0.64	1.15	0.64	1.20	2.18	0.86	1.04	1.10	0.89	1.29	1.07	0.86
Isoleucine (158)	2.12	0.90	1.15	0.92	0.54	1.28	1.02	1.02	0.51	1.36	0.87	1.33	2.67	1.70
L,L-Cystathionine (218)	0.38	0.39	0.37	0.65	0.96	0.93	1.66	0.59	0.92	0.46	0.66	0.78	0.50	0.71
Lactate (191)	1.03	0.80	0.66	0.54	0.64	0.83	0.67	1.01	0.61	0.89	0.93	0.88	1.46	1.52
L-Aspartate (232)	0.93	1.05	0.95	0.85	1.02	0.90	0.81	1.26	0.75	1.08	0.90	0.85	1.44	1.20
L-Cysteine (220)	1.28	0.71	0.54	1.40	0.42	0.76	1.99	0.97	1.02	0.95	1.54	0.97	0.93	1.50
Leucine (158)	0.72	0.82	1.01	0.25	1.41	1.23	0.31	0.86	0.07	0.37	0.20	0.43	5.37	2.93
L-Homocysteine (234)	1.39	0.54	0.60	1.02	0.43	1.11	1.90	0.59	1.15	0.74	1.47	1.24	0.64	1.28
L-Homoserine (218)	3.62	1.27	2.12	0.52	0.59	1.68	0.41	0.58	2.69	5.84	2.72	10.00	2.17	1.01
Lysine (156)	0.50	0.36	0.37	1.43	0.73	1.01	3.95	2.69	0.36	0.84	1.20	0.31	2.30	3.28
Malate (245,307)	1.21	0.51	0.91	0.43	0.75	1.78	0.85	1.92	0.45	1.33	0.71	0.69	2.94	1.56
Maleic acid (245)	0.24	ND	0.51	ND	2.12	0.00	ND	1.27	ND	1.37	ND	1.08	0.00	ND
Methionine (176)	0.90	0.49	0.58	0.51	0.64	1.18	1.03	1.06	0.58	0.86	0.87	0.81	1.49	1.51
myo-Inositol (305)	1.01	0.67	1.53	0.87	1.52	2.29	1.31	2.73	4.60	6.82	4.52	2.49	1.48	0.98
Myo-Inositol-P (318)	1.08	1.13	1.20	1.28	1.11	1.06	1.13	0.40	0.74	1.85	0.67	4.66	2.52	0.92
N-Acetyl-Aspartate (274)	2.31	3.71	11.68	0.57	5.05	3.15	0.15	4.82	19.18	61.43	1.11	12.73	3.20	0.06
N-Acetyl-L-Serine (261)	3.06	1.02	0.89	2.48	0.29	0.87	2.42	1.20	1.47	1.18	1.25	0.98	0.80	0.85
o-acetyl-L-Homoserine (202)	1.69	0.83	0.56	0.45	0.33	0.68	0.54	1.15	0.90	1.08	1.29	0.94	1.20	1.44
o-acetyl-L-Serine (132)	1.55	0.72	ND	6.55	ND	ND	9.07	3.39	1.28	1.21	0.97	0.36	0.95	0.76
Ornithine (142)	1.03	0.48	0.68	0.55	0.66	1.40	1.15	1.06	0.73	1.06	0.90	1.00	1.45	1.24
Orotic acid (254)	1.84	2.75	5.63	1.29	3.06	2.04	0.47	1.96	31.00	88.75	1.25	45.22	2.86	0.04
Panthoic acid (201)	2.28	1.11	0.97	1.43	0.42	0.87	1.29	2.35	0.93	1.10	1.11	0.47	1.18	1.19
PEP (369)	6.36	1.34	0.48	ND	0.08	0.36	ND	1.15	0.53	2.14	0.25	1.86	4.05	0.47

Tab.19.7: Identified metabolites of the ancestral and evolved strains by gas chromatography-mass spectrometry.

Phenylalanine (192)	2.46	1.10	2.30	1.88	0.94	2.09	1.70	1.31	1.21	2.20	1.88	1.68	1.82	1.56
Phosphate19.28 (299)	0.94	0.45	0.51	5.95	0.54	1.14	13.29	0.75	<b>14.23</b>	1.22	4.38	1.64	0.09	0.31
Proline (142)	1.27	<u>0.53</u>	1.28	<u>0.63</u>	<u>1.01</u>	<b>2.41</b>	1.18	1.71	<u>0.62</u>	1.41	<u>1.27</u>	<u>0.83</u>	<b>2.28</b>	<b>2.06</b>
Putrescine (142,174)	0.33	<b>0.41</b>	<b>0.43</b>	0.38	1.30	1.04	0.92	0.70	0.95	0.91	1.12	1.29	0.95	1.17
Pyruvate (174)	1.05	<b>0.47</b>	0.82	0.33	0.78	1.74	0.69	1.19	0.87	1.84	0.87	1.55	<b>2.11</b>	0.99
Ribitol	0.86	0.76	0.91	0.66	<u>1.06</u>	<u>1.19</u>	0.86	0.94	0.78	0.90	0.94	<u>0.96</u>	<u>1.15</u>	1.20
Ribose (217)	<u>0.69</u>	<b>0.39</b>	<b>0.47</b>	1.08	0.68	1.21	<b>2.80</b>	0.86	<u>0.42</u>	0.90	1.05	1.05	2.14	2.49
Ribose-5-P (315,299)	1.04	<b>0.55</b>	0.69	0.80	0.66	1.24	1.44	1.08	0.68	1.13	0.86	1.04	1.65	1.26
Salicylicacid (267)	1.03	<b>0.43</b>	0.72	0.40	0.70	<u>1.67</u>	0.92	1.51	<b>0.48</b>	1.21	0.57	0.80	<b>2.55</b>	1.19
Serine (204)	1.22	0.65	0.92	0.53	<u>0.75</u>	1.41	0.81	1.45	0.84	1.26	1.20	<u>0.87</u>	<u>1.50</u>	1.42
Shikimate (204)	0.42	0.51	<b>0.38</b>	0.11	<u>0.91</u>	<u>0.75</u>	0.22	0.52	0.55	0.37	0.11	<u>0.72</u>	<u>0.68</u>	0.21
Spermidine (144)	1.05	0.67	<b>0.60</b>	0.93	0.58	0.90	1.39	0.83	0.94	0.93	0.92	1.12	0.99	0.98
Spermine (144)	0.93	0.81	<u>1.01</u>	1.10	<u>1.09</u>	1.25	1.36	1.24	1.15	<u>0.90</u>	0.88	<u>0.72</u>	0.78	0.77
Succinate (247,409)	0.97	<u>0.45</u>	0.97	0.85	<u>1.01</u>	<u>2.17</u>	1.89	1.72	<u>0.64</u>	1.31	1.08	<u>0.76</u>	<b>2.03</b>	1.68
Sucrose (361)	0.85	0.71	0.78	0.78	<u>0.92</u>	1.10	1.10	1.06	0.77	0.89	0.77	<u>0.84</u>	1.15	1.00
Threonine (101)	<b>2.97</b>	<u>0.92</u>	1.62	0.71	<u>0.54</u>	1.75	0.77	0.60	0.70	4.26	2.33	<b>7.12</b>	6.07	3.32
Thymine (255)	<b>0.53</b>	<u>0.63</u>	<u>0.66</u>	<u>0.47</u>	<u>1.23</u>	1.03	0.74	<b>0.67</b>	<b>0.36</b>	<b>0.55</b>	<b>0.69</b>	0.82	1.51	<b>1.91</b>
Trehalose (361)	<b>2.39</b>	<b>2.00</b>	<b>6.63</b>	0.59	2.77	<b>3.32</b>	<b>0.29</b>	<b>2.01</b>	0.85	<b>3.97</b>	<b>0.53</b>	1.98	4.69	0.63
Tryptophan (202)	0.77	0.71	0.82	0.67	<u>1.05</u>	1.14	0.94	0.72	0.65	0.56	0.67	<u>0.77</u>	0.86	1.04
Tyrosine (218)	1.46	1.05	0.69	1.67	0.47	0.66	1.59	0.92	3.58	<u>0.62</u>	1.03	0.67	0.17	0.29
Unknown14.80 (228)	1.09	0.73	0.82	0.26	0.75	1.12	0.36	1.09	<b>0.32</b>	1.44	0.66	1.31	<b>4.46</b>	<b>2.06</b>
Unknown32.96 (361)	<b>183.88</b>	<u>1.32</u>	<b>231.43</b>	1.88	<u>1.26</u>	<b>175.44</b>	1.42	<b>196.78</b>	<b>4.59</b>	<b>255.29</b>	1.43	<u>1.30</u>	<b>55.62</b>	0.31
Uracil (255,241)	1.09	<b>0.46</b>	1.03	0.70	<u>0.95</u>	<b>2.24</b>	1.50	1.42	<b>0.49</b>	<b>1.88</b>	1.11	<u>1.33</u>	<b>3.83</b>	<b>2.27</b>
Urea (189)	0.60	0.06	<u>0.17</u>	0.06	<u>0.29</u>	2.81	0.99	1.43	0.30	<u>0.25</u>	0.44	<u>0.18</u>	0.84	1.46
Valine (144)	0.36	0.66	<b>0.53</b>	0.53	<u>1.47</u>	0.80	0.81	0.99	<b>0.54</b>	<b>0.44</b>	0.64	<b>0.44</b>	0.81	1.19

Tab. 19.8: Identified proteins of the ancestral and evolved strains by 2D-PAGE.

**Metabolism**

Swissprot ID	B-Number	Gene	Function	MG1655	MG_GAL	MG_ADP	MG_ADPGAL	MG_STAT	DH1655	DH_GAL	DH_ADP	DH_ADPGAL	DH_STAT
P0ADI4	b0595	entB	isochorismatase	+	+	+	+	+	+	+	+	+	+
P62707	b0755	gpmA	phosphoglyceromutase 1	+	+	+	+	+	+	+	+	+	+
P0A9C3	b0756	galM	galactose-1-epimerase (mutarotase)	+	+	+	+	+	+	+	+	+	+
P0A6R0	b1091	fabH	3-oxoacyl-[acyl-carrier-protein] synthase III	+	+	+	+	+	+	+	+	+	+
P08200	b1136	icd	e14 prophage; isocitrate dehydrogenase, specific for NADP+	+	+	+	+	+	+	+	+	+	+
P0AEE5	b2150	mgIB	methyl-galactoside transporter subunit; periplasmic-binding component of ABC superfamily	+	+	+	+	+	+	+	+	+	+
P0AEU0	b2309	hisJ	histidine/lysine/arginine/ornithine transporter subunit; periplasmic-binding component of ABC superfamily	+	+	+	+	+	+	+	+	+	+
P09551	b2310	argT	lysine/arginine/ornithine transporter subunit; periplasmic-binding component of ABC superfamily	+	+	+	+	+	+	+	+	+	+

P0ABK5	b2414	cysK	cysteine synthase A, O-acetylserine sulfhydrylase A subunit	
P69783	b2417	crr	glucose-specific enzyme IIA component of PTS	
P0A825	b2551	glyA	serine hydroxymethyltransferase	
P0A6P9	b2779	eno	enolase	
P0A6P9	b2779	eno	enolase	
P0A799	b2926	pgk	phosphoglycerate kinase	
P61889	b3236	mdh	malate dehydrogenase, NAD(P)-binding	
P0AD96	b3460	livJ	leucine/isoleucine/valine transporter subunit; periplasmic-binding component of ABC superfamily	
P0ABB4	b3732	atpD	F1 sector of membrane-bound ATP synthase, beta subunit	
P0ABB4	b3732	atpD	F1 sector of membrane-bound ATP synthase, beta subunit	
P12758	b3831	udp	uridine phosphorylase	
P0A9G6	b4015	aceA	isocitrate lyase	



P0A9G6	b4015	aceA	isocitrate lyase																	
P0A7A9	b4226	ppa	inorganic pyrophosphatase																	
P0A9B2	b1779	gapA	glyceraldehyde-3-phosphate dehydrogenase A																	
P37902	b0655	gltI	glutamate and aspartate transporter subunit; periplasmic-binding component of ABC superfamily																	
P0A870	b0008	talB	transaldolase B																	
P09147	b0759	galE	UDP-galactose-4-epimerase																	

### Information Transfer

P0A6Y8	b0014	dnaK	chaperone Hsp70, co-chaperone with DnaJ																	
P0A6Y8	b0014	dnaK	chaperone Hsp70, co-chaperone with DnaJ																	
P0A6P1	b0170	tsf	protein chain elongation factor EF-Ts																	
P16659	b0194	proS	prolyl-tRNA synthetase																	
P0ADI4	b0595	entB	isochorismatase																	

P0AEE5	b2150	mgIB	methyl-galactoside transporter subunit; periplasmic-binding component of ABC superfamily															
P0A6N1	b3339	tufA	protein chain elongation factor EF-Tu (duplicate of tufB)															
P0A6N1	b3980	tufB	protein chain elongation factor EF-Tu (duplicate of tufA)															
P0A6F5	b4143	groL	Cpn60 chaperonin GroEL, large subunit of GroESL															

### Transport

P06971	b0150	fhuA	ferrichrome outer membrane transporter															
P28635	b0197	metQ	DL-methionine transporter subunit; periplasmic-binding component of ABC superfamily															
P05825	b0584	fepA	iron-enterobactin outer membrane transporter															
P02931	b0929	ompF	outer membrane porin 1a (Ia,b;F)															
P0AEE5	b2150	mgIB	methyl-galactoside transporter subunit; periplasmic-binding component of ABC superfamily															
P17315	b2155	cirA	ferric iron-catecholate outer membrane transporter															
P06996	b2215	ompC	outer membrane porin protein C															

P0AEU0	b2309	hisJ	histidine/lysine/arginine/ornithine transporter subunit; periplasmic-binding component of ABC superfamily	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
P09551	b2310	argT	lysine/arginine/ornithine transporter subunit; periplasmic-binding component of ABC superfamily	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
P69783	b2417	crr	glucose-specific enzyme IIA component of PTS	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
P0AD96	b3460	livJ	leucine/isoleucine/valine transporter subunit; periplasmic-binding component of ABC superfamily	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
P0ABB4	b3732	atpD	F1 sector of membrane-bound ATP synthase, beta subunit	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
P0ABB4	b3732	atpD	F1 sector of membrane-bound ATP synthase, beta subunit	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
P37902	b0655	gltI	glutamate and aspartate transporter subunit; periplasmic-binding component of ABC superfamily																											

### Cell processes

P0A6Y8	b0014	dnaK	chaperone Hsp70, co-chaperone with DnaJ	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
P0A6Y8	b0014	dnaK	chaperone Hsp70, co-chaperone with DnaJ	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
P05825	b0584	fepA	iron-enterobactin outer membrane transporter	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
P0AE08	b0605	ahpC	alkyl hydroperoxide reductase, C22 subunit	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

P75780	b0805	fiu	predicted iron outer membrane transporter										
P0A862	b1324	tpx	lipid hydroperoxide peroxidase										
P0A6N1	b3339	tufA	protein chain elongation factor EF-Tu (duplicate of tufB)										
P0A6N1	b3980	tufB	protein chain elongation factor EF-Tu (duplicate of tufA)										
P0A6F5	b4143	groL	Cpn60 chaperonin GroEL, large subunit of GroESL										
P09147	b0759	galE	UDP-galactose-4-epimerase										

**Cell structure**

P06971	b0150	fhuA	ferrichrome outer membrane transporter										
P0A940	b0177	yaeT	conserved protein										
P05825	b0584	fepA	iron-enterobactin outer membrane transporter										
P09147	b0759	galE	UDP-galactose-4-epimerase										
P75780	b0805	fiu	predicted iron outer membrane transporter										

P02931	b0929	ompF	outer membrane porin 1a (la;b:F)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
P17315	b2155	cirA	ferric iron-catecholate outer membrane transporter	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
P06996	b2215	ompC	outer membrane porin protein C	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
P0ABB4	b3732	atpD	F1 sector of membrane-bound ATP synthase, beta subunit	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
P0ABB4	b3732	atpD	F1 sector of membrane-bound ATP synthase, beta subunit	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

### Extrachromosomal

P06971	b0150	fhuA	ferrichrome outer membrane transporter	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
P05825	b0584	fepA	iron-enterobactin outer membrane transporter	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
P17315	b2155	cirA	ferric iron-catecholate outer membrane transporter	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

---

# ABBREVIATIONS

---

2D-PAGE	-	Two-dimensional-polyacrylamide gel electrophoresis
CAS	-	Casamino acids
CHAPS	-	3-([3-chloramidopropyl]dimethylammonio)-1-propane-sulfonate
DDT	-	Dithiothreitol
DH	-	DH10B strain grown in glucose
DHAdp	-	DH10B strain adapted about 1000 generations in glucose
DHAdpGal	-	DHAdp (glucose evolved strain) grown in galactose
DHGal	-	DH10B strain grown in galactose
DHStat	-	DH10B strain grown in prolonged stationary phase (37 days)
FMN	-	Flavin mononucleotide
GASP	-	Growth advantage in stationary phase
GC-MS	-	Gas chromatography - mass spectroscopy
GO	-	Gene Ontology
IPG	-	Immobilized pH gradient
IPTG	-	Isopropyl $\beta$ -D-1-thiogalactopyranoside
kDa	-	Kilo dalton
l	-	Liter
LC-MS	-	Liquid chromatography - mass spectroscopy
LPS	-	Lipopolysaccharide
M	-	Molar
m	-	Milli ( $10^{-3}$ )
MALDI-TOF-MS	-	Matrix-Assisted Laser Desorption/Ionisation Time-Of-Flight Mass Spectrometry
MG	-	MG1655 strain grown in glucose
MGAdp	-	MG1655 strain adapted about 1000 generations in glucose
MGAdpGal	-	MGAdp (glucose evolved strain) grown in galactose
MGGal	-	MG1655 strain grown in galactose
MGStat	-	MG1655 strain grown in prolonged stationary phase (37 days)
MS	-	Mass spectroscopy
MW	-	Molecular weight
NADH	-	Nicotinamide adenine dinucleotide - reduced
NADPH	-	Nicotinamide adenine dinucleotide phosphate - reduced
NMR	-	Nuclear magnetic resonance
OD	-	Optical density
OM	-	Outer-membrane
ORF	-	Open reading frame
PCA	-	Principal components analysis
PCC	-	Pearson correlation coefficient
PCR	-	Polymerase chain reaction
pI	-	Isoelectric point
ppGpp	-	Guanosine-tetraphosphate
RT-PCR	-	Reverse transcription polymerase chain reaction
SDS-PAGE	-	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
TCA	-	Tricarboxylic acid cycle
TFA	-	Trifluoroacetic acid
$\mu$	-	Micro ( $10^{-6}$ )

---

## ACKNOWLEDGMENTS

---

I would like to express my gratitude to **Prof. Erwin Flaschel** for providing me the opportunity to work under his guidance in his department. Special thanks to him for providing enormous liberty and constant encouragement to try new experiments throughout my study, which gave me immense pleasure to work in numerous occasions. I would like to thank **Prof. Karsten Niehaus** for critical scientific discussions and for being my second supervisor. Furthermore, I thank **Dr. Karl Friehs** for his support and fruitful discussions.

Special thanks to **Eberhard Wunsch, Thomas Schaffer, Manuela Meyer, and Carola Eck** for their technical assistance.

My very special thanks to **Rashmi** for critically reading my thesis in detail, without her help it would have been difficult to be in a place to write this acknowledgment at this moment. Thanks a lot.

I am grateful for **Rileen** for proofreading my thesis. I thank **Subbu** for being a good friend and providing chapati's :) during crucial time of my graduation, it made me to spend more time in work rather than in the kitchen. I like to thank **Vikram** for being a good colleague and a friend. Also I am very grateful for having wonderful friends like you - **Shyam** and **Antony**.

This thesis is dedicated to my grandfather, **Mr. M.K. Bellie**. Being one of my greatest role model for various reasons without which none of this would have been even possible.

I thank and recognize that this research would not have been possible without the financial assistance from the *International NRW Graduate School in Bioinformatics and Genome Research*.

And last but not least, I would like to thank my family for both giving and encouraging me to seek for myself a demanding and meaningful education. This study could not have taken place without that precious gift.