

Metabolic Network Analysis in the Post-Genomic Era

Muhammad Rizwan Riaz

(2015-14-0009)

PhD Thesis

Advisor

Dr. Aziz Mithani



Department of Biology

Syed Babar Ali School of Science and Engineering

Lahore University of Management Sciences

Metabolic Network Analysis in the Post-Genomic Era

**Syed Babar Ali School of Science and Engineering
Lahore University of Management Sciences**

**Submitted in Partial Fulfillment of the Requirements
for the degree of Doctor of Philosophy in Biology**

by

Muhammad Rizwan Riaz

2015-14-0009

2020

Declaration

The work described in this thesis was carried out in the period from September 2015 to July 2020 at Lahore University of Management Sciences under the supervision of Dr. Aziz Mithani at the Department of Biology. I hereby, declare that this submission is my own work and that, to the best of my knowledge and belief, contains no material previously published or written by another person neither has it been accepted for the award of any other degree or diploma at a university or any other institute of higher learning, except where due acknowledgement has been made in the text.

Muhammad Rizwan Riaz

Title

Metabolic network analysis in the post-genomic era

Submitted by: Muhammad Rizwan Riaz

Advisor: Aziz Mithani

Acknowledgments

All praises, thanks, and acknowledgments to **Almighty ALLAH**, The Most Merciful and The Most Beneficent, the entire source of knowledge and wisdom endowed to the humankind, for enabling me to carry out this humble research work. All praises to Holy Prophet **Muhammad (Peace Be Upon Him)**, a torch of guidance and knowledge forever for humanity as a whole.

I owe the utmost love and respect to my beloved parents, **Mr. and Mrs. Riaz-ul-Hassan**, for their kindness, tireless devotion, and for encouraging me towards the studies and knowledge. I am thankful to my brother and sisters, who gently offer counsel and unconditional support at each turn of the road. I offer my sincerest gratitude to **Dr. Aziz Mithani**, Associate Professor, Department of Biology, Lahore University of Management Sciences, Lahore, for his advice throughout my research work. His observations and comments helped me to establish the overall direction of the research and to move forward with the investigation in depth. This thesis would not have been possible without his keen interest and expert knowledge. I just cannot thank him aptly for his unconditional support. I am highly thankful to **Dr. Muhammad Tariq** (Associate Professor) and **Dr. Safee Ullah Chaudhary** (Associate Professor), my thesis committee members, for their guidance, teachings, and support. I would also like to acknowledge **all my teachers** for their support and guidance throughout my academic life.

I would like to thank my friends and colleagues who supported me directly or indirectly through this journey. I would also like to acknowledge past and present members of Computational Genomics & Systems Biology Lab for productive discussions, countless

cups of tea and a wonderful time. I heartily appreciate my friends Muhammad Asaad Usmani, Rashid Hussain and Rahim Ullah Yousafzai, for the encouragement and support throughout my PhD. Special thanks to Ms. Iqra Manzoor for taking care of the administrative stuff, and I also appreciate the support of research and academic staff of Department of Biology, LUMS. I am thankful to the Higher Education Commission of Pakistan and Lahore University of Management Sciences for funding and support. Finally, I would like to thank my wife, Anam Fatima, for supporting me during the compilation of this thesis.

Muhammad Rizwan Riaz

Abstract

Metabolic networks are intricate systems comprising of interconnected biochemical reactions transforming source metabolites into target metabolites. This thesis presents a web-based tool called MAPPS: Metabolic network Analysis and Pathway Prediction Server (<https://mapps.lums.edu.pk>), for the prediction of metabolic pathways and comparisons of metabolic networks using traditional and ‘omics datasets. MAPPS provides an intuitive approach to answer biological questions focusing on the metabolic capabilities of an organism as well as differences between organisms or the evolution of different species by allowing pathway-based metabolic network comparisons at an organism as well as at a phylogenetic level. MAPPS also allows users to study the behavior of engineered metabolic networks and effects of metabolic availability/unavailability on metabolic pathways, identify potential drug targets, study host-microbe interactions, and build ancestral networks over a given phylogeny. MAPPS is used to understand the metabolic diversity and functional specialization in different strains of the bacteria belonging to genus *Pseudomonas* by performing whole-network and pathway-based comparisons relating to carbohydrate and energy metabolisms. Results suggest that pseudomonads with similar lifestyle tend to have a high degree of metabolic similarity and that species have adapted their metabolic networks to suit their diverse lifestyles. Finally, this thesis explores the changes occurring in the metabolic networks of two mango (*Mangifera indica*) cultivars, ‘Sindhri’ and ‘Kala Chaunsa’ during fruit maturation. For this, metabolic maps of various KEGG pathway maps are developed by assigning metabolic annotations to a mango transcriptomic reference, which are further used to analyze metabolic pathways differentially expressed between immature and mature stages in the two cultivars by

identifying differentially expressed genes. Results suggest that carbohydrates, lipids and amino acids, and secondary metabolite pathways are differentially expressed in both cultivars, demonstrating the use of ‘omic data for better understanding of metabolic networks in today’s post-genomic era.

Table of Contents

1	Introduction	1
1.1	Analysis of metabolic networks	2
1.2	Resources for metabolic network analysis	7
1.2.1	Public databases for metabolic network analysis.....	7
1.2.2	Tools for the analysis of metabolic networks	11
1.3	Metabolic network analysis in post-genomic era.....	13
1.4	Overview of the thesis.....	14
2	MAPPS: A Web-Based Tool for Metabolic Pathway Prediction and Network Analysis in the Post-Genomic Era	16
2.1	Rahnuma: hypergraph-based tool for metabolic pathway prediction and network comparison.....	21
2.2	MAPPS: Metabolic Network Analysis and Pathway Prediction Server	26
2.2.1	MAPPS architecture.....	27
2.2.2	Working of the tool.....	31
2.2.3	MAPPS output	42
2.2.4	Analyses provided by MAPPS.....	45
2.2.5	<i>In silico</i> metabolic engineering.....	62
2.2.6	Network filtering using ‘omics data	65
2.2.7	Pathway ranking.....	67

2.3	Applications of MAPPS	68
2.3.1	Predicting biologically meaningful metabolic pathways by tracing specific elements	69
2.3.2	Studying the effects of <i>in silico</i> metabolic knock-out mutation	73
2.3.3	Designing heterologous pathways using <i>in silico</i> metabolic knock-in mutations	76
2.3.4	Identification of potential drug targets.....	78
2.3.5	Studying emergent pathways resulting from host-microbe interaction	81
2.4	Discussion	83
3	Metabolic Diversity and Functional Specialization in <i>Pseudomonas</i>	87
3.1	<i>Pseudomonas</i> dataset.....	90
3.2	Comparison of metabolic networks in <i>Pseudomonas</i>	94
3.2.1	Carbohydrate and amino acid metabolisms	95
3.2.2	Energy metabolism	99
3.3	Deciphering the patterns of amino acid assimilation pathways in <i>Pseudomonas</i>	104
3.3.1	Pathway prediction from amino acids to ammonia.....	106
3.3.2	Pathway prediction from amino acids to TCA cycle intermediates	109
3.4	16S rRNA and Multilocus Sequence Analysis	113
3.5	Metabolic Similarity Analysis.....	120

3.6	Discussion	127
4	Differential Metabolic Analysis of Mango between Immature and Mature Stages.....	131
4.1	RNA-seq data for mango.....	134
4.2	Metabolic annotation of mango genes	135
4.3	Identification of genes differential expressed during fruit maturation.....	139
4.4	Assignment of genes differentially expressed during ripening on KEGG pathways	141
4.4.1	Metabolic mapping of differentially expressed genes in ‘Kala Chaunsa’ .	141
4.4.2	Metabolic mapping of differentially expressed genes in ‘Sindhri’	144
4.5	Discussion	162
5	Discussion	166
6	References	175
7	Appendices	193

List of Figures

Figure 1.1: An example of metabolic network.	3
Figure 1.2: The image shows KEGG reference map for glycolysis.	4
Figure 1.3: Genome scale metabolic network of <i>Clamydomonas reinhardtii</i>	5
Figure 1.4: Screenshot of KEGG webpage.....	8
Figure 1.5: Screenshot of BioCyc database.	9
Figure 2.1: <i>Rahnuma</i> interface for job submission.	22
Figure 2.2: An example pathway to demonstrate limitation of pathway prediction algorithm of <i>Rahnuma</i>	24
Figure 2.3: Plain text output of pathway-based comparison of <i>Rahnuma</i>	25
Figure 2.4: Screenshot of <i>Rahnuma</i> interface to select organisms and metabolites.	26
Figure 2.5: Overview of MAPPS.	28
Figure 2.6: MAPPS architecture.	29
Figure 2.7: MAPPS login page.	32
Figure 2.8: MAPPS interface for user account registration.	32
Figure 2.9: MAPPS interface to reload previously saved jobs.	33
Figure 2.10: Flowchart describing MAPPS workflow.	34
Figure 2.11: MAPPS graphical user interface for job submission.....	35
Figure 2.12: Building metabolic networks by specifying organism(s) using search box or taxonomy viewer.....	36

Figure 2.13: Specifying one or more organism set(s) to build a metabolic network(s) in MAPPS.	37
Figure 2.14: Building metabolic network over a phylogeny.	38
Figure 2.15: MAPPS interface for specifying pathway-related parameters.	39
Figure 2.16: Review and submit the job.	40
Figure 2.17: MAPPS interface showing list of completed jobs for registered user.	41
Figure 2.18: MAPPS interface showing the job status for guest users.	42
Figure 2.19: An example of the interactive graphical output for pathway prediction.	43
Figure 2.20: An example of plain text output for pathway prediction of MAPPS.	43
Figure 2.21: An example of HTML output for pathway prediction.	44
Figure 2.22: An example of SBML output for pathway prediction.	45
Figure 2.23: Graphical output for metabolic reachability.	50
Figure 2.24: MAPPS interface for entering job parameters for reachability analysis.	51
Figure 2.25: MAPPS interface for ancestral network building.	54
Figure 2.26: MAPPS interface for network enumeration/comparison.	55
Figure 2.27: MAPPS interface for metabolic similarity analysis.	57
Figure 2.28: An example dendrogram resulting from metabolic similarity analysis between three organisms.	58
Figure 2.29: MAPPS job submission interface for host-microbe interaction.	59

Figure 2.30: Graphical output showing pathways resulting due to host-microbe interactions.....	61
Figure 2.31: MAPPS job submission interface for estimate evolution parameters analysis.	63
Figure 2.32: Performing <i>in silico</i> metabolic engineering in MAPPS.	64
Figure 2.33: Network filtering using ‘omics data.	66
Figure 2.34: An example of pathway score calculation using reaction thermodynamics and structural similarity in MAPPS.....	68
Figure 2.35: KEGG pathway map of sulfur metabolism.	69
Figure 2.36: Sulfate assimilation pathway in <i>Arabidopsis thaliana</i>	70
Figure 2.37: An alternate pathway from sulfate to L-cysteine predicted by MAPPS.	71
Figure 2.38: Examples of incorrect pathways predicted by FMM from sulfate to L-cysteine.	72
Figure 2.39: Leloir pathway in humans metabolizing α -D-galactose to produce UDP-glucose.	74
Figure 2.40: Effect of <i>in silico</i> knockouts on predicted pathways between α -D-galactose and UDP-glucose in human galactose metabolic network.	75
Figure 2.41: A heterologous pathway for flavonoid production in <i>Escherichia coli</i>	76
Figure 2.42: Comparing <i>Escherichia coli</i> metabolic network with and without <i>in silico</i> knock-in mutations.....	77
Figure 2.43: An application of potential drug target identification in MAPPS.	79

Figure 2.44: Identification of Enzyme II ^{Glc} (EC 2.7.1.11) as a potential drug target in <i>Salmonella enterica</i> serovar Typhimurium.	80
Figure 2.45: Prediction of emergent pathways arising due to the interaction between <i>Rickettsia parkeri</i> and humans.	82
Figure 3.1: Number of reactions involved in carbohydrate and amino acid metabolisms predicted to be present in <i>Pseudomonas</i>	97
Figure 3.2: Distribution of KEGG reactions involved in carbohydrate and amino acid metabolism across different <i>Pseudomonas</i> strains.	98
Figure 3.3: Distribution of KEGG reactions involved in energy metabolism across different <i>Pseudomonas</i> strains.	104
Figure 3.4: Number of pathways predicted in different pseudomonads from amino acids to ammonia.	108
Figure 3.5: KEGG map of Tricarboxylic Acid (TCA) cycle.	110
Figure 3.6: Number of pathways predicted in different pseudomonads from amino acids to TCA Cycle intermediates.	112
Figure 3.7: A schematic diagram showing concatenation of housekeeping gene sequences for MLSA.	114
Figure 3.8: Phylogenetic tree of the 111 pseudomonads constructed using 16S rRNA sequences.	116
Figure 3.9: Phylogenetic tree of the 111 pseudomonads constructed using MLSA.	117

Figure 3.10: Enzyme-based metabolic similarity analysis of <i>Pseudomonas</i> using single linkage clustering.	121
Figure 3.11: Reaction-based metabolic similarity analysis of <i>Pseudomonas</i> using single linkage clustering.	122
Figure 3.12: Enzyme-based metabolic similarity analysis of <i>Pseudomonas</i> using average linkage clustering.	123
Figure 3.13: Reaction-based metabolic similarity analysis of <i>Pseudomonas</i> using average linkage clustering.	124
Figure 3.14: Enzyme-based metabolic similarity analysis of <i>Pseudomonas</i> using maximum linkage clustering.	125
Figure 3.15: Reaction-based metabolic similarity analysis of <i>Pseudomonas</i> using maximum linkage clustering.	126
Figure 4.1: Distribution of reference transcripts assigned to different KEGG pathway sets.	136
Figure 4.2: Mapping of reference transcripts classified according to BRITE hierarchy of protein families.	139
Figure 4.3: Number of differentially expressed genes identified during mango fruit development.	141
Figure 4.4: Distribution of differentially expressed genes of ‘Kala Chaunsa’ assigned to different KEGG pathway sets.	142

Figure 4.5: Number of genes differentially expressed during fruit maturation in various KEGG pathway maps in ‘Kala Chaunsa’.	143
Figure 4.6: Metabolic mapping of genes differentially expressed during fruit maturation on KEGG pathway map of Glycolysis.	146
Figure 4.7: Metabolic mapping of genes differentially expressed during fruit maturation on pyruvate metabolism.	147
Figure 4.8: Metabolic mapping of genes differentially expressed during fruit ripening on glutathione metabolism.	148
Figure 4.9: Metabolic mapping of genes differentially expressed during fruit maturation on starch and sucrose metabolism.	149
Figure 4.10: Metabolic mapping of genes differentially expressed during fruit maturation on arginine biosynthesis.	150
Figure 4.11: Metabolic mapping of genes differentially expressed during fruit maturation on nitrogen metabolism.	151
Figure 4.12: Metabolic mapping of genes differentially expressed during fruit maturation on brassinosteroid biosynthesis.	152
Figure 4.13: Metabolic mapping of genes differentially expressed during fruit maturation on phenylpropanoid biosynthesis.	153
Figure 4.14: Distribution of differentially expressed genes of ‘Sindhri’ assigned to KEGG pathway sets.	154

Figure 4.15: Number of genes differentially expressed during fruit maturation in various KEGG pathway maps in ‘Sindhri’	155
Figure 4.16: Metabolic mapping of genes differentially expressed during fruit maturation on flavonoid biosynthesis.	159
Figure 4.17: Metabolic mapping of genes differentially expressed during fruit maturation on carotenoid biosynthesis.	160
Figure 4.18: Metabolic mapping of genes differentially expressed during fruit maturation on riboflavin metabolism.	161
Figure 5.1: A schematic representation of coupled metabolic-regulatory network (Shlomi et al., 2007)	172

List of Tables

Table 2.1: Tools available for metabolic network analysis	19
Table 3.1: <i>Pseudomonas</i> strains used for metabolic network analysis	91
Table 3.2: List of pathway maps from KEGG relating to carbohydrate and amino acid metabolisms used for metabolic comparisons in <i>Pseudomonas</i>	96
Table 3.3: Reactions involved in carbohydrate and amino acid metabolisms predicted to be present in only one pseudomonad but absent from all the others	100
Table 3.4: List of pathway maps from KEGG relating to energy metabolism used for metabolic comparisons in <i>Pseudomonas</i>	102
Table 3.5: Reactions involved in from energy metabolism predicted to be present in only one pseudomonad but absent from the others	103
Table 3.6: Number of distinct pathways of various lengths predicted from amino acids to ammonia across all <i>Pseudomonas</i> strains	105
Table 3.7: Number of distinct pathways of various lengths predicted from amino acids to the TCA cycle intermediate	109
Table 4.1: Alignment mapping summary for ‘Sindhri’ and ‘Kala Chaunsa’ samples....	135
Table 4.2: Number of transcripts mapped to carbohydrate, amino acid, lipid and energy metabolisms	137
Table 4.3: Metabolic mapping of upregulated genes in ‘Kala Chaunsa’	144
Table 4.4: Metabolic mapping of downregulated genes in ‘Kala Chaunsa’	145

Table 4.6: Metabolic mapping of top 20 downregulated genes in ‘Sindhri’	157
Table 4.5: Metabolic mapping of top 20 upregulated genes in ‘Sindhri’	156

List of Pseudocodes

Box 2.1: Pseudocode for pathway prediction algorithm.....	49
Box 2.2: Pseudocode for metabolite reachability algorithm.....	52
Box 2.3: Pseudocode for identifying metabolite-specific reactions	53
Box 2.4: Pseudocode for host-microbe interaction analysis algorithm	60
Box 2.5: Pseudocode for potential drug-targets identification algorithm.....	62

List of Appendices

Appendix A: Database schema of MAPPS.....	193
Appendix B: List of ubiquitous metabolites ignored by default during pathway prediction in MAPPS	194
Appendix C: List of conserved reactions related to carbohydrate and amino acid metabolism.....	196
Appendix D: List of variable reactions of carbohydrate and amino acid metabolism.....	200
Appendix E: List of conserved reactions related to energy metabolism	221
Appendix F: List of variable reactions related to energy metabolism	222
Appendix G: Number of pathways predicted from Amino Acids to Ammonia in <i>Pseudomonas</i>	225
Appendix H: Number of unique pathways predicted from amino acids to Ammonia	233
Appendix I: Number of pathways predicted from amino acids to TCA Cycle intermediates in <i>Pseudomonas</i>	234
Appendix J: Number of unique pathways predicted from different amino acids to four metabolites of TCA Cycle in <i>Pseudomonas</i>	256
Appendix K: List of organisms used as representative set in KAAS for assigning metabolic annotations to mango genes	257
Appendix L: List of enzymes predicted to be upregulated in ‘Sindhri’	258
Appendix M: List of enzymes predicted to be downregulated in ‘Sindhri’	262

Publication

Riaz, M. R., Preston, G. M., & Mithani, A. (2020). MAPPS: A Web-Based Tool for Metabolic Pathway Prediction and Network Analysis in the Postgenomic Era. *ACS Synthetic Biology*, 9(5), 1069-1082. <https://doi.org/10.1021/acssynbio.9b00397>

1 Introduction

Metabolism is an essential process to sustain life. It is categorized into two parts; breaking down of food and other nutrients into small building blocks (catabolism) and combining them to form complex molecules (anabolism). This is achieved via a network of interconnected enzyme-catalyzed or spontaneous biochemical reactions, which is called a metabolic network (**Figure 1.1**) (Jeong et al., 2000). A biochemical reaction connects one or more metabolites called substrates, which combine to give one or more metabolites called products. A sequence of biochemical reactions in a metabolic network transforming a source metabolite into a target metabolite is known as a metabolic pathway (Jeong et al., 2000). Over the years, novel pathways have been deciphered using experimental protocols (Fell, 1992), which have provided insights into the metabolic capabilities of different organisms. These findings have helped to unravel the mystery that despite the presence of many possible paths from one metabolite to another, some organisms have evolved to favor a particular pathway to produce or consume a compound (Planes and Beasley, 2009). Studies have found that most organisms have a core set of enzymes that are involved in energy metabolism and catalyze essential processes such as protein synthesis and DNA replication. However, a significant proportion of the enzymes present in different organisms are specific to the needs of individual organisms or tissues (Dandekar et al., 1999; Smith and Morowitz, 2004; Mentzen et al., 2008).

Traditionally, metabolic networks have been studied by grouping the reactions into smaller networks based on their involvement in different biological processes. An example network corresponding to glycolysis in humans is shown in **Figure 1.2**. However, with the emphasis on studying networks as a whole has led to the development of genome-scale

metabolic networks (**Figure 1.3**). Genome-scale metabolic networks are beneficial to understand the complex network of metabolic reactions happening inside cells (Reed et al., 2003). A genome-scale metabolic network typically consists of thousands of reactions and metabolites, which capture vital metabolic pathways such as the energy metabolism, biosynthesis of amino acids and lipids, as well as transport of molecules inside the cells (Oberhardt et al., 2009). In recent years, the construction of such networks has increased tremendously, in part also due to the growing number of genomes sequenced (Monk et al., 2014), providing insights into the metabolic activities of various organisms. This is aided by the availability of functional annotations and transcriptomic, proteomic and metabolomic ('omic) datasets, which facilitate in providing a better understanding of the metabolic capabilities of organisms at a systems level.

1.1 Analysis of metabolic networks

Comparative and evolutionary analyses of metabolic networks have a broad range of applications, ranging from research into metabolic evolution through to practical applications in drug development, synthetic biology, and biodegradation. Comparative analyses of metabolic networks, for example, can be used to provide insight into the evolution of metabolic pathways, the environments that organisms occupy, and the selective pressures that have shaped their evolution.

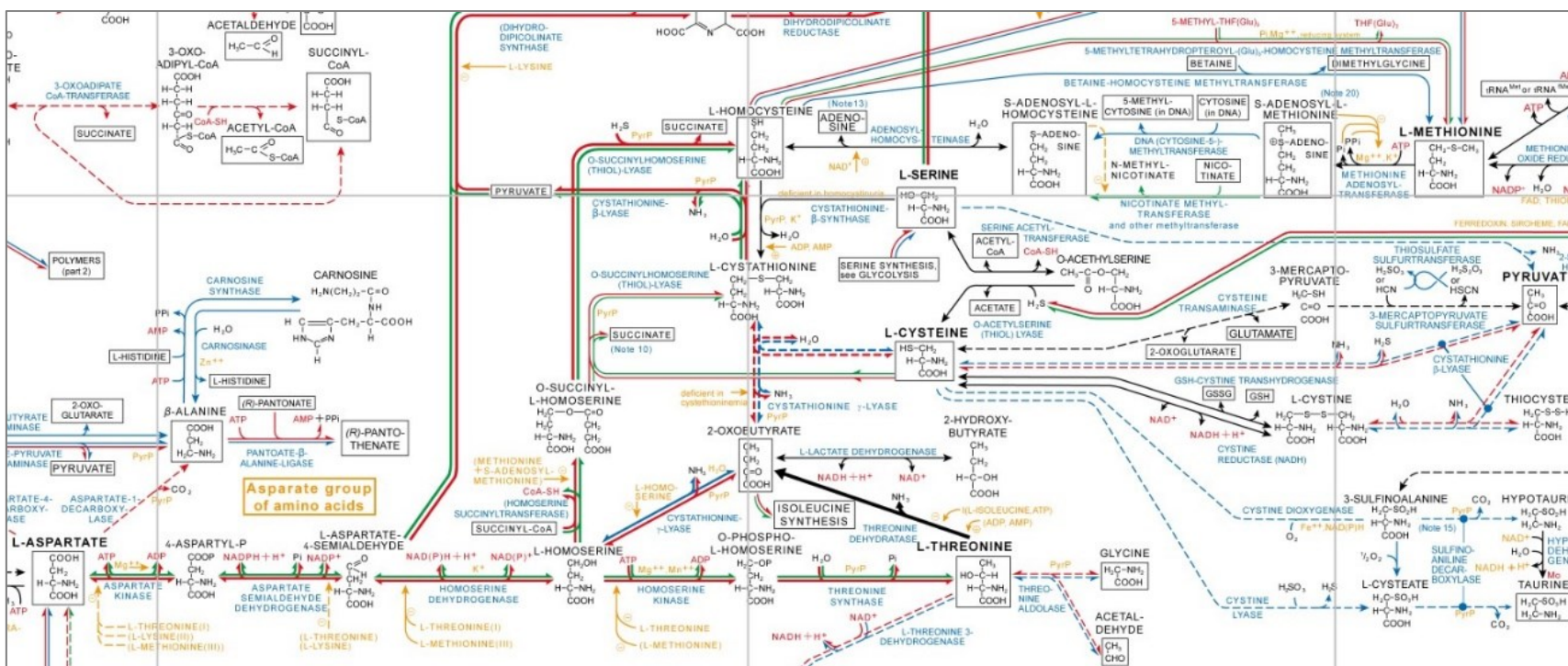


Figure 1.1: An example of metabolic network. A metabolic network consists of enzymatic as well as spontaneous reactions which convert source metabolites into target metabolites. Image taken from Roche biochemical pathways (Lee, 2012).

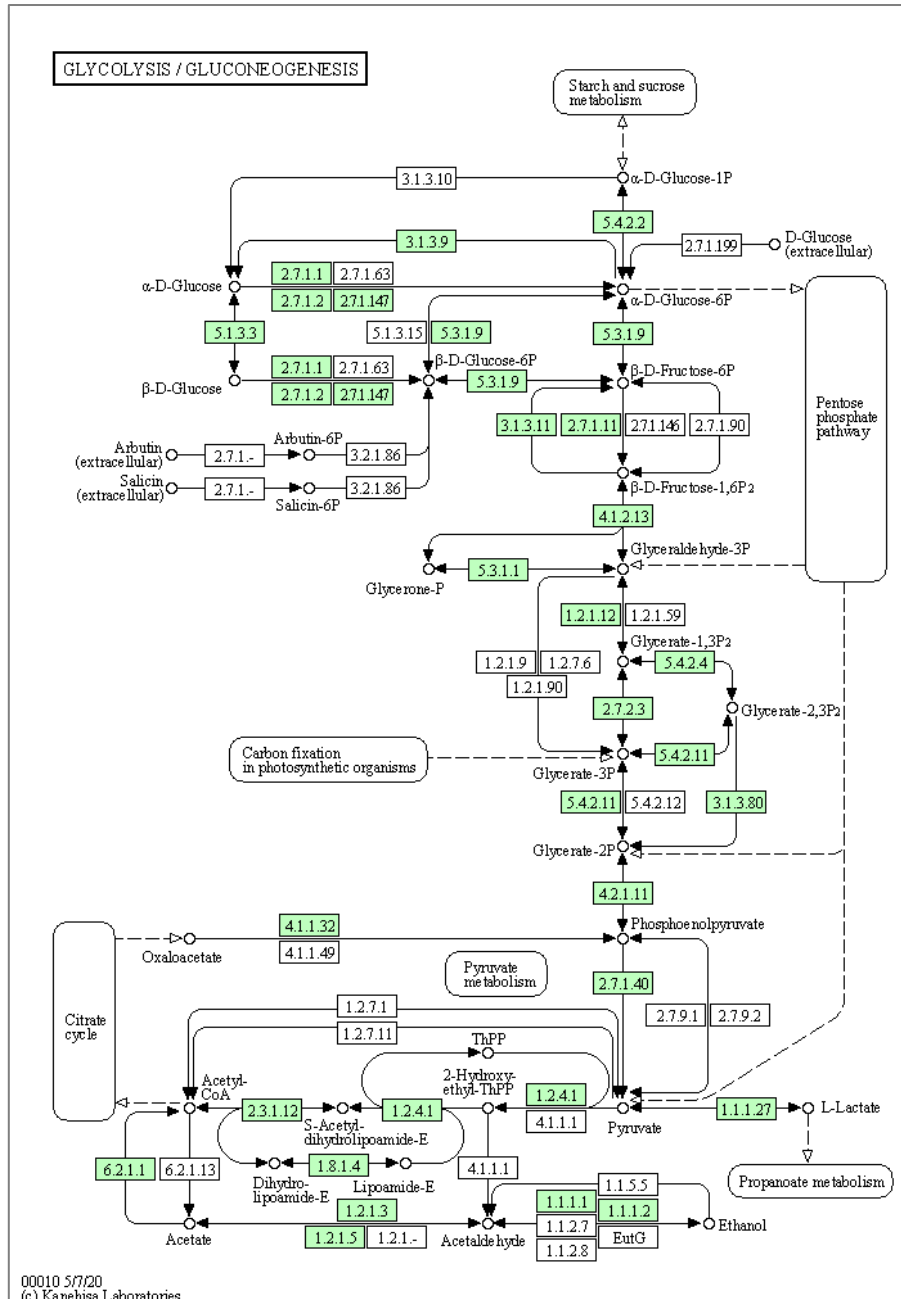


Figure 1.2: The image shows KEGG reference map for glycolysis. Enzymes present in the human metabolic network are highlighted in green.

Besides this, by performing *in silico* analysis of metabolic networks based on transcriptomic data or *in silico* insertion and knock out experiments, it might be possible to develop testable hypotheses that explain experimental results or stimulate further investigation, thereby providing an insight into the metabolic functionality of organisms.

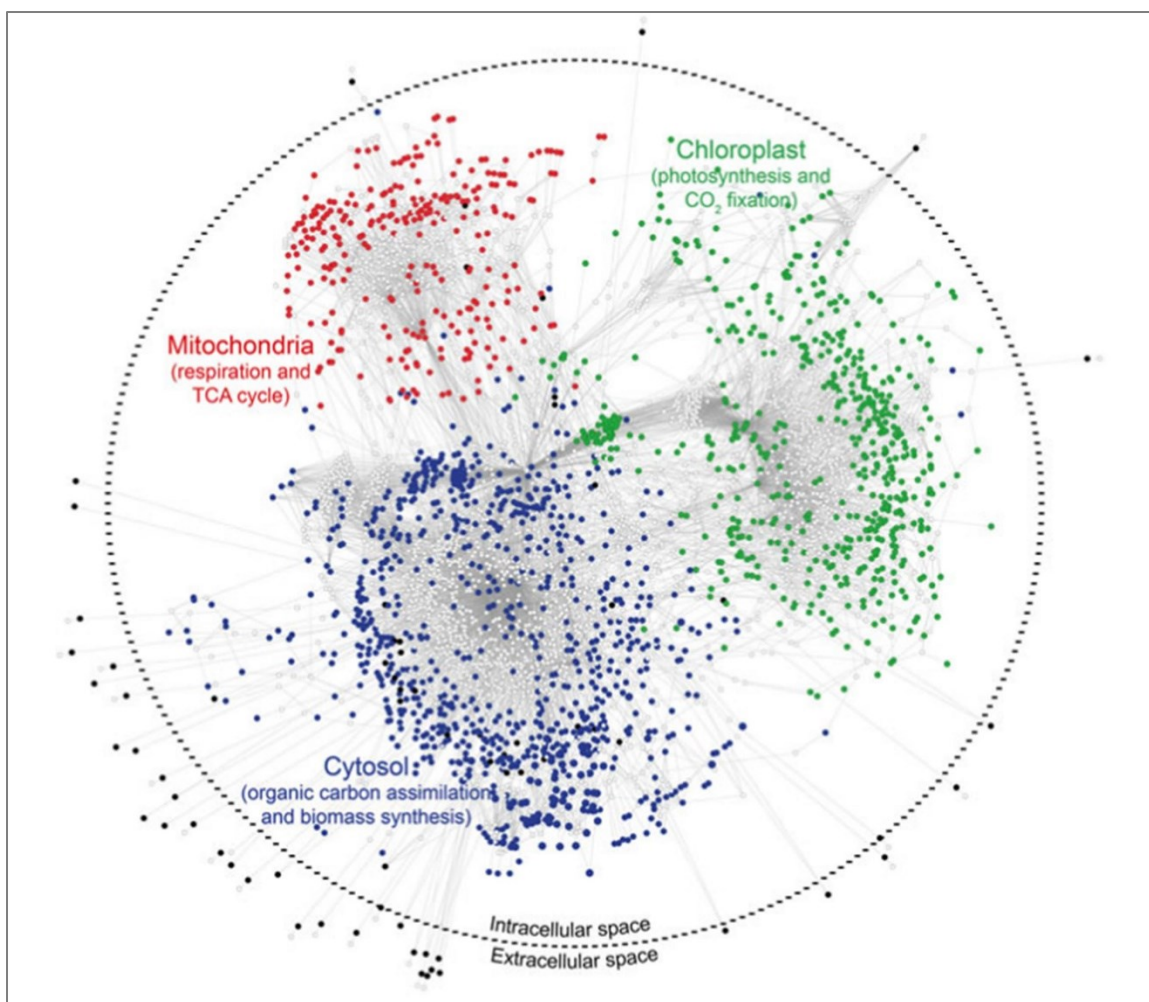


Figure 1.3: Genome scale metabolic network of *Clamydomonas reinhardtii*. Metabolites are divided into different compartments i.e., mitochondria (red), cytosol (blue), chloroplast (green) and extracellular (black). Reactions are shown in grey color. Image taken from (Perez-Garcia et al., 2016).

Several studies have used data from publicly available data repositories to compare metabolic networks across genomes and to investigate biological questions (Ma and Zeng 2003; Light et al. 2005; Forst et al. 2006; Wang et al. 2006). However, due to the complex nature of metabolic networks and the unavailability of suitable tools for the analyses, most comparative analyses have been restricted to analyzing and summarizing network properties such as degree distribution, clustering coefficient, and average path length, or comparing small, well-characterized regions within metabolic networks.

The graphical representation of metabolic network also provides a general framework to perform a variety of specialized analyses. Some of these include prediction of chokepoints for drug-target identification, which identifies the reactions which either consumes a unique substrate or produce a unique product and, therefore, can be used as potential drug targets (Taylor et al., 2013), prediction of the metabolic state of the pathogen during infection of the host by generating a combined host-pathogen metabolic model and adjusting the boundary conditions based on experimental data (Rienksma et al., 2019), understanding cooperative and competitive relationships between microbial species by analyzing the metabolic exchange and biosynthetic capabilities of each microbial species in the microenvironment (Ponomarova and Patil, 2015), and producing beneficial compounds in the microbes with the help of ‘omics data, *in silico* gene knockout/knock-in strategies, pathway prediction and enzyme engineering for metabolic networks (Chae et al., 2017).

Another area that has gained interest in recent years is the study of metabolic evolution. Metabolic networks, like all other biological networks, are under a process of continuous evolution. However, evolutionary mechanisms are not yet known. It is unclear how these networks evolve and if there is a correlation between the evolution of metabolic capabilities and factors such as the network structure or the environment in which the organisms live. The availability of genomes for many closely related species offers the possibility of an in-depth analysis of metabolic pathways within an organism as well as between different organisms to comprehend the processes and attributes that affect the evolution of metabolic networks (Kreimer et al., 2008; Mithani et al., 2010). It is now possible to study how genomic events such as gene duplications may increase the metabolic

flux in the corresponding part of the network due to the availability of extra copy of the enzyme-coding gene, or how the evolution of an enzyme to catalyze a new substrate causes the edges in the metabolic network to be rewired (Chae et al., 2012). Thus, evolutionary analysis of metabolic networks can provide insights into the evolution of metabolic pathways, the environments that organisms occupy and the selective pressures that have shaped their evolution.

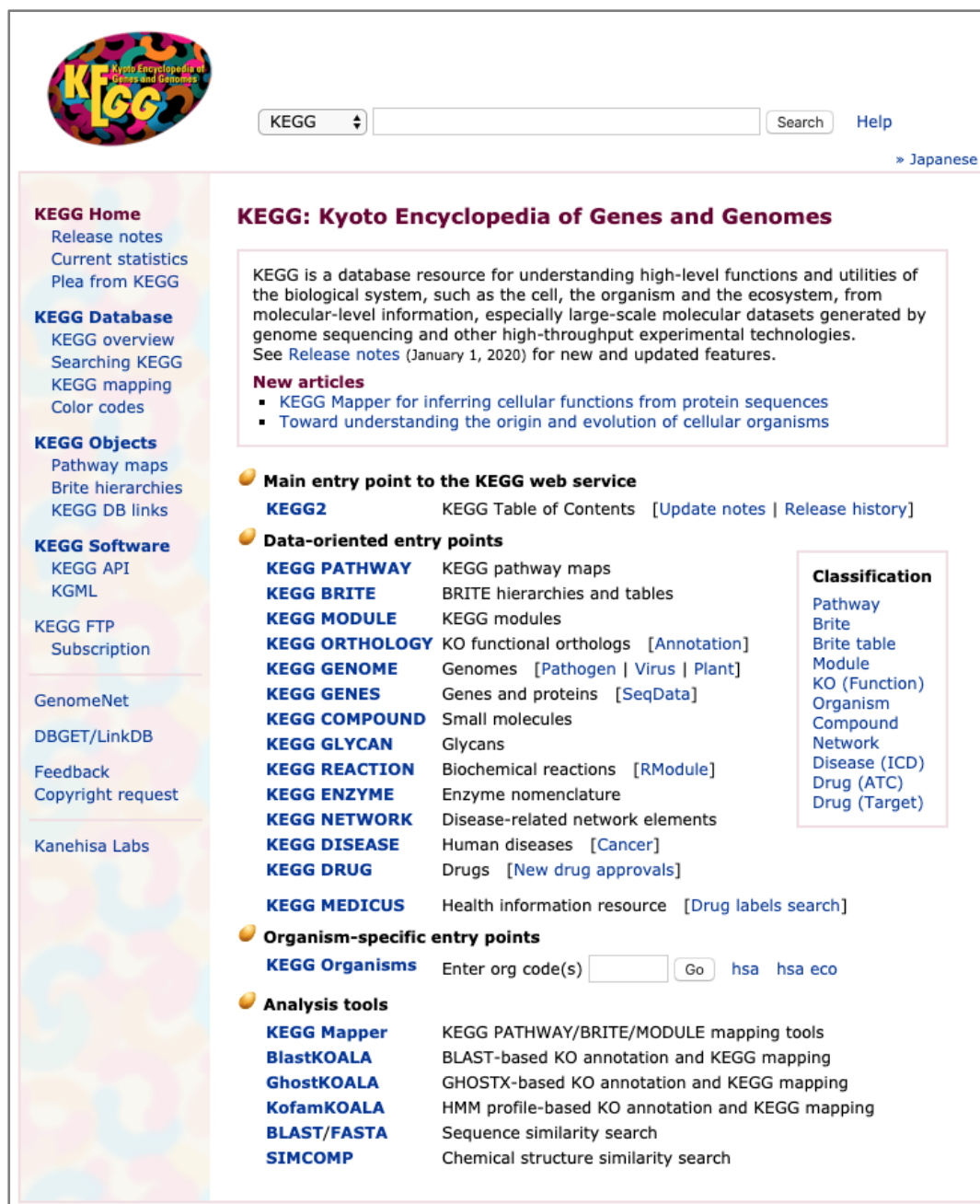
1.2 Resources for metabolic network analysis

Over the years, researchers have developed comprehensive databases that catalog genes, enzymes, reactions, and pathways that are predicted to be present in the genome-sequenced organisms and have developed tools for mining these datasets and predicting possible pathways for assimilation, synthesis and transformation of biological molecules (Jing et al., 2014). These are discussed in subsequent subsections.

1.2.1 Public databases for metabolic network analysis

A number of public databases exist for metabolic networks, which contain curated/non-curved information on metabolic pathways, enzymes, and metabolites and are a very useful resource in studying metabolic networks. Some of the most popular databases are discussed below.

Kyoto Encyclopedia of Genes and Genomes (KEGG), one of the most popular resources for metabolic networks, is a reference knowledge base and is widely used for analyzing not only genomics data but also transcriptomics, proteomics, glycomics, metabolomics, metagenomics, and other high-throughput data (Kanehisa et al., 2017). It facilitates researchers to study the conservation and variation of genes and genomes by



KEGG Home
 Release notes
 Current statistics
 Plea from KEGG

KEGG Database
 KEGG overview
 Searching KEGG
 KEGG mapping
 Color codes

KEGG Objects
 Pathway maps
 Brite hierarchies
 KEGG DB links

KEGG Software
 KEGG API
 KGML
 KEGG FTP
 Subscription

GenomeNet
 DBGET/LinkDB
 Feedback
 Copyright request

Kanehisa Labs

KEGG: Kyoto Encyclopedia of Genes and Genomes

KEGG is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism and the ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies. See [Release notes](#) (January 1, 2020) for new and updated features.

New articles

- KEGG Mapper for inferring cellular functions from protein sequences
- Toward understanding the origin and evolution of cellular organisms

Main entry point to the KEGG web service
KEGG2 KEGG Table of Contents [[Update notes](#) | [Release history](#)]

Data-oriented entry points

KEGG PATHWAY	KEGG pathway maps
KEGG BRITE	BRITE hierarchies and tables
KEGG MODULE	KEGG modules
KEGG ORTHOLOGY	KO functional orthologs [Annotation]
KEGG GENOME	Genomes [Pathogen Virus Plant]
KEGG GENES	Genes and proteins [SeqData]
KEGG COMPOUND	Small molecules
KEGG GLYCAN	Glycans
KEGG REACTION	Biochemical reactions [RModule]
KEGG ENZYME	Enzyme nomenclature
KEGG NETWORK	Disease-related network elements
KEGG DISEASE	Human diseases [Cancer]
KEGG DRUG	Drugs [New drug approvals]
KEGG MEDICUS	Health information resource [Drug labels search]

Organism-specific entry points
KEGG Organisms Enter org code(s) [hsa](#) [hsa eco](#)

Analysis tools

KEGG Mapper	KEGG PATHWAY/BRITE/MODULE mapping tools
BlastKOALA	BLAST-based KO annotation and KEGG mapping
GhostKOALA	GHOSTX-based KO annotation and KEGG mapping
KofamKOALA	HMM profile-based KO annotation and KEGG mapping
BLAST/FASTA	Sequence similarity search
SIMCOMP	Chemical structure similarity search

Classification
 Pathway
 Brite
 Brite table
 Module
 KO (Function)
 Organism
 Compound
 Network
 Disease (ICD)
 Drug (ATC)
 Drug (Target)

Figure 1.4: Screenshot of KEGG webpage. Main page of KEGG shows list of databases and tools.

complementing the experimental protocols (**Figure 1.4**). Having started in 1995 with only four databases which included KEGG Pathway containing information of manually drawn pathway maps, KEGG Genes containing gene information collected from publicly available databases, KEGG Enzyme containing information of Enzyme Classification and

BioCyc Database Collection

BioCyc is a collection of 16,822 Pathway/Genome Databases (PGDBs), plus software tools for exploring them [Karp17]. Key aspects of BioCyc data:

- Quality data curated from tens of thousands of publications, including curated databases for *E. coli*, *B. subtilis*, *H. sapiens*, and *S. cerevisiae*.
- Computationally predicted metabolic pathways and operons.
- Data integrated from other databases including gene essentiality, regulatory networks, protein features, and GO annotations.

Current version: 24.0, released on May 14, 2020. [Update History](#)

Subscriptions are required to access most of BioCyc beyond a limited period of free use. For more information on obtaining a subscription, click [here](#). Subscriptions support curation of BioCyc content on a non-profit basis. You do not need a subscription to access EcoCyc or MetaCyc, or if your institution already subscribes (institution name is in the top right corner of this page, or appears [here](#).)

New to BioCyc? Typical usage is:

- Select a database (genome) to search by clicking "change organism database" in the box in the top right of every page.
- Search for a gene, pathway, metabolite, or enzyme using the Quick Search in that same box... or see the Search menu for more options.

Explore Metabolic Maps for Thousands of Organisms
The Cellular Overview enables you to browse through zoomable metabolic map diagrams that are customized to each BioCyc organism.

[Learn More](#)

paused... 1 2 3 4 5 6 7 8 9 10

Figure 1.5: Screenshot of BioCyc database. User can search for metabolic pathway, enzyme, reaction, or metabolite. It also allows users to search in organism specific databases called Pathway/Genome Databases (PGDBs)

relation hierarchy for reactions and KEGG Compound consisting of chemical information of compounds including chemical structures, KEGG now has eighteen databases which contain a diverse set of information including genomes, orthologs, glycans, drugs, and health and disease-related information (Kanehisa et al., 2019) (**Figure 1.4**). Besides this, KEGG also provides different tools to analyze data available in the KEGG databases or user-provided inputs. These include PathPred for computing pathways between metabolites (Moriya et al., 2010), SIMCOMP/SUBCOMP for chemical similarity and substructure search (Hattori et al., 2010) and KEGG Automated Annotation Server for metabolic annotations of draft and incomplete genomes (Moriya et al., 2007).

Another widely used resource for metabolic networks is BioCyc Database Collection (**Figure 1.5**). It is a collection of 16,822 Pathway/Genome Databases (PGDBs) in addition to different software tools for exploring them. Each PGDB contain the full

genome and predicted metabolic network of one organism, including metabolites, enzymes, reactions, and metabolic pathways (Karp et al., 2018). Unlike KEGG which groups the reactions into traditional pathway maps based on their involvement in different biological processes, BioCyc provides information at individual pathway level. Besides this, BioCyc databases are categorized based on quality. Databases having the most accurate curated data with experimental support and manual curations are present in the Tier 1 while Tiers 2 and 3 consists of databases which have been predicted using different computational methods (Mishra et al., 2019). MetaCyc, a curated database of small-molecule metabolism, which is part of the BioCyc Database Collection, is designated as the reference database. It is also used by the Pathway Tools software, available as a part of BioCyc, to predict the metabolic networks computationally and to develop PGDB of organisms that have sequenced and annotated genomes (Caspi et al., 2016).

Besides KEGG and BioCyc, other specialized databases for metabolic networks are also available including BRENDA, REACTOME, BiGG and PubChem. BRENDA, the comprehensive enzyme database, provides molecular and biochemical information including functional parameters, enzyme structure and other useful information about enzymes involved in metabolic networks (Placzek et al., 2017). It also provides links to external protein sources like UniProt, and literature references. It compiles the information by manually extracting the information from the literature and using text mining and prediction algorithms. Similarly, ENZYME database also provides information about enzymes and facilitate the development of other metabolic databases (Bairoch, 2000). REACTOME is an organism specific database and focuses on different biological processes, including metabolism, of human cell (Jassal et al., 2020). BiGG is a database

which stores high quality genome-scale metabolic models built from manually curated information available in public databases (King et al., 2016). PubChem, on the other hand, contains information about compounds and their biological activity, and it is complemented by various tools such as compound search and chemical structure sketcher (S. Kim et al., 2016). These and other databases provide necessary data for analyzing metabolic networks and pathways in a biologically meaningful way.

1.2.2 Tools for the analysis of metabolic networks

Metabolic databases have been complemented by the development of tools for analyzing and comparing metabolic networks. These tools have many applications: from understanding biological processes and studying the evolution of metabolic networks by predicting and comparing metabolic pathways, through to providing a platform technology to perform *in silico* metabolic engineering, and design novel pathways for biosynthesis and biodegradation enabling researchers to understand the adaptation and functional specialization in different species (Tomar and De, 2013). Metabolic network analysis tools provide a theoretical framework to analyze the intertwined metabolic routes and functional capabilities of an organism or metabolic network. Some of these tools include Pathway Hunter Tool (Rahman et al., 2005), Pathway Analyst (Pireddu et al., 2006), MetaPath Online (Handorf and Ebenhöf, 2007), MetaRoute (Blum and Kohlbacher, 2008), From Metabolite to Metabolite (Chou et al., 2009), Rahnuma (Mithani et al., 2009a), PathPred (Moriya et al., 2010), Metabolic Route Search and Design (Xia et al., 2011), Metabolic Tinker (McClymont and Soyer, 2013), FogLight (Khosraviani et al., 2015) and Metabolic Route Explorer (Kuwahara et al., 2016). These are further discussed in **Chapter 2**.

Currently available tools for analyzing and comparing metabolic networks and metabolic pathways have several limitations:

- Many of the pathway prediction tools available typically compute shortest pathways between metabolites and do not report multiple pathways between metabolites
- Pathway predictions and network comparisons are performed using graph-based models of metabolic networks and therefore do not capture dependencies between metabolites for reactions that involve more than one substrate and product.
- Some pathway prediction tools fail to take reaction direction into account and treat all reactions as reversible, resulting in biologically improbable predictions.
- Pathway prediction tools are typically limited to simple queries based on a single pair of metabolites in a single organism, and cannot easily be used to investigate complex questions such as the prediction of possible pathways for carbon or nitrogen assimilation starting from a specific nutrient source.
- Many of the available tools do not allow metabolic pathway predictions and network comparisons on custom data or draft organisms thereby limiting their usability on only genome-sequenced organisms.
- Although metabolic databases such as KEGG and BioCyc provide a user-friendly interface for visualizing metabolic network predictions for genome-sequenced organisms, they do not provide an intuitive way to ask biological questions focusing on differences between organisms, interactions between organisms or the evolution of metabolic networks.

- Current tools are unable to process data from the wide range of sources that users may want to analyze, such as transcriptomic and expressed sequence tag data, or genomic data generated using next-generation sequencing methods.

A tool is, therefore, required that can overcome these limitations and provide an interactive and user-friendly interface for the analysis and comparison of metabolic networks allowing users to focus on the differences between organism, study their interactions and evolution at a metabolic level in today's post-genomic era.

1.3 Metabolic network analysis in post-genomic era

With an increased focus on the understanding of the system as a whole, the challenge of predicting and analyzing the properties of metabolic networks based on 'omics data has gained much attention in recent years (Faust et al., 2011). Since the completion of the first genome-scale metabolic model of *Haemophilus influenza* in 1995, many analytical and computational tools/approaches have been designed for the construction of metabolic models (Saha et al., 2014). The availability of genome-scale data has enabled scientists to decipher the links between different metabolic pathways and the underlying genes which code for the enzymes catalyzing the reactions involved in these pathways. The post-genomic era has opened lots of avenues and challenges, which requires sophisticated approaches to understand the spatial and temporal variations and the interplay between different layers and components of biological systems. For example, integration of phenomics (Herrgård et al., 2006) and 'omics data into metabolic networks (Ramon et al., 2018) can help to understand the connections between different layers of cellular interactions. In recent years, integration of expression data with metabolic networks has been used to understand the effect of variation in gene expression on the genome-scale

metabolic networks (Blazier and Papin, 2012), extract the subset of reactions from a genome-scale metabolic network to perform specific analyses or build core models of plants (Pfau et al., 2018) and to understand the metabolic profiles in diseased networks (Graudenzi et al., 2018) in addition to various other studies. However, this has been done manually and for limited datasets. There is an extensive scope to extend the capabilities and interconnectivity of tools and databases, particularly with respect to integration with transcriptomic, proteomic and metabolomic datasets to provide a better understanding of metabolic networks. Thus, the integration of ‘omics data to better understand the functioning of metabolic networks is one of the key challenges of post-genomic era (Huang et al., 2017).

1.4 Overview of the thesis

This thesis focuses on the development of a tool for metabolic network analysis and pathway prediction that can provide an interactive interface for metabolic network analysis, and integrate ‘omic datasets with existing models of metabolic networks in order to predict metabolic pathways and to analyze and compare pathways and metabolic networks across different species in today’s post-genomic era. It also investigates the evolution and functional specialization of metabolic pathways in the bacteria belonging to genus *Pseudomonas*. Finally, it identifies the metabolic changes that occur during ripening of mango fruit by performing metabolic pathway mapping of genes which are found to be differentially expressed during mango ripening. Each chapter ends with a discussion summarizing the work and outlining the possible extensions to the work presented in the chapter. The final chapter summarizes the overall thesis and presents the general extensions to the work presented in this thesis.

The thesis is organized as follows.

- **Chapter 1** outlines the problem statement and provides a general introduction to the thesis.
- **Chapter 2** presents a Metabolic network Analysis and Pathway Prediction Server (MAPPS), a web-based tool for performing a variety of analyses ranging from simple pathway prediction and comparisons to specialized analyses including identification of potential drug targets, *in silico* metabolic engineering, ‘omic filtering of metabolic networks, host-microbe interactions and ancestral network building. It also provides a number of case studies to demonstrate the applications of the tool.
- **Chapter 3** uses MAPPS and other bioinformatics analyses to investigate the evolution and functional specialization of metabolic pathways in the bacteria belonging to genus *Pseudomonas* by systematically performing pathway specific and whole network comparisons and evolutionary analysis of 111 genome-sequenced pseudomonads available in KEGG.
- **Chapter 4** investigates the metabolic changes that occur during fruit maturation in Mango (*Mangifera Indica*) by performing differential expression analysis between immature and mature stages of two elite South Asian cultivars, namely Sindhri and Kala Chaunsa, and studying the differences in metabolic pathway mapping of the identified differentially expressed genes.
- **Chapter 5** summarizes the work presented in this thesis and discusses future directions.

2 MAPPS: A Web-Based Tool for Metabolic Pathway Prediction and Network Analysis in the Post-Genomic Era

The fundamental processes of life, such as the generation of energy from light and food and the synthesis of biological molecules, are carried out by the enzymes that form the metabolic networks present in each living cell. Most organisms have a core set of reactions that catalyze essential processes such as protein synthesis and DNA replication, but a large proportion of the enzymes present in different organisms are specific to the needs of individual organisms or tissues. The availability of genome sequence data allows us to predict the functions of the enzymes present in different organisms, and how these enzymes might be organized into a metabolic network in which the product of one reaction becomes the substrate for another (Mithani et al., 2010). It also offers the possibility of an in-depth analysis of these metabolic pathways within an organism as well as between different organisms to comprehend the processes and attributes that affect the capabilities and evolution of metabolic networks.

Over the years, researchers have developed comprehensive databases that catalog experimentally discovered genes, enzymes, reactions, and pathways of genome-sequenced organisms (Jing et al., 2014), such as the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al., 2017), Reactome (Fabregat et al., 2016), and BioCyc (Caspi et al., 2016), to study metabolic networks at a systems level. These databases have been

complemented by the development of tools for predicting and analyzing metabolic networks, comparing metabolic networks, performing *in silico* metabolic engineering, and designing novel pathways for biosynthesis and biodegradation to understand the adaptation and functional specialization in different species (Tomar and De, 2013). Some of the commonly used tools for metabolic network analysis are listed in **Table 2.1**.

Although the available tools serve as useful resources for metabolic pathway prediction, their usability is limited by the fact that most of them are designed to allow users to examine the metabolism of a single organism or reference metabolic network rather than analyzing data for multiple organisms, which is essential to carry out metabolic comparisons. In addition, many tools including Pathway Hunter Tool (PHT) (Rahman et al., 2005), From Metabolite to Metabolite (FMM) (Chou et al., 2009), Metabolic Route Search and Design (MRSD) (Xia et al., 2011), PathComp (Ogata et al., 1998), PathPred (Moriya et al., 2010) and Metabolic Route Explorer (MRE) (Kuwahara et al., 2016) do not allow users to simultaneously predict pathways between multiple pairs of source and target metabolites making metabolic pathway analysis and comparison even within a single organism a tedious task for the end-user. Another limitation of the currently available tools is the lack of options to filter predicted pathways and to refine pathway searches. Although some tools allow users to define a limited number of constraints during pathway prediction, their usability is limited by the flexibility provided by these tools. For example, PHT allows a provision of requiring a metabolite to be present in the predicted pathways but does not allow metabolites to be avoided during pathway prediction. Similarly, MRSD only provides a provision of intermediary metabolites to be required during pathway prediction. MRE, on the other hand, allows the exclusion of multiple metabolites during pathway

prediction but does not provide an option to require one or more metabolites. Only MetaRoute allows the provision of both required and excluded metabolites/reactions but it is provided as a pathway filtering option once the pathways have been predicted between source and target metabolites and not at the time of pathway prediction itself. Besides this, there is limited support for specialized analyses such as effects of enzyme/reaction insertion and knock-out (MetaPath Online and PHT) on metabolic pathways, support for custom networks (PHT, NetCooperate, MetaRoute and MetQuest), and host-pathogen interactions (NetCooperate) in currently available tools.

Metabolic networks, like all other biological networks, are under a process of continuous evolution. However, the evolutionary mechanisms of these networks are not well understood. It is unclear how these networks evolve and if there is a correlation between the evolution of metabolic capabilities and various factors such as the network structure and/or the environment in which these organisms thrive. The availability of genomes for many closely related species offers the possibility of tracing metabolic evolution on a phylogeny relating the genomes to understand the evolutionary processes and constraints that affect the evolution of metabolic networks. To the best of our knowledge, none of the currently available tools allows phylogeny-based analyses focusing on the evolution of metabolic networks. Furthermore, with an increased focus on the understanding of the system as a whole, the challenge of predicting and analyzing the properties of metabolic networks based on transcriptomic, proteomic and metabolomic data has gained much attention in recent years (Faust et al., 2011).

Table 2.1: Tools available for metabolic network analysis

Tool Name	Web-based Tool	Active	Database Last Updated	Supported Analyses						Reference
				Pathway Prediction	Comparative Analysis	Support for Custom Networks	Evolutionary Analysis	Drug Target Identification	Host-Microbe Interaction	
ATLAS	✓	✓	2015 [†]	✓						(Hadadi et al., 2016)
BPAT-S/ BPAT-M	✓	✓	20/06/2011	✓						(Heath et al., 2011)
FMM	✓	✓	01/10/2008	✓	✓					(Chou et al., 2009)
FogLight		✓	Unknown*	✓						(Khosraviani et al., 2015)
Metabolic Tinker			N/A	✓						(McClymont and Soyer, 2013)
MetaPath Online	✓	✓	13/2/2007	✓		✓				(Handorf and Ebenhöf, 2007)
MetaRoute	✓	✓	21/10/2007	✓	✓	✓				(Blum and Kohlbacher, 2008)
MetQuest		✓	Unknown*	✓		✓				(Ravikrishnan et al., 2018)
MRE	✓	✓	01/10/2015	✓						(Kuwahara et al., 2016)
MRSD	✓		N/A	✓						(Xia et al., 2011)
NetCooperate		✓	Unknown*						✓	(Levy et al., 2015)
PathComp	✓	✓	Real-time	✓						(Ogata et al., 1998)
PathPred	✓	✓	Real-time	✓						(Moriya et al., 2010)
PHT	✓	✓	05/04/2011	✓		✓		✓		(Rahman et al., 2005)
Rahnuma	✓		N/A	✓	✓		✓			(Mithani et al., 2009a)

[†]ATLAS database has been developed using KEGG 2015 reactions

*Last updated date not available

Most of the currently available pathway prediction tools do not allow the user to map ‘omics data on the metabolic network and compare metabolic networks to see the effects of ‘omics data on the pathways between specific metabolites. Finally, currently available tools do not provide an interactive interface for visualizing pathway prediction results or exporting the results into standard Systems Biology Markup Language (SBML) for further analyses. A tool is, therefore, required that addresses these limitations and allows users to analyze a wide range of metabolic analyses, integrate ‘omic datasets in order to predict and compare metabolic networks, visualize pathway predictions in a biologically meaningful way and obtain insight into the functioning and evolution of metabolic networks.

This chapter describes a tool called MAPPS: Metabolic network Analysis and Pathway Prediction Server that provides an interactive platform to analyze metabolic networks using traditional as well as ‘omics data. MAPPS is a web-based tool available at <https://mapps.lums.edu.pk> and builds upon the existing architecture of *Rahnuma* (Mithani et al., 2009a), a tool which was published over a decade ago but is no longer available for use. Although *Rahnuma* had a variety of distinctive features that are not available in many of the available tools including multi-network pathway prediction, and network comparison between two or more organisms or at different levels of a phylogeny, its utility as a tool for studying network evolution, predicting metabolic capabilities and analyzing ‘omic data was limited by its basic user interface, text-based output, and reliance on KEGG-derived annotations of completely sequenced genomes (see **Section 2.1**). MAPPS has been designed to overcome these limitations by providing an interactive user interface for job submission and graphical result visualization and allowing users to upload custom

data to enable analyses on draft and custom genomes in addition to providing diverse functionalities. Like its predecessor, MAPPS represents metabolic networks as hypergraphs, rather than the commonly used graph representation. A hypergraph is a generalization of a regular graph where an edge, called a hyperedge, can connect more than two vertices (Berge and Minieka, 1973). Since a reaction is treated as a single entity in a hypergraph, it can be used to capture relationships between multiple metabolites involved in a reaction, unlike regular graphs where each edge is independent. Hypergraphs have been used to represent metabolic networks in different studies (Yeung et al., 2007; Mithani et al., 2009b). Overall, MAPPS aims to provide a single powerful resource for the analysis and comparison of metabolic networks and for the study of metabolic evolution by allowing pathway-based metabolic network comparisons at the organism as well as phylogenetic levels.

2.1 Rahnuma: hypergraph-based tool for metabolic pathway prediction and network comparison

Rahnuma, a web-based tool for metabolic pathway prediction and network comparison (**Figure 2.1**), was published in 2009 (Mithani et al., 2009a) but is no longer available for use. *Rahnuma* represented metabolic networks as hypergraphs rather than the commonly used graph representation and considered reaction directions while predicting pathways between metabolites by using directed hyperedges. Analyses provided by *Rahnuma* were grouped into three categories: (i) Pathway Analysis for pathway predictions on more or more organisms, (ii) Comparative Analysis for pathway based and full-network comparison between two networks, and (iii) Network Analysis for phylogenetic-based analysis. *Rahnuma* provided a variety of distinctive features including multi-network

Rahnuma: hypergraph-based tool for metabolic pathway prediction and network comparison

Job Detail

Job Name * [?]

Email * [?]

Dataset

Select KEGG Pathways [?]

☒ All Available KEGG Pathways

☐ All Curated KEGG Pathways

☐ Select

Job Specification

Analysis [?]

☐ Pathway Analysis

☒ Comparative Analysis

☐ Network Analysis

Analysis Options

Type [?]

☒ Standard Comparison

☐ All but One Comparison

Mode [?]

☒ Organism

☐ Phylogeny

Comparison [?]

☐ Pathway Based

☒ Full Network

Output Options

Type [?]

☒ Tabular

☐ Descriptive

Format [?]

☒ Text

☐ HTML

Figure 2.1: *Rahnuma* interface for job submission. *Rahnuma*, a web-based tool for metabolic pathway prediction and network comparison, was published in 2009 (Mithani et al., 2009b) and provided a basic interface for job submission and had limited options available.

pathway prediction, and network comparison between two or more organisms or at different levels of a phylogeny.

Rahnuma's utility as a tool for predicting metabolic capabilities, comparing metabolic networks and studying network evolution was, however, limited due to its strong reliance on KEGG-derived annotations of completely sequenced genomes. *Rahnuma*, for example, allowed user to perform simple pathway prediction between two or more

metabolites in addition to identifying reactions involved in pathways from only one of the given start metabolites and reactions that, if deleted from the network, would result in all pathways being removed between the specified metabolites. This, however, was restricted to the networks built using underlying data downloaded from KEGG. It did not allow user to alter metabolic networks by allowing addition or removal of metabolites or reactions from the network or run these analyses on user-defined networks thus limiting its capabilities for answering diverse biological questions. Similarly, comparative and phylogenetic analyses were allowed only on the KEGG data with no option to perform *in silico* modification of the network. Comparison of modified network with the original network is an important aspect of metabolic network analysis in today's post-genomic era enabling user to study the behavior of diseased or engineered metabolic network (Tomar and De, 2013). Moreover, the all-but-one comparison (see **Section 2.2.4.1**), which facilitates the identification of reactions or pathways that are present in only one organism in a group but absent in all other organisms and vice versa was only available for single organisms in *Rahnuma* thereby limiting its usability. Allowing all-but-one comparison between groups of organisms can enable users to compare metabolic networks between different species or genera thereby enabling them to study metabolic capabilities of organisms at a broader level. Besides this, metabolic network analysis at phylogenetic level, a unique aspect of *Rahnuma*, provided only three algebraic modes (union, intersection and reaction neighborhood) to build metabolic networks on the user-defined phylogeny. It did not provide any provision to infer phylogeny based on metabolic network data.

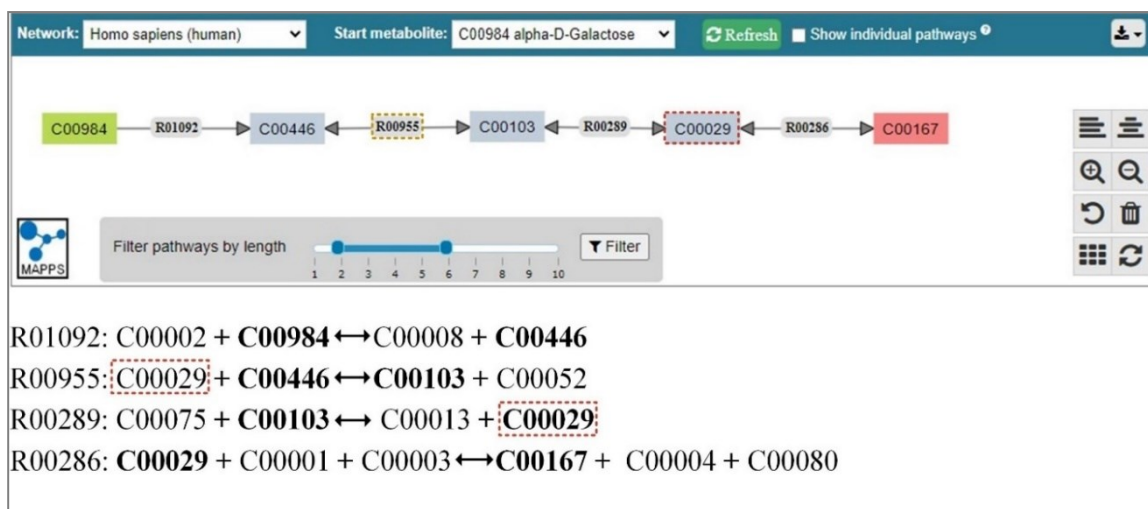


Figure 2.2: An example pathway to demonstrate limitation of pathway prediction algorithm of *Rahnuma*. It shows MAPPS output of pathway prediction α -D-galactose to UDP-glucuronate. Metabolite which acts as a substrate in already added reaction (yellow box) is highlighted (red box).

Besides this, since *Rahnuma* did not allow any metabolite to be added as an intermediate metabolite if it was present as a substrate in any of the reactions of the predicted pathway, it failed to capture many pathways by ignoring substrates of participating reactions. For example, consider the four-step metabolic pathway from α -D-galactose (KEGG ID: C00984) to UDP-glucuronate (KEGG ID: C00167) shown in **Figure 2.2**. In this pathway, α -D-galactose is converted into α -D-galactose-1-phosphate (KEGG ID: C00446) via reaction R01092, which is subsequently converted into D-glucose-1-phosphate (KEGG ID: C00103) by the reaction R00955. In the next step, UDP-glucose (KEGG ID: C00029) is added as an intermediate metabolite in the pathway by the reaction R00289 to reach to the destination metabolite (UDP-glucuronate). *Rahnuma* could not predict this pathway since UDP-glucose (KEGG ID: C00029) is also present as a substrate in the reaction R00955 (**Figure 2.2**). MAPPS overcomes this limitation by allowing a metabolite to be added as an intermediate metabolite even if it is present as a substrate in any of the participating reactions of the predicted pathway (see **Section 2.2.4.1**).

```

Job Name: Metabolism of polyamines - Pathways
DATASET
KEGG Pathway Maps: All Available
JOB DETAILS
Analysis: Comparative Analysis
Analysis Type: Standard Comparison
Analysis Mode: Organism
Comparison Mode: Pathway Based
OUTPUT DETAILS
Output Type: Tabular
Output Format: Text
JOB PARAMETER(S)
Organism(s) 1:Pseudomonas syringae pv. syringae B728a [psb], Pseudomonas syringae pv. phaseolicola 1448A [pap], Pseudomonas syringae pv. tomato DC3000 [pst]
Organism(s) 1 Mode:All Reactions
Organism(s) 2:Pseudomonas fluorescens Pf-5 [pf1], Pseudomonas fluorescens Pf0-1 [pf0], Pseudomonas fluorescens SBW25 [pfs], Pseudomonas putida F1 [ppf], Pseudomonas putida GB-1 [ppg], Pseudomonas putida KT2440 [ppu], Pseudomonas putida W619 [ppw]
Organism(s) 2 Mode:All Reactions
PATHWAY PARAMETERS
Start Metabolite(s): L-Arginine [C00062], L-Ornithine [C00077]
End Metabolite(s): Putrescine [C00134], Spermidine [C00315], Spermine [C00750]
Pathway Prediction Mode: Connection
Connection Option: KEGG RPAIRs
Cutoff Length: 7
Ignore Metabolites: Yes
Return Value: Pathway

Start Metabolites Length Pathway Pseudomonas syringae pv. syringae B728a + Pseudomonas syringae pv. phaseolicola 1448A + Pseudomonas syringae pv. tomato DC3000 [psb+pap+pst] Pseudomonas
fluorescens Pf-5 + Pseudomonas fluorescens Pf0-1 + Pseudomonas fluorescens SBW25 + Pseudomonas putida F1 + Pseudomonas putida GB-1 + Pseudomonas putida KT2440 + Pseudomonas putida W619
[ppf+ppg+ppu+ppw]
L-Arginine [C00062] 2 L-Arginine [C00062] <--(*R00565*)--> L-Ornithine [C00077] <--(R00670)--> Putrescine [C00134] Y N
L-Arginine [C00062] 2 L-Arginine [C00062] <--(*R01989*)--> L-Ornithine [C00077] <--(R00670)--> Putrescine [C00134] Y N
L-Arginine [C00062] 7 L-Arginine [C00062] <--(*R00551*)--> Urea [C00086] <--(*R00774*)--> Urea-1-carboxylate [C01010] <--(*R00005*)--> NH3 [C00014] <--(R00243)--> L-Glutamate [C00025]
<--(R02283)--> N-Acetylornithine [C00437] <--(R00669)--> L-Ornithine [C00077] <--(R00670)--> Putrescine [C00134] Y N
L-Arginine [C00062] 7 L-Arginine [C00062] <--(*R00551*)--> Urea [C00086] <--(*R00774*)--> Urea-1-carboxylate [C01010] <--(*R00005*)--> NH3 [C00014] <--(R00243)--> L-Glutamate [C00025]
<--(R02283)--> N-Acetylornithine [C00437] <--(R02283)--> L-Ornithine [C00077] <--(R00670)--> Putrescine [C00134] Y N
L-Arginine [C00062] 7 L-Arginine [C00062] <--(*R00565*)--> L-Ornithine [C00077] <--(R00669)--> N-Acetylornithine [C00437] <--(R02283)--> L-Glutamate [C00025] <--(R07396)--> L-Methionine [C00073]
<--(R00177)--> S-Adenosyl-L-methionine [C00019] <--(*R00178)--> S-Adenosylmethioninamine [C01137] <--(R01920)--> Spermidine [C00315] Y N
L-Arginine [C00062] 7 L-Arginine [C00062] <--(*R00565*)--> L-Ornithine [C00077] <--(R00669)--> N-Acetylornithine [C00437] <--(R02283)--> L-Glutamate [C00025] <--(R07396)--> L-Methionine [C00073]
<--(R00177)--> S-Adenosyl-L-methionine [C00019] <--(*R00178)--> S-Adenosylmethioninamine [C01137] <--(R02869)--> Spermine [C00750] Y N
L-Arginine [C00062] 7 L-Arginine [C00062] <--(*R00565*)--> L-Ornithine [C00077] <--(R00669)--> N-Acetylornithine [C00437] <--(R02283)--> L-Glutamate [C00025] <--(R07396)--> L-Methionine [C00073]
<--(R00178)--> S-Adenosyl-L-methionine [C00019] <--(*R00178)--> S-Adenosylmethioninamine [C01137] <--(R01920)--> Spermidine [C00315] Y N
L-Arginine [C00062] 7 L-Arginine [C00062] <--(*R00565*)--> L-Ornithine [C00077] <--(R00669)--> N-Acetylornithine [C00437] <--(R02283)--> L-Glutamate [C00025] <--(R07396)--> L-Methionine [C00073]
<--(R00178)--> S-Adenosyl-L-methionine [C00019] <--(*R00178)--> S-Adenosylmethioninamine [C01137] <--(R02869)--> Spermine [C00750] Y N
L-Arginine [C00062] 7 L-Arginine [C00062] <--(*R00565*)--> L-Ornithine [C00077] <--(R00669)--> N-Acetylornithine [C00437] <--(R02283)--> L-Glutamate [C00025] <--(R07396)--> L-Methionine [C00073]
<--(R06895)--> S-Adenosyl-L-methionine [C00019] <--(*R00178)--> S-Adenosylmethioninamine [C01137] <--(R01920)--> Spermidine [C00315] Y N
L-Arginine [C00062] 7 L-Arginine [C00062] <--(*R00565*)--> L-Ornithine [C00077] <--(R00669)--> N-Acetylornithine [C00437] <--(R02283)--> L-Glutamate [C00025] <--(R07396)--> L-Methionine [C00073]
<--(R06895)--> S-Adenosyl-L-methionine [C00019] <--(*R00178)--> S-Adenosylmethioninamine [C01137] <--(R02869)--> Spermine [C00750] Y N
L-Arginine [C00062] 7 L-Arginine [C00062] <--(*R00565*)--> L-Ornithine [C00077] <--(R00669)--> N-Acetylornithine [C00437] <--(R02283)--> L-Glutamate [C00025] <--(R07396)--> L-Methionine [C00073]
<--(R07767)--> S-Adenosyl-L-methionine [C00019] <--(*R00178)--> S-Adenosylmethioninamine [C01137] <--(R01920)--> Spermidine [C00315] Y N
L-Arginine [C00062] 7 L-Arginine [C00062] <--(*R00565*)--> L-Ornithine [C00077] <--(R00669)--> N-Acetylornithine [C00437] <--(R02283)--> L-Glutamate [C00025] <--(R07396)--> L-Methionine [C00073]
<--(R07767)--> S-Adenosyl-L-methionine [C00019] <--(*R00178)--> S-Adenosylmethioninamine [C01137] <--(R02869)--> Spermine [C00750] Y N

```

Figure 2.3: Plain text output of pathway-based comparison of *Rahnuma*. It did not provide a graphical interface for results visualization results.

In addition, *Rahnuma*'s utility as a tool was also limited by its basic user interface (Figure 2.1), which prevented the user from studying complex questions focusing on metabolic capabilities of different organisms, and text-based output (Figure 2.3) preventing user to visualize the results in an interactive graphical user interface. *Rahnuma* did not allow the user to build metabolic network for multiple organism sets in a single job (Figure 2.4) forcing them to submit multiple jobs with same parameters for each organism set. In addition, although the user could choose multiple start and end metabolites in *Rahnuma*, it lacked the flexibility to allow user-defined constraints on intermediary metabolites (Figure 2.4, see Section 2.2.4.1).

In summary, *Rahnuma* allowed answering biological questions focusing on metabolic differences between organisms by allowing reaction and pathway based metabolic network comparisons in an organismal and phylogenetic context and provided

Parameters

Organism(s) * [?]

☐ Reference

☒ Select

Show Subset: Eukaryotes > Animals > Ascidians

Available:		Selected:
Ciona intestinalis (sea squirt) [cin]	Add >	
Ciona savignyi (EST) [ecsv]	Add All >>	
Molgula tectiformis (EST) [emte]	< Remove	
	<< Remove All	

Pathway Specific Parameters

Start Metabolites * [?]

Show Subset: Amino Acid

Available:		Selected:
4-Aminobutanoate [C00334]	Add >	
D-Alanine [C00133]	Add All >>	
D-Arginine [C00792]	< Remove	
D-Aspartate [C00402]	<< Remove All	
D-Cysteine [C00793]		
D-Glutamate [C00217]		
D-Glutamine [C00819]		
D-Lysine [C00739]		
D-Phenylalanine [C02265]		
D-Proline [C00763]		
D-Serine [C00740]		
Glycine [C00037]		

Note: Only metabolites present in selected pathway maps are available for pathway prediction

End Metabolites * [?]

Show Subset: Amino Acid

Available:	Selected:

Figure 2.4: Screenshot of *Rahnuma* interface to select organisms and metabolites. User can select multiple organisms to build a metabolic network, and choose multiple start and end metabolites for pathway prediction

basic architecture for it but it lacked the scalability and user-centric design required to answer complex biological questions by specifying constraints at metabolite-, reaction- and enzyme-levels, and did not provide a graphical interface for results visualization.

2.2 MAPPS: Metabolic Network Analysis and Pathway Prediction Server

MAPPS: Metabolic network Analysis and Pathway Prediction Server is a web-based tool available at <https://mapps.lums.edu.pk>. MAPPS presents an interactive platform to analyze

metabolic networks using traditional as well as ‘omics data by offering an intuitive, easy to use and biologically meaningful graphical interface to the user. It provides a powerful resource for metabolic pathway prediction and comparison as well as specialized analyses including identification of metabolic drug targets, detection of metabolite-specific reactions, analyzing the effects of host-microbe interactions, and studying metabolic evolution using traditional as well as stochastic models (**Figure 2.5**). MAPPS also allows users to upload custom data in addition to using data from KEGG, thus enabling metabolic analyses on draft and custom genomes, enables *in silico* metabolic engineering by adding/removing metabolic reactions or enzymes and has an ‘omics pipeline to refine the pathway results using transcriptomic, proteomic, or metabolomic data to provide a greater insight into the metabolic capabilities of organisms, making it relevant in today’s postgenomic era. The tool is described in detail in subsequent sections followed by a few case studies highlighting important features of the tool.

2.2.1 MAPPS architecture

MAPPS is built on .Net framework and is hosted on a Dell PowerEdge R740 Server with two Intel(R) Xeon Silver 4110 2.1 GHz CPUs with eight cores each and 64GB of Memory running Windows Server 2012 attached to Dell Power Vault MD1200 storage box with 10TB of storage space. MAPPS is designed in a way to minimize the dependency among its components and enhance scalability (**Figure 2.6**). It is divided into the following parts.

2.2.1.1 Database

MAPPS uses KEGG as the primary data source for organisms, compounds, reactions, and metabolic pathway maps (Kanehisa et al., 2017). External mapping for compounds and

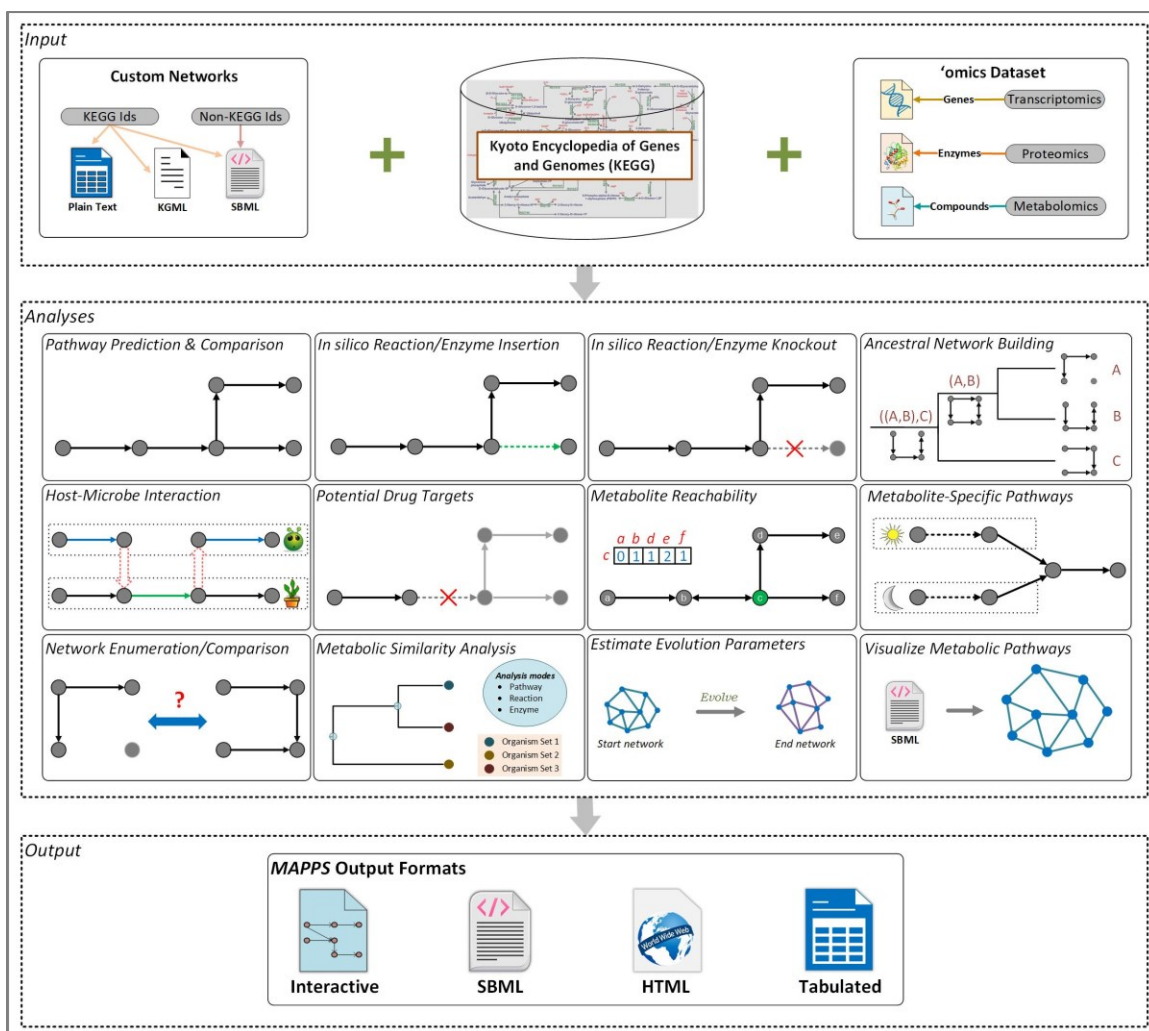


Figure 2.5: Overview of MAPPS. MAPPS uses data from the KEGG database in addition to allowing users to upload custom metabolic networks containing KEGG or non-KEGG identifiers. KEGG-based networks can be refined based on ‘omics (transcriptomics, proteomics and metabolomics) datasets. MAPPS provides several analyses including pathway prediction and comparison, ancestral network building/comparison, host-microbe interaction, identifying potential drug targets, metabolite reachability, network enumeration and comparison, metabolic similarity analysis, metabolite-specific pathways, estimation of evolution parameters and interactive network viewer. Users can also modify networks using in silico enzyme/reaction insertion or knock-out. MAPPS output can be generated in multiple output formats including Hyper Text Markup Language (HTML) file containing hyperlinks to KEGG for compound and reaction entities, Systems Biology Markup Language (SBML) file, tab-delimited text file and interactive graphical format.

enzymes to other public databases is obtained through the KEGG API and reaction directions are extracted by parsing KEGG Markup Language (KGML) files of reference reaction maps. MAPPS provides an option to use reaction energies and metabolic structural

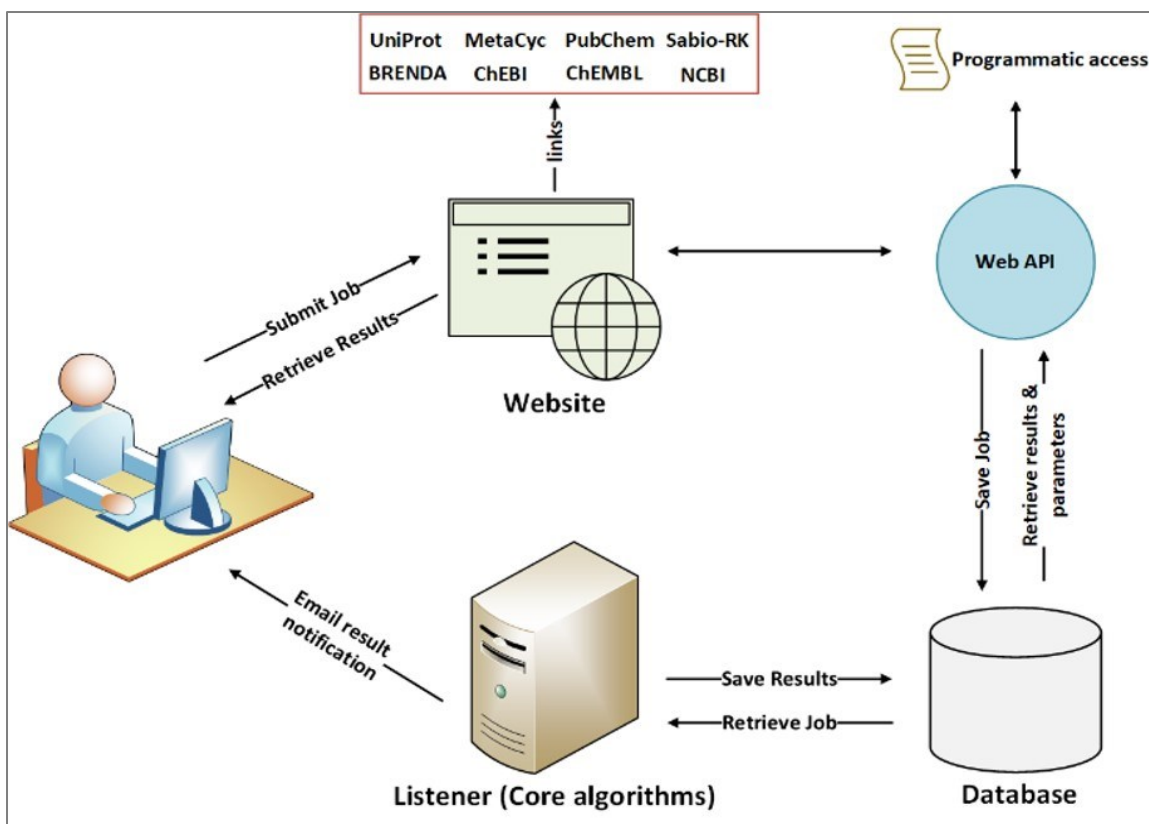


Figure 2.6: MAPPS architecture. MAPPS has a modular architecture to minimize the dependency among its components and enhance scalability. It consists of a listener program which is written in C#, a MySQL database for storing data and job-related information, Web Application Programming Interface (API) which communicates between the MAPPS website and the database, in addition to providing a public interface for programmatic access, and the interactive MAPPS website for job submission and result visualization as well as quick access to relevant public databases.

similarity into the pathway scoring scheme (see **Section 2.2.7**). For this, standard Gibbs energies of KEGG reactions for pH values ranging from 5 to 9 were downloaded from *eQuilibrator* (Flamholz et al., 2012) and structural similarity scores of metabolites were obtained through the REST API of SIMCOMP2 (Hattori et al., 2010). Data is stored in a relational database (see **Appendix A**) running on a MySQL server. The database also stores external links, login information, and job parameters. KEGG data is updated in the database fortnightly. The results are stored in the database for three days for guest users and fifteen days for registered users.

2.2.1.2 Website

The MAPPS website provides an interactive platform to submit complex queries. It is developed in ASP.net, HTML, CSS, and various JavaScript libraries (JQuery, AJAX, JointJS, Bootstrap, Dagre, and Vectorizer) are used to make it more interactive. MAPPS provides a seamless four-step job submission process for various metabolic network analyses. Jobs can be submitted as a registered user or guest user; registered users can save their job parameters to submit later and retrieve the results of previous jobs. MAPPS website communicates with the database through web API for retrieving data and submitting jobs (**Figure 2.6**) and provides quick access to relevant public databases through an interactive interface.

2.2.1.3 Application Programming Interface (API)

The API provides a public interface between client and server. In MAPPS, API facilitates the communication between the website and the MAPPS database. It retrieves the data from the database and sends it to the website and takes the job parameters submitted by the user from the website, validates it, and submits the job in the database. The Listener takes this job and executes it; when the execution is completed, and job status is updated in the database, the API fetches the result from the database and sends it to the website on the client's browser.

2.2.1.4 Listener

A desktop application written in C#, named as *Listener*, runs all the core algorithms of MAPPS. This application runs as a multithread process and performs three primary functions, database polling, job status monitoring, and job execution. The database polling thread samples the database at regular intervals to fetch newly submitted jobs. If the user

submits a new job, the database polling thread adds it to the job queue, which is a priority queue. Another thread actively monitors the status of queued entries, and it starts an independent thread to execute the job and remove it from the job queue. Now the job is moved to the ‘active jobs’ list. Currently, MAPPS can execute ten jobs in parallel using a multithreading approach. When a job is completed, job status is updated in the database and results are saved on the server. If the number of active jobs is less than 10, a new job is executed from the job queue, and this process goes on. Multiple threads work together to fetch the job from the database, execute it, update its status in the database and move to the next job.

2.2.2 Working of the tool

2.2.2.1 Registration

User can use MAPPS by creating an account or as a “Guest User” (**Figure 2.7**). MAPPS allows user to create an account using a simple registration process (**Figure 2.8**). An email address is used as a unique identifier to create a new account, after submitting the registration form an activation link will be sent to the given email address for verification. Once the email address is verified, the account will be activated, and user can login to use MAPPS. Registered users can use additional features of MAPPS like they can save the job parameters of unsubmitted jobs to edit and submit later (**Figure 2.9**), and results of completed jobs will be accessible for 15 days. While on the other hand, guest users can access the results only for three days.

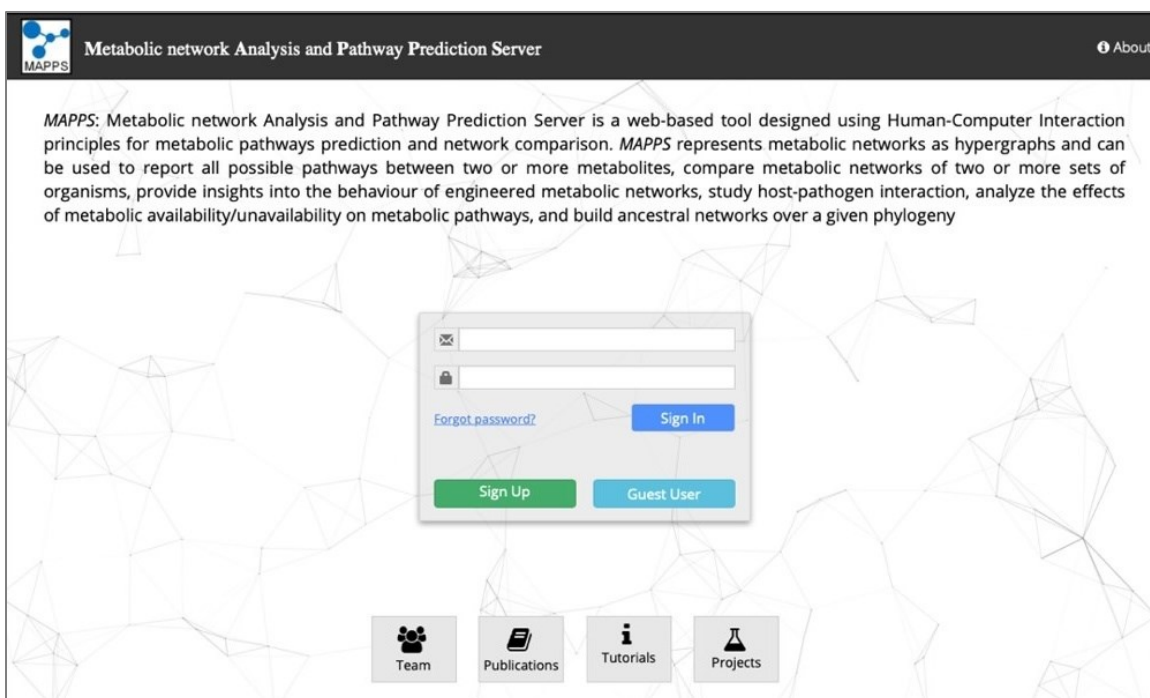


Figure 2.7: MAPPS login page. Users can log in with a registered email address or use MAPPS as a guest user. Registered users can save job parameters of unsubmitted jobs, and access completed jobs.

Metabolic network Analysis and Pathway Prediction Server

MAPPS

Home **Feedback** **About**

Register

All fields are required.

Name

Organization

Email address

you will receive account activation link on this email address

Password minimum 6 characters

Confirm Password


☐ I'm not a robot 

Figure 2.8: MAPPS interface for user account registration. User can register using an email address, which will be used as user ID for sign in.

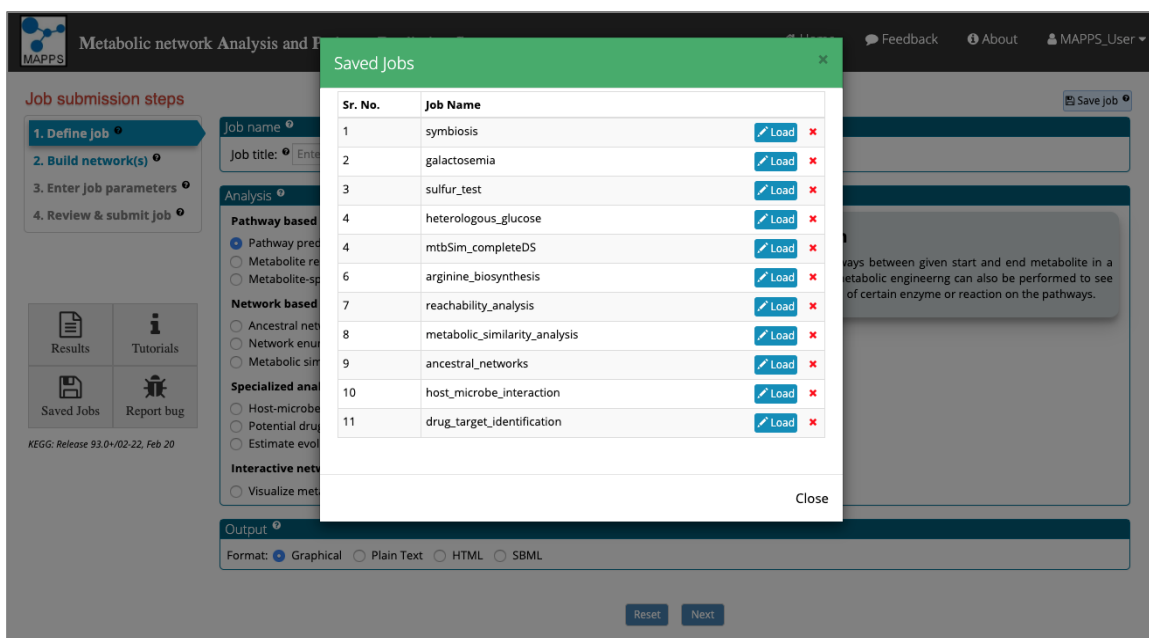


Figure 2.9: MAPPS interface to reload previously saved jobs. The user can save multiple jobs which can be subsequently reload. Reloading the previously saved jobs allows the user to edit the job parameters and submit the jobs.

2.2.2.2 Job submission

Once logged in, user can submit jobs in MAPPS. Job submission in MAPPS consists of four main steps (**Figure 2.10**):

2.2.2.2.1 Define job

At the first step of job submission (**Figure 2.11**), the user provides a job name and selects the analysis to be performed. MAPPS provides ten different analyses (see **Section 2.2.4**), which are grouped into four categories namely pathway-based analyses, network-based analyses, specialized analysis and interactive network viewer. Users also chooses the output format for the results (see **Section 2.2.3**).

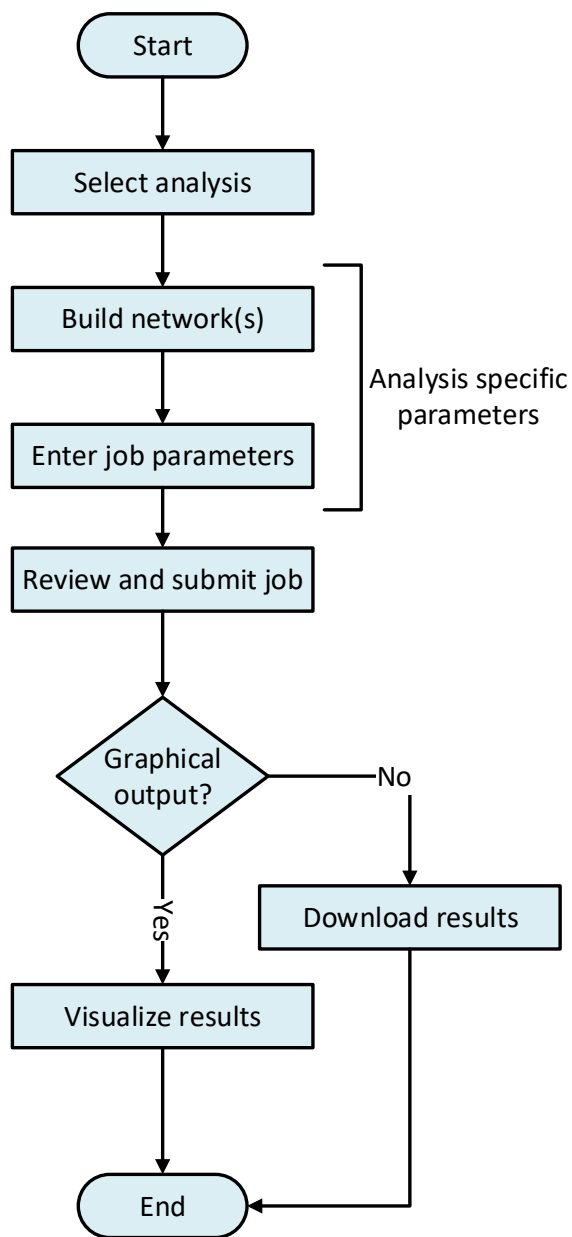


Figure 2.10: Flowchart describing MAPPS workflow. The job submission process starts with the selection of analysis followed by the user definition of the dataset and organism set(s) to build metabolic network(s). The user then enters job-specific parameters (if applicable), and finally reviews and submits the job. After the job completes, if the selected output format is graphical, the user will be redirected to the interactive result visualizer; otherwise the user can download the result (HTML, SBML or text) file.

Metabolic network Analysis and Pathway Prediction Server

Home Feedback About GuestUser

Job submission steps

1. Define job (selected)
2. Build network(s)
3. Enter job parameters
4. Review & submit job

Results Tutorials
Saved Jobs Report bug

KEGG: Release 93.0+02-22, Feb 20

Job name
Job title: Enter job name...

Analysis

Pathway based analysis

- ☒ Pathway prediction and comparison
- ☐ Metabolite reachability
- ☐ Metabolite-specific pathways

Network based analysis

- ☐ Ancestral network building/comparison
- ☐ Network enumeration/comparison
- ☐ Metabolic similarity analysis

Specialized analysis

- ☐ Host-microbe interaction
- ☐ Potential drug targets
- ☐ Estimate evolution parameters

Interactive network viewer

- ☐ Visualize metabolic network/pathways

Pathway Prediction

Predict all the possible pathways between given start and end metabolite in a metabolic network. *In silico* metabolic engineering can also be performed to see the effect of addition/removal of certain enzyme or reaction on the pathways.

Output

Format: ☒ Graphical ☐ Plain Text ☐ HTML ☐ SBML

Reset Next

Figure 2.11: MAPPS graphical user interface for job submission. The screenshot shows the first step of job submission listing the different analyses available in MAPPS. Job submission in MAPPS consists of four steps: (1) define job (2) build network (3) enter job parameters (if applicable), and (4) review and submit the job. Users can provide a meaningful job name and select the desired output format.

2.2.2.2.2 Build network

In the second step of job submission, user selects the data source, i.e. KEGG pathways or uploads custom data to build a metabolic network(s). In the case of custom data, users can upload metabolic network(s) in SBML format, and for the KEGG pathways, networks can be built at an organism level by specifying one or more organism sets, each set can have multiple organisms, using search box or taxonomy viewer to search for the organism(s) (**Figure 2.12** and **Figure 2.13**) or over a phylogeny (**Figure 2.14**). MAPPS allows user to choose up to 10 organisms sets in a single job to build metabolic networks, in addition to compare the results it helps to avoid the hassle of submitting multiple jobs for the same parameters. Autofill feature of search box can help user to find the organism intuitively and the taxonomy view is to facilitate the specie identification by their taxonomic

Metabolic network Analysis and Pathway Prediction Server

Home
Feedback
About
GuestUser

Job submission steps

1. Define job
2. Build network(s)
3. Enter job parameters
4. Review & submit job

Analysis:
Pathway prediction and comparison
Show job summary

ResultsTutorials

Saved JobsReport bug

KEGG: Release 93.0+/02-22, Feb 20

Data source

KEGG
Custom data

MAPPS output is dependent on the accuracy of KEGG annotations

Dataset

All pathways
Select pathways

MAPPS may not report all known pathways depending on the dataset used

Pathway set: Carbohydrate metabolism
Search pathways...

ID	Pathway Name	Pathway Set
00660	C5-Branched dibasic acid metabolism	Carbohydrate metabolism
00020	Citrate cycle (TCA cycle)	Carbohydrate metabolism
00051	Fructose and mannose metabolism	Carbohydrate metabolism
00052	Galactose metabolism	Carbohydrate metabolism
00010	Glycolysis / Gluconeogenesis	Carbohydrate metabolism
00430	Gluconate and dicarboxylate metabolism	Carbohydrate metabolism

Amino sugar and nucleotide sugar metabolism
Citrate cycle (TCA cycle)
Inositol phosphate metabolism
Starch and sucrose metabolism

Ascorbate and aldarate metabolism
Fructose and mannose metabolism
Galactose metabolism
Pentose and glucuronate interconversions

Butanoate metabolism
Glycolysis / Gluconeogenesis
Pentose phosphate pathway
Propanoate metabolism

C5-Branched dibasic acid metabolism
Glyoxylate and dicarboxylate metabolism
Pyruvate metabolism

Network building mode

Organism set(s)
Phylogeny

Organism set(s)

Organism set 1

Use
Organism-based network
Reference network
Custom network (KEGG-based)

Search Box
Taxonomy View

Pseudomonas aeruginosa PAO1 (ID:pae) fluor

Build network using: All reactions

Perform in silico metabolic engineering
Filter network using 'omics dataset'

Pseudomonas fluorescens LBUM223 (ID:pfb)
Pseudomonas fluorescens A506 (ID:pfk)
Pseudomonas fluorescens F113 (ID:pfh)
Pseudomonas fluorescens PICF7 (ID:pfj)
Pseudomonas fluorescens UK4 (ID:pfk)
Pseudomonas fluorescens Pf0-1 (ID:pfh)
Pseudomonas fluorescens SBW25 (ID:pfk)
Pseudomonas fluorescens PCL1751 (ID:pfw)
Pseudomonas fluorescens L228 (ID:pfk)

Organism set 2

Use
Organism-based network
Reference network
Custom network (KEGG-based)

Search Box
Taxonomy View

Search Tree...
Reset

Prokaryotes
Bacteria
Betaproteobacteria
Firmicutes - Bacilli
Aquificae
Gammaproteobacteria - Others
Actinobacteria
Alphaproteobacteria
Firmicutes - Clostridia


Build network using: All reactions

Perform in silico metabolic engineering
Filter network using 'omics dataset'
Perform comparative analysis

Add organism set

Figure 2.12: Building metabolic networks by specifying organism(s) using search box or taxonomy viewer. Each organism set can be defined on one or more species. MAPPS allows user to select one or more organism(s) using search box or hierarchical checkboxes which represent the taxonomical classification of organism.

classification of KEGG (Figure 2.12). MAPPS allows the user to make changes in the metabolic networks by adding or removing reactions/enzymes from these networks using


Metabolic network Analysis and Pathway Prediction Server
Home
Feedback
About
GuestUser

Job submission steps

1. Define job
2. **Build network(s)**
3. Enter job parameters
4. Review & submit job

Analysis:
Pathway prediction and comparison
[Show job summary](#)

Results

Tutorials

Saved Jobs

Report bug

KEGG: Release 93.0-/02-22, Feb 20

Data source

☒ KEGG
☐ Custom data

MAPPS output is dependent on the accuracy of KEGG annotations

Dataset

☐ All pathways
☒ Select pathways

MAPPS may not report all known pathways depending on the dataset used

Pathway set:
Carbohydrate metabolism

Search pathways...

ID	Pathway Name	Pathway Set
<input checked="" type="checkbox"/> 00020	Citrate cycle (TCA cycle)	Carbohydrate metabolism
<input type="checkbox"/> 00051	Fructose and mannose metabolism	Carbohydrate metabolism
<input checked="" type="checkbox"/> 00052	Galactose metabolism	Carbohydrate metabolism
<input checked="" type="checkbox"/> 00010	Glycolysis / Gluconeogenesis	Carbohydrate metabolism
<input type="checkbox"/> 00630	Glyoxylate and dicarboxylate metabolism	Carbohydrate metabolism
<input type="checkbox"/> 00663	Inositol phosphate metabolism	Carbohydrate metabolism

Citrate cycle (TCA cycle)
Glycolysis / Gluconeogenesis
Galactose metabolism

Network building mode

☒ Organism set(s)
☐ Phylogeny

Organism set(s)

Organism set 1

Use
☒ Organism-based network
☐ Reference network
☐ Custom network (KEGG-based)

Search Box
Taxonomy View

Escherichia coli K-12 MG1655 (ID:eco)

Search organism...

Build network using:
All reactions

☐ Perform *in silico* metabolic engineering
☐ Filter network using 'omics dataset'

Organism set 2

Use
☒ Organism-based network
☐ Reference network
☐ Custom network (KEGG-based)

Search Box
Taxonomy View

Brucella suis 1330 (ID:bsi)

Search organism...

Build network using:
All reactions

☐ Perform *in silico* metabolic engineering
☐ Filter network using 'omics dataset'

Organism set 3

Use
☒ Organism-based network
☐ Reference network
☐ Custom network (KEGG-based)

Search Box
Taxonomy View

Paenibacillus sp. JDR-2 (ID:pjd)

Search organism...

Build network using:
All reactions

☐ Perform *in silico* metabolic engineering
☐ Filter network using 'omics dataset'

☒ Perform comparative analysis

Add organism set

Back
Reset
Next

Figure 2.13: Specifying one or more organism set(s) to build a metabolic network(s) in MAPPS. If more than one organism sets are created, users can perform comparative analysis. Users can also choose to build a network using all pathway maps available in KEGG or select one or more pathway maps by selecting the appropriate dataset.

in silico metabolic engineering (see **Section 2.2.5**) and/or filtered by user-provided ‘omics datasets (see **Section 2.2.6**).

Metabolic network Analysis and Pathway Prediction Server

Home
Feedback
About
GuestUser

Job submission steps

1. Define job
2. Build network(s)
3. Enter job parameters
4. Review & submit job

Analysis: Pathway prediction and comparison

Show job summary

ResultsTutorials

Saved JobsReport bug

KEGG: Release 93.0+02-22, Feb 20

Data source

KEGG
Custom data

MAPPS output is dependent on the accuracy of KEGG annotations

Dataset

All pathways
Select pathways

MAPPS may not report all known pathways depending on the dataset used

Pathway set: Carbohydrate metabolism

Search pathways...

ID	Pathway Name	Pathway Set
00660	C5-Branched dibasic acid metabolism	Carbohydrate metabolism
00020	Citrate cycle (TCA cycle)	Carbohydrate metabolism
00051	Fructose and mannose metabolism	Carbohydrate metabolism
00052	Galactose metabolism	Carbohydrate metabolism
00010	Glycolysis / Gluconeogenesis	Carbohydrate metabolism
00630	Gluconate and dicarboxylate metabolism	Carbohydrate metabolism

Amino sugar and nucleotide sugar metabolism

Ascorbate and aldarate metabolism

Butanoate metabolism

C5-Branched dibasic acid metabolism

Citrate cycle (TCA cycle)

Fructose and mannose metabolism

Galactose metabolism

Glycolysis / Gluconeogenesis

Glyoxylate and dicarboxylate metabolism

Inositol phosphate metabolism

Pentose and glucuronate interconversions

Pentose phosphate pathway

Propanoate metabolism

Pyruvate metabolism

Starch and sucrose metabolism

Network building mode

Organism set(s)
Phylogeny

Phylogeny mode

Parsimony

- Fitch
- Sankoff
- Dollo
- Polymorphism

Algebraic

- Union
- Intersection
- Reaction Neighbourhood

Stochastic

- Independent edge model
- Neighbour dependent model
- Hybrid model

Phylogeny

Enter phylogeny in Newick format using KEGG organism codes

(((hsa_ptr),ggo),mcc)

Back

Reset

Next

Figure 2.14: Building metabolic network over a phylogeny. MAPPS allows user to enter phylogeny in strict Newick format using KEGG organism codes.

2.2.2.2.3 Enter job parameters

User enters the job parameters required to perform the analysis on the metabolic network(s). Users can select one or more source/target metabolites, define pathway length and other parameters for pathway prediction. Users can apply constraints at metabolite and/or reaction/enzyme and/or element level for computing pathways (**Figure 2.15**). This step is only available for pathway-based analyses.

Metabolic network Analysis and Pathway Prediction Server

Home
Feedback
About
GuestUser

Job submission steps

1. Define job
2. Build network(s)
3. Enter job parameters
4. Review & submit job

Analysis:
Pathway prediction and comparison
Show job summary

Results

Tutorials

Saved Jobs

Report bug

KEGG: Release 93.0+/02-22, Feb 20

Metabolites

Select pathway: Citrate cycle (TCA cycle)

Start metabolite(s)

ID	Name
C00026	2-Oxoglutarate

Required metabolite(s)

ID	Name
----	------

List of metabolites

ID	Name
C00022	Pyruvate
C00024	Acetyl-CoA
C00036	Oxaloacetate
C00068	Thiamin diphosphate
C00074	Phosphoenolpyruvate
C00091	Succinyl-CoA
C00149	(S)-Malate
C00158	Citrate
C00311	Isocitrate
C00417	cis-Aconitate
C05125	2-(alpha-Hydroxyethyl)thiamine diphosphate
C05379	Oxalosuccinate
C05381	3-Carboxy-1-hydroxypropyl-ThPP
C15972	Enzyme N6-(lipoyl)lysine
C15973	Enzyme N6-(dihydrolipoyl)lysine
C16254	[Dihydrolipoyllysine-residue succinyltransferase] S-a
C16255	[Dihydrolipoyllysine-residue acetyltransferase] S-a

End metabolite(s)

ID	Name
C00042	Succinate
C00122	Fumarate

Ignore metabolite(s)

ID	Name
C00001	H2O
C00002	ATP
C00003	NAD+
C00004	NADH
C00005	NADPH
C00006	NADP+
C00007	Oxygen
C00008	ADP
C00009	Orthophosphate
C00010	CoA
C00011	CO2
C00012	Peptide
C00013	Diphosphate
C00015	UDP
C00016	FAD
C00017	Protein

☐ Use default metabolic constraint(s)

Reaction/Enzyme

☒ Apply additional constraints on network(s)

Define constraint(s) on

Reactions
Enzymes

Select pathway: Citrate cycle (TCA cycle)

ID	Definition	Ignore	Required
R00014	Pyruvate + Thiamin diphosphate <=> 2-(alpha-Hydroxyethyl)thiamine diphosphate + CO2	<input type="radio"/>	<input type="radio"/>
R00268	Oxalosuccinate <=> 2-Oxoglutarate + CO2	<input type="radio"/>	<input type="radio"/>
R00341	ATP + Oxaloacetate <=> ADP + Phosphoenolpyruvate + CO2	<input type="radio"/>	<input type="radio"/>
R00342	(S)-Malate + NAD+ <=> Oxaloacetate + NADH + H+	<input type="radio"/>	<input type="radio"/>
R00344	ATP + Pyruvate + HCO3- <=> ADP + Orthophosphate + Oxaloacetate	<input type="radio"/>	<input type="radio"/>

Elements

Nitrogen	<input type="radio"/> Require	<input type="radio"/> Ignore
Oxygen	<input type="radio"/> Require	<input type="radio"/> Ignore
Phosphorus	<input type="radio"/> Require	<input type="radio"/> Ignore
Sulphur	<input type="radio"/> Require	<input type="radio"/> Ignore
Bromine	<input type="radio"/> Require	<input type="radio"/> Ignore
Manganese	<input type="radio"/> Require	<input type="radio"/> Ignore
Zinc	<input type="radio"/> Require	<input type="radio"/> Ignore

Other parameters

Connections: Reaction Class (KEGG)

Structural similarity (SIMCOMP2) cutoff: 0.0

Metabolite connectivity cutoff: ☐ Yes ☒ No

Reaction energies (eQuilibrator): ☐ Yes ☒ No

Pathway Length: 1 2 3 4 5 6 7 8 9 10


Back

Reset

Next





Figure 2.15: MAPPS interface for specifying pathway-related parameters. In addition to specifying one or more start/end metabolites, the user can also define constraints at metabolite and/or enzyme/reaction level during pathway-based analyses requiring them to be present or ignored during pathway prediction. Besides, the user can also specify which elements are required or to be ignored in the pathway prediction. The user can use metabolite pairing where all possible pairs between substrates and products of a given reaction are considered, or select KEGG reactant pairs to establish the connection between reactions and choose to incorporate structural similarity between consecutive metabolites while predicting metabolic pathways, in addition to specifying a range for the length of predicted pathways.

39



Metabolic network Analysis and Pathway Prediction Server
Home
Feedback
About
GuestUser

Job submission steps

1. Define job
2. Build network(s)
3. Enter job parameters
4. Review & submit job

 Results
 Tutorials
 Saved Jobs
 Report bug

KEGG: Release 93.0-02-22, Feb 20

 Save job

Job summary

Define Job

Job name:	pathway_job
Output format:	Graphical
Analysis:	Pathway prediction and comparison

Build Network(s)

Organism set: 1

Dataset:	Citrate cycle (TCA cycle), Galactose metabolism, Glycolysis / Gluconeogenesis
Define network on:	Organism set(s)
Build network using:	Organism-based network
Organism(s):	Escherichia coli K-12 MG1655 (ID:eco)
Network building mode:	All reactions

Organism set: 2

Build network using:	Organism-based network
Organism(s):	Brucella suis 1330 (ID:bsi)
Network building mode:	All reactions

Organism set: 3

Build network using:	Organism-based network
Organism(s):	Paenibacillus sp. JDR-2 (ID:pjd)
Network building mode:	All reactions

Job parameters


Start metabolite(s):	C00026
End metabolite(s):	C00042, C00122
Ignored metabolite(s):	C00001, C00002, C00003, C00004, C00005, C00006, C00007, C00008, C00009, C00010, C00011, C00012, C00013, C00015, C00016, C00017, C00018, C00019, C00020, C00027, C00034, C00035, C00044, C00053, C00054, C00055, C00063, C00066, C00075, C00080, C00081, C00087, C00104, C00130, C00131, C00206, C00237, C00286, C00320, C00857, C01322, C01342, C01352, C01528, C02987, C04432, C05336, C05684, C05697, C06481, C06482, C09306, C14818, C14819
Connections:	Reaction Class (KEGG)
Structural similarity:	0.0
Pathway length:	2 - 6

Submit job

Figure 2.16: Review and submit the job. The user can review the job parameters, edit the parameters for the selected analysis (if required) and submit the job.

2.2.2.2.4 Review and submit job

This is the last step in the job submission process, the user can review the job parameters and submits the job (**Figure 2.16**).



Metabolic network Analysis and Pathway Prediction Server

[Home](#)
[Feedback](#)
[About](#)
[MAPPS_User](#)

Results

(This table will refresh after 9 seconds)

Sr.	Job Name		Output Format	Submission Date/Time	Completion Date/Time
1	test	Show	graphical	2/26/2020 3:59:07 PM	2/26/2020 3:59:25 PM
2	test	Show	graphical	2/26/2020 3:49:35 PM	2/26/2020 3:49:46 PM
3	test	Show	graphical	2/26/2020 3:48:17 PM	2/26/2020 3:48:26 PM
4	sulfur_test	Download	text	2/26/2020 3:16:48 PM	2/26/2020 3:18:32 PM
5	ref_bug_test	Show	graphical	2/26/2020 2:13:28 PM	2/26/2020 2:16:07 PM
6	alpha_2_9	Show	graphical	2/26/2020 1:45:52 PM	2/26/2020 2:06:09 PM
7	ref_2_8	Show	graphical	2/26/2020 1:45:20 PM	2/26/2020 2:45:56 PM
8	test_allPws_2_9	Show	graphical	2/26/2020 12:22:40 PM	2/26/2020 4:29:30 PM
9	test_carbAA_2_9	Show	graphical	2/26/2020 12:21:54 PM	2/26/2020 1:08:31 PM
10	test_2_9_lysineDS	Show	graphical	2/26/2020 12:20:29 PM	2/26/2020 12:23:59 PM
11	alpha_CarbAA_2_10	Download	text	2/26/2020 11:17:29 AM	2/26/2020 11:26:25 AM
12	alpha_CarbAA_2_9	Download	text	2/26/2020 11:01:45 AM	2/26/2020 12:15:31 PM
13	alpha_selectAll_2_10	Download	text	2/26/2020 10:29:38 AM	2/26/2020 11:24:29 AM
14	lysineDS_2_10	Show	graphical	2/26/2020 9:50:36 AM	2/26/2020 9:50:46 AM
15	lysineDS_2_9	Show	graphical	2/26/2020 9:49:51 AM	2/26/2020 12:05:05 PM
16	alphanproteobacteria_2_10	Download	text	2/25/2020 11:10:26 PM	2/26/2020 12:09:19 AM
17	alphanproteobacteria_2_10	Show	graphical	2/25/2020 9:07:07 PM	2/25/2020 10:09:49 PM
18	alphanproteobacteria_2_9	Download	text	2/25/2020 9:05:42 PM	2/25/2020 9:20:05 PM
19	alpha_allpws_2_9	Show	graphical	2/25/2020 7:56:59 PM	2/25/2020 8:19:34 PM
20	alpha_2_10_lysineDS	Show	graphical	2/25/2020 7:54:28 PM	2/25/2020 7:54:36 PM
21	alpha_2_9	Show	graphical	2/25/2020 7:53:30 PM	2/25/2020 7:53:41 PM
22	alphanproteobacteria_1_9	Show	graphical	2/25/2020 6:07:06 PM	2/25/2020 6:29:46 PM
23	alphanproteobacteria_2_9	Show	graphical	2/25/2020 6:03:30 PM	2/25/2020 6:03:49 PM
24	alphanproteobacteria_small	Show	graphical	2/25/2020 5:35:57 PM	2/25/2020 5:43:46 PM
25	test	Show	graphical	2/25/2020 5:12:43 PM	2/25/2020 5:12:52 PM
26	alphanproteobacteria	Show	graphical	2/25/2020 4:47:14 PM	2/25/2020 5:56:19 PM
27	alpha_2_10	Show	graphical	2/25/2020 4:35:48 PM	2/25/2020 4:35:58 PM
28	ref_2_10	Show	graphical	2/25/2020 4:34:06 PM	2/25/2020 4:34:16 PM
29	ref_2_8_asp2Lys	Show	graphical	2/25/2020 12:27:44 PM	2/25/2020 3:57:38 PM
30	alphanproteobacteria_2_9	Show	graphical	2/25/2020 12:26:43 PM	2/25/2020 3:20:09 PM
31	asp2lys_test	Show	graphical	2/24/2020 4:19:39 PM	2/24/2020 4:19:49 PM
32	asp2lys_test2	Show	graphical	2/24/2020 4:16:52 PM	2/24/2020 4:17:03 PM
33	asp2lys_test	Show	graphical	2/24/2020 4:08:33 PM	2/24/2020 4:08:43 PM
34	asp2lys_ref_test	Show	graphical	2/24/2020 10:19:35 AM	2/24/2020 1:29:57 PM
35	lysine_biosynthesis_full	Show	graphical	2/19/2020 2:40:28 PM	2/19/2020 2:54:07 PM
36	lysine_biosynthesis2	Show	graphical	2/19/2020 2:39:27 PM	2/19/2020 2:39:35 PM
37	lysine_biosynthesis	Show	graphical	2/19/2020 2:36:38 PM	2/19/2020 2:36:50 PM

Figure 2.17: MAPPS interface showing list of completed jobs for registered user. User can download the results (plain text, HTML or SBML) or visualize the results (graphical output) by clicking on the link button (green color).

2.2.2.3 Results retrieval

When a user submits a job successfully, Listener (see **Section 2.2.1.4**) starts the job processing and the user will be redirected to results page if the user is signed in with a registered account, where all the jobs completed within the past 15 days can be accessed (**Figure 2.17**). The user is also notified via email when the current job is completed. If the

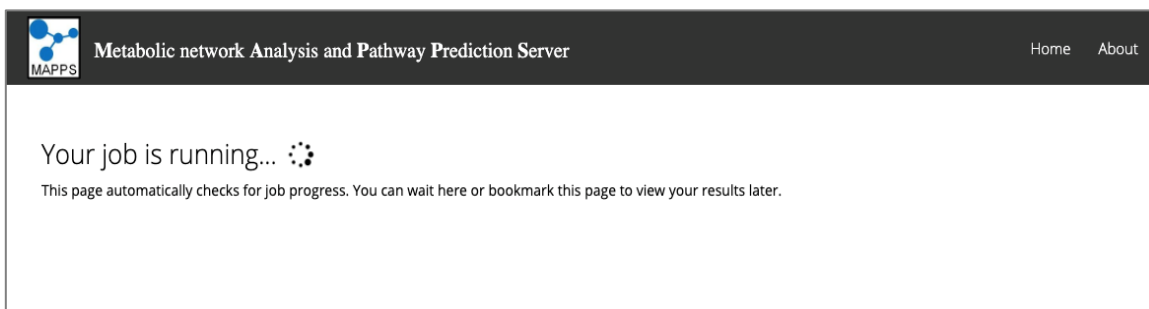


Figure 2.18: MAPPS interface showing the job status for guest users. This page refreshes automatically and checks for job progress, the user can retrieve the results by using this page link.

job is, however, submitted as a guest user, they will be redirected to the job progress page which checks the job status from the database every ten seconds (**Figure 2.18**). The user can wait for the job to complete or save the page link to access the results later. Results for the guest user are stored for three days after the job completion. MAPPS generates results in various formats including iterative graphical output, plain text, HTML and SBML.

2.2.3 MAPPS output

MAPPS offers multiple output formats including graphical, plain text, HTML, SBML. These are described below.

2.2.3.1 Graphical output

User can visualize results in a graphical format to navigate through the results easily and explore the output by manipulating the parameters. MAPPS allows the user to visualize predicted pathways as a merged subnetwork or as individual pathways sorted on pathway score. User can filter the predicted pathways by length, delete one or more metabolites and/or reactions to analyze the effects of reported pathways, visualize metabolite and reaction details and provides single-click access to related information in relevant

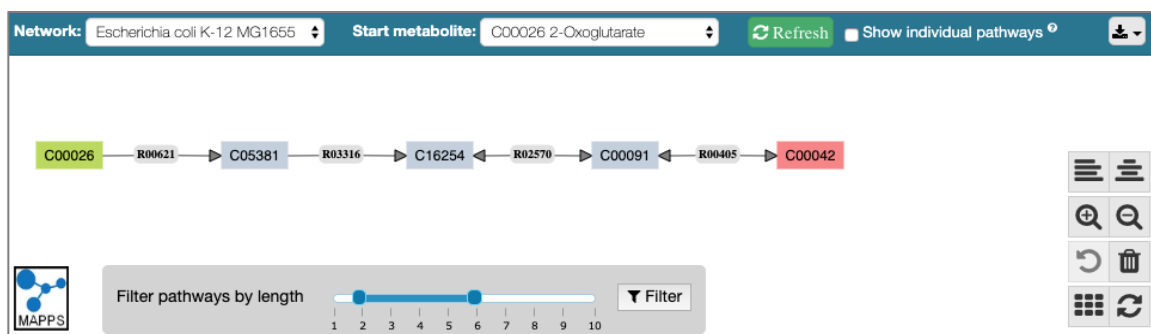


Figure 2.19: An example of the interactive graphical output for pathway prediction. Green box represents the start metabolite, blue boxes show the intermediate metabolites, and the red box shows the end metabolite. Arrows between the metabolites correspond to the reactions. The graphical result can also be saved in PNG and PDF formats.

```

Job Name: test job
Submission Date Time: 6/24/2020 12:24:57 PM
Completion Date Time: 6/24/2020 12:25:05 PM

DATASET
KEGG Pathway Maps: Glycolysis / Gluconeogenesis [MAP00010], Citrate cycle (TCA cycle) [MAP00020], Galactose metabolism [MAP00052]

OUTPUT DETAILS
Output Format: TEXT

JOB DETAILS
Analysis: Pathway prediction
Comparison Mode: Individual

JOB PARAMETERS(S)
ORGANISM SET: 1
Organism(s): Escherichia coli K-12 MG1655 [eco]
Network building mode: Union
ORGANISM SET: 2
Organism(s): Brucella suis 1330 [bsi]
Network building mode: Union
ORGANISM SET: 3
Organism(s): Paenibacillus sp. JDR-2 [pjd]
Network building mode: Union

PATHWAY PARAMETERS
Start Metabolite(s): 2-Oxoglutarate [C00026]
End Metabolite(s): Succinate [C00042], Fumarate [C00122]
Ignore Metabolite(s): H2O [C00001], ATP [C00002], NAD+ [C00003], NADH [C00004], NADPH [C00005], NADP+ [C00006], Oxygen [C00007], ADP [C00008], Orthophosphate [C00009], CoA [C00010], CO2 [C00011], Peptide [C00012], Diphosphate [C00013], UDP [C00015], FAD [C00016], Protein [C00017], Pyridoxal phosphate [C00018], S-Adenosyl-L-methionine [C00019], AMP [C00020], Hydrogen peroxide [C00027], Manganese [C00034], GDP [C00035], GTP [C00044], 3'-Phosphoadenylyl sulfate [C00053], Adenosine 3',5'-bisphosphate [C00054], CMP [C00055], CTP [C00063], tRNA [C00066], UTP [C00075], H+ [C00080], ITP [C00081], Sulfur [C00087], IDP [C00104], IMP [C00130], dATP [C00131], dADP [C00206], CD [C00237], dGTP [C00286], Thiosulfate [C00320], Deamino-NAD+ [C00857], RK [C01322], NH4+ [C01342], FADH2 [C01352], Hydrogen selenide [C01528], L-Glutamyl-tRNA(Glu) [C02587], tRNA containing 6-isopentenyladenosine [C04432], Selenomethionyl-tRNA(Met) [C05336], Selenite [C05684], Selenate [C05697], L-Seryl-tRNA(Sec) [C06481], L-Selenocysteinyl-tRNA(Sec) [C06482], Sulfur dioxide [C09306], Fe2+ [C14818], Fe3+ [C14819]
Connection Option: REPAIR
Pathway Length Range: 2-6
Metabolic Similarity Cutoff: 0
Return Value: Pathways

Start Metabolite Id Start Metabolite Network Pathway Length Pathway Score Reversible Reaction(s)
C00026 2-Oxoglutarate Escherichia coli K-12 MG1655 [eco] 2-Oxoglutarate [C00026] --(R00621 - 1.2.4.2)--> 3-Carboxy-1-hydroxypropyl-ThPP [C05381] --(R03316 - 1.2.4.2)--> [Dihydrodipicolyllysine-residue succinyltransferase] S-succinylidihydrodipicolyllysine [C16254] <--(R02570 - 2.3.1.61)--> Succinyl-CoA [C00091] <--(R00405 - 6.2.1.5)--> Succinate [C00042] 4 0.643217625 2

C00026 2-Oxoglutarate Brucella suis 1330 [bsi] 2-Oxoglutarate [C00026] --(R00621 - 1.2.4.2)--> 3-Carboxy-1-hydroxypropyl-ThPP [C05381] --(R03316 - 1.2.4.2)--> [Dihydrodipicolyllysine-residue succinyltransferase] S-succinylidihydrodipicolyllysine [C16254] <--(R02570 - 2.3.1.61)--> Succinyl-CoA [C00091] <--(R00405 - 6.2.1.5)--> Succinate [C00042] 4 0.643217625 2

C00026 2-Oxoglutarate Paenibacillus sp. JDR-2 [pjd] 2-Oxoglutarate [C00026] --(R00621 - 1.2.4.2)--> 3-Carboxy-1-hydroxypropyl-ThPP [C05381] --(R03316 - 1.2.4.2)--> [Dihydrodipicolyllysine-residue succinyltransferase] S-succinylidihydrodipicolyllysine [C16254] <--(R02570 - 2.3.1.61)--> Succinyl-CoA [C00091] <--(R10343 - 2.8.3.18)--> Succinate [C00042] 4 0.643217625 2

C00026 2-Oxoglutarate Paenibacillus sp. JDR-2 [pjd] 2-Oxoglutarate [C00026] <--(R01197 - 1.2.7.11,1.2.7.3)--> Succinyl-CoA [C00091] <--(R00405 - 6.2.1.5)--> Succinate [C00042] 2 0.6300925 2

C00026 2-Oxoglutarate Paenibacillus sp. JDR-2 [pjd] 2-Oxoglutarate [C00026] --(R00621 - 1.2.4.2)--> 3-Carboxy-1-hydroxypropyl-ThPP [C05381] --(R03316 - 1.2.4.2)--> [Dihydrodipicolyllysine-residue succinyltransferase] S-succinylidihydrodipicolyllysine [C16254] <--(R02570 - 2.3.1.61)--> Succinyl-CoA [C00091] <--(R00405 - 6.2.1.5)--> Succinate [C00042] 4 0.643217625 2

Pathway Statistics
Length Count
4 4
2 1
Total reversible reactions:10

```

Figure 2.20: An example of plain text output for pathway prediction of MAPPS. The output is a descriptive text format which contains information about the job parameters, predicted pathways, pathway scores, number of pathways predicted of each length, and number of reversible reactions in plain text.

databases. **Figure 2.19** shows an example of the graphical output of pathway prediction analysis where the start metabolite is shown as green rectangles, intermediary metabolites


```

1 <?xml version='1.0' encoding='utf-8' standalone='no' ?><sbml xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:dcterms="http://purl.org/dc/terms/" xmlns:cvCard="http://www.w
2
3 <model id="Escherichia_coli_K-12_M61655" >
4 <listOfCompartments> <compartment id="c1" constant="false" /></listOfCompartments>
5 <listOfSpecies>
6 <species id="C00066" name="2-Oxoglutarate" compartment="c1" hasOnlySubstanceUnits="false" boundaryCondition="false" initialConcentration="1" constant="false" />
7 <species id="C00068" name="Thiamin diphosphate" compartment="c1" hasOnlySubstanceUnits="false" boundaryCondition="false" initialConcentration="1" constant="false" />
8 <species id="C05381" name="3-Carboxy-1-hydroxypropyl-ThPP" compartment="c1" hasOnlySubstanceUnits="false" boundaryCondition="false" initialConcentration="1" constant="false" />
9 <species id="C15972" name="Enzyme M6-(lipoyl)lysine" compartment="c1" hasOnlySubstanceUnits="false" boundaryCondition="false" initialConcentration="1" constant="false" />
10 <species id="C16254" name="(Dihydrolipoyl)lysine-residue succinyltransferase" S-succinyl-dihydrolipoyllysine" compartment="c1" hasOnlySubstanceUnits="false" boundaryCondition="false" initialConcentration="1" constant="false" />
11 <species id="C00091" name="Succinyl-CoA" compartment="c1" hasOnlySubstanceUnits="false" boundaryCondition="false" initialConcentration="1" constant="false" />
12 <species id="C15973" name="Enzyme M6-(dihydrolipoyl)lysine" compartment="c1" hasOnlySubstanceUnits="false" boundaryCondition="false" initialConcentration="1" constant="false" />
13 <species id="C00042" name="Succinate" compartment="c1" hasOnlySubstanceUnits="false" boundaryCondition="false" initialConcentration="1" constant="false" />
14 </listOfSpecies>
15 <listOfReactions>
16 <reaction id="R00621" name="R00621" reversible="false" fast="false">
17 <listOfReactants>
18 <speciesReference species="C00066" stoichiometry="1" constant="true"/>
19 <speciesReference species="C00068" stoichiometry="1" constant="true"/>
20 </listOfReactants>
21 <listOfProducts>
22 <speciesReference species="C05381" stoichiometry="1" constant="true"/>
23 </listOfProducts>
24 </reaction>
25 <reaction id="R03316" name="R03316" reversible="false" fast="false">
26 <listOfReactants>
27 <speciesReference species="C05381" stoichiometry="1" constant="true"/>
28 <speciesReference species="C15972" stoichiometry="1" constant="true"/>
29 </listOfReactants>
30 <listOfProducts>
31 <speciesReference species="C16254" stoichiometry="1" constant="true"/>
32 <speciesReference species="C00068" stoichiometry="1" constant="true"/>
33 </listOfProducts>
34 </reaction>
35 <reaction id="R02570" name="R02570" reversible="true" fast="false">
36 <listOfReactants>
37 <speciesReference species="C00091" stoichiometry="1" constant="true"/>
38 <speciesReference species="C15973" stoichiometry="1" constant="true"/>
39 </listOfReactants>
40 <listOfProducts>
41 <speciesReference species="C16254" stoichiometry="1" constant="true"/>
42 </listOfProducts>
43 </reaction>
44 <reaction id="R00405" name="R00405" reversible="true" fast="false">
45 <listOfReactants>
46 <speciesReference species="C00042" stoichiometry="1" constant="true"/>
47 </listOfReactants>
48 <listOfProducts>
49 <speciesReference species="C00091" stoichiometry="1" constant="true"/>
50 </listOfProducts>
51 </reaction>
52 </listOfReactions>
53 </model>
54 <model id="Brucella_suis_1330" >
55 <listOfCompartments> <compartment id="c1" constant="false" /></listOfCompartments>
56 <listOfSpecies>
57 <species id="C00066" name="2-Oxoglutarate" compartment="c1" hasOnlySubstanceUnits="false" boundaryCondition="false" initialConcentration="1" constant="false" />
58 <species id="C00068" name="Thiamin diphosphate" compartment="c1" hasOnlySubstanceUnits="false" boundaryCondition="false" initialConcentration="1" constant="false" />
59 <species id="C15972" name="3-Carboxy-1-hydroxypropyl-ThPP" compartment="c1" hasOnlySubstanceUnits="false" boundaryCondition="false" initialConcentration="1" constant="false" />
60 <species id="C15973" name="Enzyme M6-(lipoyl)lysine" compartment="c1" hasOnlySubstanceUnits="false" boundaryCondition="false" initialConcentration="1" constant="false" />
61 <species id="C16254" name="(Dihydrolipoyl)lysine-residue succinyltransferase" S-succinyl-dihydrolipoyllysine" compartment="c1" hasOnlySubstanceUnits="false" boundaryCondition="false" initialConcentration="1" constant="false" />
62 <species id="C00091" name="Succinyl-CoA" compartment="c1" hasOnlySubstanceUnits="false" boundaryCondition="false" initialConcentration="1" constant="false" />
63 <species id="C15973" name="Enzyme M6-(dihydrolipoyl)lysine" compartment="c1" hasOnlySubstanceUnits="false" boundaryCondition="false" initialConcentration="1" constant="false" />
64 <species id="C00042" name="Succinate" compartment="c1" hasOnlySubstanceUnits="false" boundaryCondition="false" initialConcentration="1" constant="false" />
65 <species id="C00033" name="Acetate" compartment="c1" hasOnlySubstanceUnits="false" boundaryCondition="false" initialConcentration="1" constant="false" />
66 <species id="C00024" name="Acetyl-CoA" compartment="c1" hasOnlySubstanceUnits="false" boundaryCondition="false" initialConcentration="1" constant="false" />
67 </listOfSpecies>
68 <listOfReactions>
69 <reaction id="R00621" name="R00621" reversible="false" fast="false">

```

Figure 2.22: An example of SBML output for pathway prediction. Each organism set is represented as a model, participating metabolites are represented as listOfSpecies and reactions as listOfReactions. SBML file can be used for further analysis or visualization in any other SBML-compatible tool.

2.2.3.4 SBML

MAPPS also generate output in the Systems Biology Markup Language (SBML) format, which can be uploaded to MAPPS or other software packages such as Cytoscape (Shannon et al., 2003) for further analysis. An example of SBML output of the pathway prediction analysis is shown in **Figure 2.22**.

2.2.4 Analyses provided by MAPPS

MAPPS is a versatile tool, which offers a distinct range of analyses for metabolic pathway prediction and comparison and network analysis (see **Figure 2.5**). These analyses are

described in the subsequent sections. Tutorials highlighting various features of MAPPS are available at <https://mapps.lums.edu.pk/tool/docs.aspx>.

2.2.4.1 Pathway prediction and comparison

MAPPS allows the user to predict metabolic pathways over one or more organisms sets based on KEGG reference network or custom data (see **Section 2.2.2.2.2**), specify multiple organism sets and select more than one source/target metabolite(s) to in a single job. Furthermore, MAPPS provides an option to perform a pathway-based comparison between two or more organism sets to identify pathways or reactions involved in the predicted pathways that are present in one organism set but absent in other organism sets and vice versa. MAPPS also predicts metabolic pathways on a user-defined phylogeny. In this mode, users can perform pathway prediction on leaf nodes as well as ancestral nodes predicted using one of many available phylogeny building modes (see **Section 2.2.4.4**). Pathway prediction on a phylogeny can help in identifying the functional differences at various levels of the phylogeny and provide clues about the metabolic evolution of various species (Mithani et al., 2011).

MAPPS computes pathways using a depth-first search of metabolic networks and reports all possible pathways between the source and target metabolites. A pathway between two metabolites is defined as a connected sequence of reactions such that the product of one reaction acts as a substrate in the next reaction that satisfies user-defined constraints. User can select KEGG Reaction Class (RCLASS) (Kanehisa et al., 2017) which contains information relating to the main reactant pairs of KEGG reactions to compute connections between metabolites thus allowing pathway searches to be optimized (Kanehisa, 2006) or use metabolite pairing where all possible pairs between substrates and

products of a given reaction are considered to establish the connection between reactions. While predicting metabolic pathways, MAPPS takes into account reaction directions and computes all possible pathways between the source and target metabolites. MAPPS does not allow metabolites or reactions to be repeated in a pathway, thus avoiding cycles and reports only direct routes between metabolites. Unlike *Rahnuma*, it allows pathways through metabolites which are used as substrates in one or more reactions enabling it to report pathways missed by *Rahnuma* (see **Section 2.1**). While submitting a pathway prediction job, users must specify minimum and maximum lengths of the pathways to be reported. All pathways outside of the specified range are ignored. Pathways up to 10 reactions in length can be computed by MAPPS.

To refine the pathway search, users can define additional constraints on metabolites and/or reactions/enzymes during pathway prediction (**Figure 2.15**). At the metabolite level, users can specify a list of metabolites to be avoided or required during pathway prediction. Pathways passing through metabolites which are designated to be ignored are not reported. By default, ubiquitous metabolites such as ATP, AMP, O₂, H₂O and CO₂ are ignored during pathway prediction. The complete list of metabolites ignored by default in MAPPS is provided in **Appendix B**. If one or more metabolites are stipulated to be required, then only those pathways that contain these required metabolite(s) are reported. In addition, MAPPS also provides an option to filter metabolites based on their connectivity scores (see **Section 2.2.7**) since it has been previously reported that metabolite filtering by assigning weights to metabolites based on their connectivity in the metabolic network narrows down the search space and helps in reporting biologically relevant pathways (Croes et al., 2006). To this end, user can specify a metabolite connectivity cut-

off value to filter compounds based on their degree in the underlying metabolic network. Besides metabolites, users can also specify whether one or more reactions/enzymes must be present or are to be avoided during pathway prediction. To enhance the reporting of biologically meaningful pathways, MAPPS also provides an option of filtering intermediate metabolites by presence or absence of constituent elements including nitrogen, oxygen, phosphorus, sulfur, bromine, manganese and zinc to enable element tracing while computing pathways between source and target metabolites. The user has the flexibility to define one or more chemical elements as required (element(s) must be present in all intermediate metabolites of the predicted pathway) or ignored (element(s) must not be present in the intermediary metabolites). By default, all elements are regarded as optional (**Figure 2.15**). Finally, users can also choose to incorporate structural similarity between consecutive metabolites while predicting metabolic pathways. Structural similarity score are calculated using SIMCOMP2 (Hattori et al., 2010) and a user-specified cutoff is used to filter out intermediary metabolites during pathway prediction. The pseudocode for the pathway prediction algorithm is given in **Box 2.1**.

2.2.4.2 Metabolite reachability

The reachability of a metabolite is defined as a set of metabolites reachable from the start metabolite in given number of steps and is very useful in identifying the essential metabolites being produced from pre-cursor metabolites in a given metabolic network (Karp et al., 2009). This analysis is particularly useful in focusing on nutrient assimilation pathways (Ferguson et al., 2010; Mithani et al., 2011; Koprivova and Kopriva, 2014) since it allows users to analyze the connectedness and scope of the metabolites in the metabolic network(s) under study. Traditionally, ^{13}C tracer experiments have been used to identify

```

Input:
n = metabolic network of the current organism set
Sm = List of start metabolites
Im = List of ignore metabolites
Ie = List of ignore elements
Rm = List of required metabolites
Rr = List of required reactions
Rc = List of required enzymes
Re = List of required elements
Tm = List of target metabolites
Output:
P = List of predicted pathways
Algorithm:
Let Rs denotes the list of reactions of metabolic network n in which s ∈ Sm is acting as
a substrate
for each reaction r ∈ Rs do
    Let p denotes the predicted pathway
    If r is present in p then
        continue to next r
    Let Cs denotes the list of outgoing connections from s in r
    for each connection c ∈ Cs do
        Let m denotes the end metabolite of connection c
        Let Lp denotes the length of p
        Let Lmin denotes minimum length cutoff and Lmax denotes maximum length
        cutoff
        if Tm contains m then
            if Lmin ≤ Lp ≤ Lmax and p contains any Rm, Rr and Rc then
                calculatePathwayScore(p)
                save the pathway p in P
                go to next start metabolite
            else
                if m is already present in p as intermediate node then
                    continue to next c
                if m is present in Im then
                    continue to next c
                if m contains any element of Ie then
                    continue to next c
                if m does not contain any element of Re then
                    continue to next c

                Let Xsim denotes the similarity score between m and s
                Let Lsim denotes the similarity score cutoff defined by user

                if Xsim < Lsim then
                    continue to next c

                Add m as new node in the pathway p

                CALL pathwayPrediction function with s = m

```

Box 2.1: Pseudocode for pathway prediction algorithm

intermediary metabolites produced from a given start metabolite with metabolomic data being used in the recent years (Buescher et al., 2015; Fuhrer et al., 2005; Zhang et al., 2016). However, this has primarily been done in prokaryotes or for subsets of metabolic networks due to high experimental costs for large networks (Sauer, 2006).

MAPPS provides comprehensive functionality for users to identify metabolites reachable from one or more start metabolites in a given range of steps (see **Figure 2.23**)

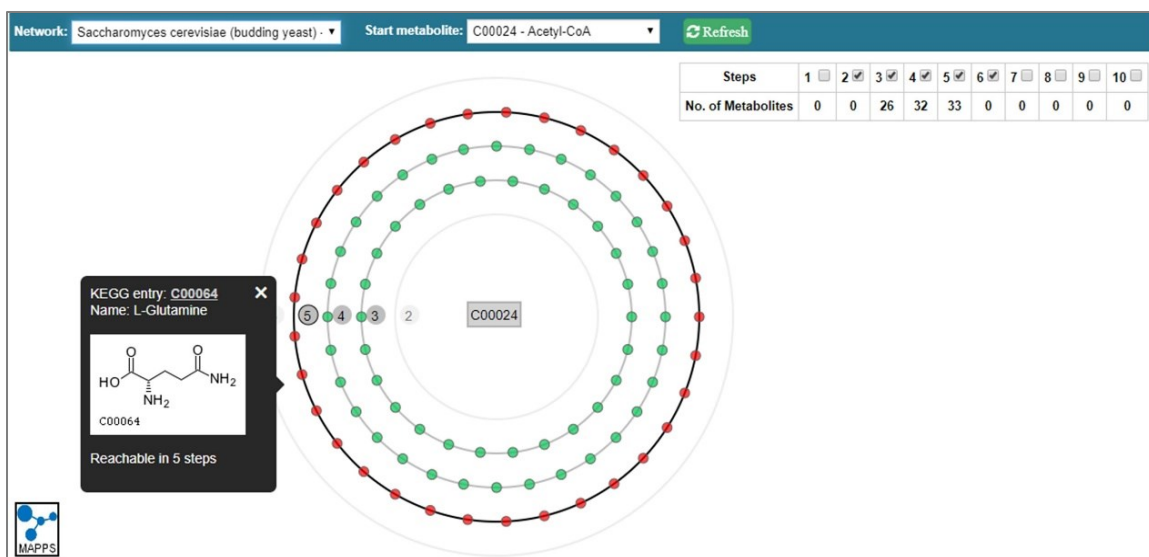


Figure 2.23: Graphical output for metabolic reachability. Green circles represent metabolites that are reachable from the selected metabolite shown in the middle (C00024: Acetyl-CoA). Gray rings represent the number of steps/reactions (shown in grey circles) required to produce the respective metabolites. Users can also select a particular number of steps by clicking on the gray circle. Metabolites reachable in the selected number of steps are shown in red color. Users can turn on or off one or more steps using the checkboxes at the top right corner.

taking into account all the constraints available for pathway prediction (see **Section 2.2.4.1**). Like pathway prediction, user can build up to 10 metabolic networks in a single job and perform reachability analysis simultaneously in addition to applying constraints at metabolite, reaction, enzyme and element levels for KEGG based metabolic networks (see **Figure 2.24**). Unlike pathway prediction, no end metabolite is required for this analysis. For each metabolic network, MAPPS searches the network space starting from the source metabolite and extracts the products of all those reactions in which the source metabolite is participating and the list of reachable compounds is returned to the user as output. The pseudocode of metabolic reachability is given in **Box 2.2**.

2.2.4.3 Metabolite-specific reactions

Metabolite-specific reactions are defined as reactions that are involved in pathways from only one of the given start metabolites (Mithani et al., 2009a). Exploring a metabolic


```

Input:
N = list of the metabolic network of organism set
Sm = List of start metabolites
Ie = List of ignore metabolites
Ie = List of ignore elements
Re = List of required elements

Output:
Mreach = List of reachable compounds

Algorithm:
Let Lmin and Lmax denotes the reachability range defined by the user
Let Rs denotes the list of reactions of metabolic network n in which s is acting as
substrate
Let Mproducts denotes the list of product metabolites of each step
for each metabolic network n ∈ N do
    for each start metabolite s ∈ Sm do
        step = 1
        Mproducts = FindProducts(s)
        while step is less than or equal to Lmax do
            for each product metabolite b ∈ Mbranches do
                append results of FindProducts(b) to Mproducts

            if step is less than or equal to Lmax and greater than or equal
            to Lmin then
                for each product metabolite b ∈ Mbranches do
                    if b contains any element of Ie then
                        continue to next b
                    if b does not contain any element of Re then
                        continue to next b
                    Let Xsim denotes the similarity score between b and s
                    Let Lsim denotes the similarity score cutoff defined by
                    user
                    if Xsim < Lsim then
                        continue to next b
                    add b to the list of reachable compounds, Mreach, for
                    start metabolite s

                step=step+1

```

Box 2.2: Pseudocode for metabolite reachability algorithm

in the pathways from only one start metabolite but absent from all others. The pseudocode for identifying metabolite-specific reactions is given in **Box 2.3**.

2.2.4.4 Ancestral network-building/comparison

Metabolic networks are under a process of continuous evolution. The availability of genomes for many closely related species offers the possibility of tracing metabolic evolution on a phylogeny relating the genomes to understand the evolutionary processes and constraints that affect the evolution of metabolic networks (Mithani et al., 2009a, 2011). MAPPS provides an interface for studying metabolic evolution where the user can build and compare metabolic networks at internal nodes of phylogeny using various

```

Input:
N = List of metabolic networks
Sm = List of start metabolites
Im = List of ignore metabolites
Ie = List of ignore elements
Rm = List of required metabolites
Rr = List of required reactions
Rc = List of required enzymes
Re = List of required elements
Tm = List of target metabolites
Output:
E = List of metabolite-specific reactions

Algorithm:
Let ExRxn denotes the list of reactions extracted from predicted pathways
Let n denotes a metabolic network of an organism set
for each network  $n \in N$  do
    for each substrate  $s \in S_m$  do
        CALL predictPathways(n)
        Let Ps denotes the pathways predicted from start metabolite s
        Let Rp denotes the list of reactions extracted from predicted pathways
        for start metabolite s

        for each pathway  $p \in P_s$  do
            extractReactions(p)
            Let r denotes the list of reactions extracted from pathway p
            Add r in Rp
            Save Rp in ExRxn

        for each Rp in ExRxn do
            save reactions  $r_x$  to E which are only present in Rp

```

Box 2.3: Pseudocode for identifying metabolite-specific reactions

methods (**Figure 2.25**). Various phylogeny building modes are being used at the sequence level, but their application is not much exploited at the network level (Felsenstein, 2004). MAPPS allows user to utilize these existing methods for metabolic networks. This includes maximum parsimony and its variants (Sankoff, Dollo and Polymorphism parsimony), algebraic methods (union, intersection and reaction neighborhood) (Forst et al., 2006; Mithani et al., 2009a) distance-based methods (UPGMA and neighbor-joining) (Felsenstein, 2004), and stochastic models of metabolic evolution (Mithani et al., 2010). For parsimony and its variants, algebraic methods and stochastic models, MAPPS takes a user-defined phylogeny as input and builds metabolic networks at the internal nodes of the phylogeny whereas for the distance-based method, MAPPS takes three or more Organism Sets as input instead of phylogeny and constructs a phylogeny based on metabolic networks

Metabolic network Analysis and Pathway Prediction Server

Home Feedback About GuestUser

Job submission steps

1. Define job
2. Build network(s)
 3. Enter job parameters
 4. Review & submit job

Analysis: Ancestral network building/comparison
 Show job summary

Results Tutorials
 Saved Jobs Report bug

KEGG: Release 93.0+02-22, Feb 20

Data source
☒ KEGG ☐ Custom data
MAPPS output is dependent on the accuracy of KEGG annotations

Dataset
☒ All pathways ☐ Select pathways

Phylogeny mode

Parsimony <input type="radio"/> Fitch <input type="radio"/> Sankoff <input type="radio"/> Dollo <input type="radio"/> Polymorphism	Algebraic <input checked="" type="radio"/> Union <input type="radio"/> Intersection <input type="radio"/> Reaction Neighbourhood	Distance-based <input type="radio"/> UPGMA <input type="radio"/> Neighbour joining	Stochastic <input type="radio"/> Independent edge model <input type="radio"/> Neighbour dependent model <input type="radio"/> Hybrid model
---	--	---	--

☒ Compare two phylogeny modes?

Phylogeny mode for comparison

Parsimony <input type="radio"/> Fitch <input type="radio"/> Sankoff <input type="radio"/> Dollo <input type="radio"/> Polymorphism	Algebraic <input type="radio"/> Union <input checked="" type="radio"/> Intersection <input type="radio"/> Reaction Neighbourhood	Distance-based <input type="radio"/> UPGMA <input type="radio"/> Neighbour joining	Stochastic <input type="radio"/> Independent edge model <input type="radio"/> Neighbour dependent model <input type="radio"/> Hybrid model
---	--	---	--

Phylogeny
 Enter phylogeny in Newick format using [KEGG organism codes](#)
 Example: (((hsa,ptr),ggo),mcc)

Back Reset Next

Figure 2.25: MAPPS interface for ancestral network building. Users can choose from multiple phylogeny modes, and also can compare two different phylogeny modes.

and then build metabolic networks at the internal nodes of the phylogeny. Metabolic networks at the leaf nodes can be built using KEGG data using all or a subset of KEGG pathways or using the custom data provided by the user. MAPPS also allows the user to compare two ancestral metabolic network building modes; for example, metabolic networks built at internal nodes using union and intersection modes can be compared and analyzed (**Figure 2.25**). It is important to note that comparing ancestral networks is different from pathway-based comparison on ancestral networks (see **Section 2.2.4.1**) as the latter is pathway specific and is only able to identify pathway-specific differences between given source/target metabolites on the given phylogeny whereas the former provides an overall comparison of the metabolic networks in terms of reactions present or absent at various levels of the phylogeny under different phylogenetic modes.

Metabolic network Analysis and Pathway Prediction Server

Home Feedback About GuestUser

Job submission steps

1. Define job
2. Build network(s)
3. Enter job parameters
4. Review & submit job

Analysis:
Network enumeration/comparison
Show job summary

Results Tutorials
Saved Jobs Report bug
KEGG: Release 93.0-102-22, Feb 20

Data source
☒ KEGG ☐ Custom data
 MAPPS output is dependent on the accuracy of KEGG annotations

Dataset
☒ All pathways ☐ Select pathways

Organism set(s)

Organism set 1
 Use ☒ Organism-based network ☐ Reference network ☐ Custom network (KEGG-based)
 Search Box Taxonomy View
 Escherichia coli K-12 MG1655 (ID:eco) Search organism...
 Build network using: All reactions
☐ Perform *in silico* metabolic engineering
☐ Filter network using 'omics dataset

Organism set 2
 Use ☒ Organism-based network ☐ Reference network ☐ Custom network (KEGG-based)
 Search Box Taxonomy View
 Escherichia coli K-12 DH10B (ID:ecd) Search organism...
 Build network using: All reactions
☐ Perform *in silico* metabolic engineering
☐ Filter network using 'omics dataset
 + Add organism set

Parameters
☒ Enumerate ☐ Compare

Figure 2.26: MAPPS interface for network enumeration/comparison. Users can build networks on multiple organism sets to enumerate the metabolic networks or compare reactions in the metabolic networks.

2.2.4.5 Network enumeration/comparison

The network enumeration and comparison options in MAPPS allow users to enumerate and/or compare one or more metabolic networks (**Figure 2.26**). Network enumeration reports the reactions present in one or more organism set(s) along with the metabolites involved in these reactions in one of the allowed output formats. This option provides users an opportunity to exploit functionalities offered by MAPPS for generating input data for other tools, for example by performing *in silico* metabolic engineering on a metabolic network created using KEGG data and exporting the resulting network in SBML format for visualization and topological analysis in Cytoscape (Shannon et al., 2003).

Network comparison compares reactions present between metabolic networks built over two or more organism set(s) irrespective of their involvement in a particular metabolic pathway. If two organism sets are provided then a standard comparison is performed, which identifies reactions present or absent in their respective metabolic networks whereas for more than two organism sets an all but one comparison is performed, which identifies reactions present (or absent) in only one organism set but absent (or present) in all the others.

2.2.4.6 Metabolic similarity analysis

Metabolic network comparison is a powerful method in comparative genomics providing insights into the characteristic metabolic features of organisms under study (Mithani et al., 2011). By grouping the organisms based on their metabolic capabilities, specific hypotheses relating to specialization of metabolic networks can be generated which can then be experimentally tested in the lab. MAPPS offers metabolic similarity analysis by allowing users to perform agglomerative hierarchical clustering of metabolic networks of three or more organism sets based on KEGG data only (**Figure 2.27**). Hierarchical clustering is an unsupervised machine learning technique that iteratively groups the data into clusters based on a similarity measure (Sarle et al., 1990). In MAPPS, metabolic networks can be clustered based on their similarity between reactions, enzymes or metabolic pathways between a given source and target metabolites. The user can choose from three different clustering methods, single, average or maximum linkage method enabling users to explore the effects of selecting different methods on the resulting dendrogram. Single linkage selects the clusters containing the closest pair of elements for joining together, whereas maximum linkage selects the clusters with the farthest pair of

Metabolic network Analysis and Pathway Prediction Server

Home Feedback About GuestUser

Job submission steps

1. Define job
2. Build network(s)
3. Enter job parameters
4. Review & submit job

Analysis:
Metabolic similarity analysis
Show job summary

Results Tutorials
Saved Jobs Report bug

KEGG: Release 93.0+ / 02-22, Feb 20

Data source
☒ KEGG ☐ Custom data
 ⓘ MAPPS output is dependent on the accuracy of KEGG annotations

Dataset
☐ All pathways ☒ Select pathways
 ⓘ MAPPS may not report all known pathways depending on the dataset used
 Pathway set: All Pathways Search pathways...

ID	Pathway Name	Pathway Set
00525	Acarbose and validamycin biosynthesis	Biosynthesis of other secondary metabolites
01058	Acridone alkaloid biosynthesis	Biosynthesis of other secondary metabolites
00254	Aflatoxin biosynthesis	Biosynthesis of other secondary metabolites
00250	Alanine, aspartate and glutamate metabolism	Amino acid metabolism
00592	alpha-Linolenic acid metabolism	Lipid metabolism
00630	Amino sugar and nucleotide sugar metabolism	Carbohydrate metabolism

Metabolic similarity parameters
 Analysis mode: ☒ Pathways ☐ Reactions ☐ Enzymes
 Linkage mode: ☒ Single ☐ Average ☐ Maximum

Organism set(s)

Organism set 1
 Use ☒ Organism-based network ☐ Reference network ☐ Custom network (KEGG-based)
 Search Box Taxonomy View
 Search organism...
 Build network using: All reactions
 — Perform *in silico* metabolic engineering
 — Filter network using 'omics dataset

Organism set 2
 Use ☒ Organism-based network ☐ Reference network ☐ Custom network (KEGG-based)
 Search Box Taxonomy View
 Search organism...
 Build network using: All reactions
 — Perform *in silico* metabolic engineering
 — Filter network using 'omics dataset

Organism set 3
 Use ☒ Organism-based network ☐ Reference network ☐ Custom network (KEGG-based)
 Search Box Taxonomy View
 Search organism...
 Build network using: All reactions
 — Perform *in silico* metabolic engineering
 — Filter network using 'omics dataset

+ Add organism set

Figure 2.27: MAPPS interface for metabolic similarity analysis. MAPPS only allows KEGG as data source for metabolic similarity analysis and it requires at least three organism sets to perform this analysis.

elements. Average linkage, on the other hand, combines the clusters with a minimum average distance between all pairs of elements present in the clusters. These methods are used to calculate distances between the metabolic networks based on the occurrence of reactions, enzymes or pathways from specified source and target metabolites in the given networks. Subsequently, the metabolic networks using the distance matrix and are clustered together hierarchically and output is in the form of a dendrogram (see **Figure 2.28**).

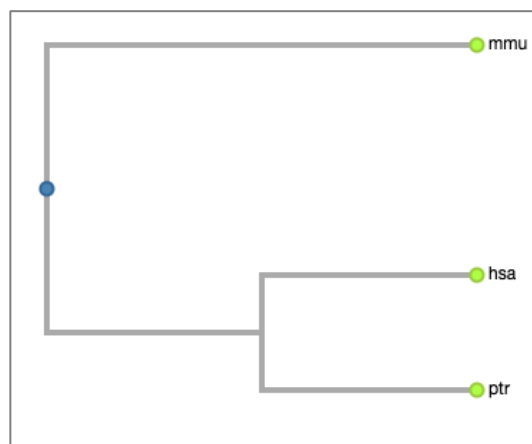


Figure 2.28: An example dendrogram resulting from metabolic similarity analysis between three organisms. MAPPS performs agglomerative hierarchical clustering of three or more organism sets based on reactions, enzymes or pathways between one or more source and target metabolites.

2.2.4.7 Host-microbe interaction

The interaction between host and microbe sharing an environment leads to the sharing of metabolites and can result in emergent pathways, for example, as shown by a detailed analysis of co-metabolism among host and its commensal microbe in a recent study (Heinken et al., 2013). MAPPS provides a methodical approach to study interactions between different organisms, such as host and microbe, at a metabolic level.

Host-microbe interaction analysis in MAPPS takes two organism sets, host set and microbe set (**Figure 2.29**), and identifies emergent pathways in the combined (host & microbe) metabolic network. MAPPS uses the user-defined parameters to build host metabolic network, microbe metabolic network, and a combined metabolic network by merging the reactions and metabolites of both host and microbe metabolic networks. Metabolic pathways are computed from the given source metabolite(s) to target metabolite in all the three metabolic networks by taking into account all the constraints available for pathway prediction (see **Section 2.2.4.1**). By systematically comparing the predicted

Figure 2.29: MAPPS job submission interface for host-microbe interaction. User can study emergent pathways resulting due to host-microbe interaction on KEGG or custom data. Two organism sets, one for the host and the other for the microbe, are required for this analysis.

pathways, MAPPS extracts those pathways which are predicted to be present in the combined network but absent in host network and microbe network. These pathways are termed as emergent pathways, which are formed due to the interaction of host and microbe metabolic networks. In addition to reporting emergent pathways, MAPPS also reports the metabolic pathways predicted in the host and the microbe networks thus allowing the user to study and compare metabolic pathways in the original as well as the combined networks (Figure 2.30). The pseudocode of host-microbe interaction is given in Box 2.4.

2.2.4.8 Potential drug targets

Identification of potential drug targets in a metabolic network is an analysis that is useful from a drug discovery point of view. A number of studies have reported key metabolic enzymes as potential drug targets since they uniquely produce or consume a metabolite,

```

Input:
 $N_h$  = Host network
 $N_m$  = Microbe network
 $S_m$  = List of start metabolites
 $I_m$  = List of ignore metabolites
 $I_e$  = List of ignore elements
 $R_m$  = List of required metabolites
 $R_r$  = List of required reactions
 $R_c$  = List of required enzymes
 $R_e$  = List of required elements
 $T_m$  = List of target metabolites
Output:
 $P$  = List of predicted pathways

Algorithm:
Let  $N_c$  denotes the combined (host-microbe) network
 $N_c = N_h \cup N_m$ 
CALL predictPathways ( $N_h$ )
CALL predictPathways ( $N_m$ )
CALL predictPathways ( $N_c$ )

Let  $P_{ch}$  denotes the list of pathways present in  $N_c$  but absent in  $N_h$ 
Let  $P_{cm}$  denotes the list of pathways present in  $N_c$  but absent in  $N_m$ 

 $P_{ch} = \text{compareNetworksByPathways}(N_c, N_h)$ 
 $P_{cm} = \text{compareNetworksByPathways}(N_c, N_m)$ 

Let  $P_{emergent}$  denotes the list of emergent pathways which are only possible due to interaction
 $P_{emergent} = P_{ch} \cap P_{cm}$ 

```

Box 2.4: Pseudocode for host-microbe interaction analysis algorithm

and their disruption leads to all related pathways being rendered as dysfunctional (Martz et al., 1996; Yeh et al., 2004; Taylor et al., 2013). To this end, MAPPS provides an interface to search for potential drug targets by identifying reactions acting as bridges in a metabolic network (see **Section 2.3.4**). Bridge reactions are the reactions which if removed from the metabolic network will result in all pathways being eliminated between the specified source and target metabolites (Mithani et al., 2009a). These reactions are termed as potential drug-targets in MAPPS because their removal from the underlying metabolic network can be used to disrupt the desired metabolic capabilities of an organism. Like pathway prediction, the user can choose one or more organism set(s) to build metabolic networks allowing them to compare the vulnerabilities in the metabolic architectures of multiple organisms sets in a single job in addition to specifying one or more start and end metabolites. All the pathways predicted from any given start and end metabolites are

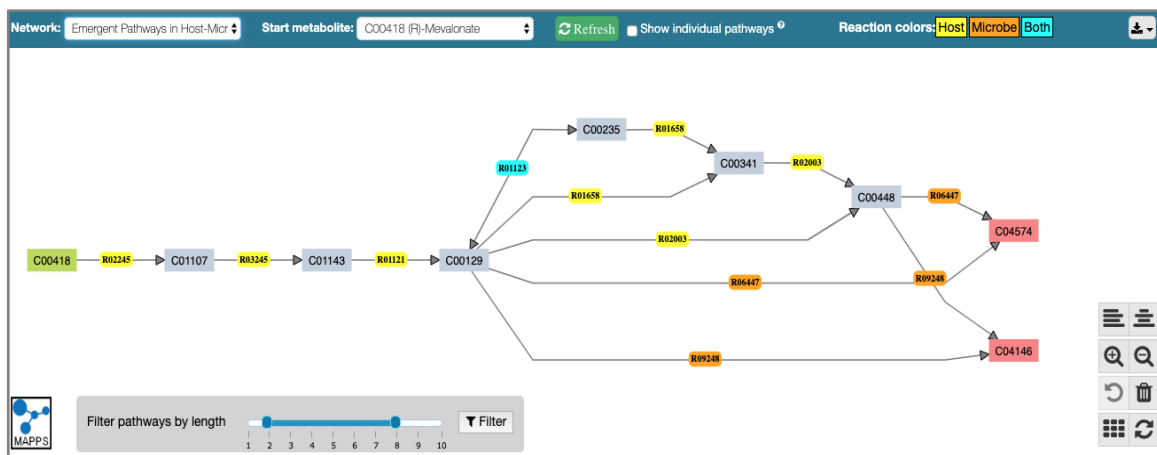


Figure 2.30: Graphical output showing pathways resulting due to host-microbe interactions. Reactions belong to the only host are shown in yellow color, reactions contributed by the microbe are shown in orange color. In this example, only one reaction is present in both host and microbe, shown in blue color.

further processed to identify bridge reactions in the underlying metabolic network, if present. The pseudocode of this analysis is given in **Box 2.5**.

2.2.4.9 Estimate evolution parameters

The evolution of metabolic networks is characterized by loss and gain of reactions (or enzymes) connecting two or more metabolites (Mithani et al., 2009b, 2010). It is possible to study how metabolic networks evolve by using simple (independent loss/gain of reactions) or complex (incorporating dependencies among reactions) stochastic models of metabolic evolution. Stochastic models of metabolic network evolution which describe metabolic evolution as a continuous-time Markov chain (Mithani et al., 2009b, 2010) are incorporated into MAPPS, which allows users to estimate evolution parameters (insertion rate, deletion rate, reaction dependencies) between two organism set(s) as well as over a phylogeny providing better insights into the evolution mechanisms of metabolic networks (**Figure 2.31**). Using statistical models of network evolution to analyze metabolic networks will also enable users to test various biological hypotheses such as specialization of genomes and identification of regions of metabolic networks that are under high selection


```

Input:
N = List of metabolic networks
Sm = List of start metabolites
Im = List of ignore metabolites
Ie = List of ignore elements
Rm = List of required metabolites
Rr = List of required reactions
Rc = List of required enzymes
Re = List of required element
Tm = List of target metabolites
Output:
drRxn = List of potential drug targets

Algorithm:
Let n denotes a metabolic network of an organism set
for each network  $n \in N$  do
  for each substrate  $s \in S_m$  do
    CALL predictPathways(n)
      Let Ps denotes the pathways predicted from start metabolite s
      Let Rp denotes the list of reactions with frequencies extracted from predicted
      pathways for start metabolite s

      for each  $r \in R_p$  do
        Let rFreq denotes the frequency of reaction r in Ps
        if rFreq is equal to count(Ps) then
          Save r in drRxn

```

Box 2.5: Pseudocode for potential drug-targets identification algorithm

and to investigate how the evolution of metabolic networks relates to the evolution of underlying genomes and the environment.

2.2.4.10 Visualize metabolic networks/pathways

MAPPs provides an option for interactive visualization of metabolic networks and pathways. This option is not limited to the visualization of MAPPs output and can also be used to visualize user-defined metabolic networks or pathways. It takes as input an SBML file containing KEGG or custom data and allows user to visualize and manipulate the resulting network/pathways in an interactive environment with an option to download the result in PNG and PDF formats.

2.2.5 *In silico* metabolic engineering

With the recent advances in high-throughput sequencing, metabolic engineering has opened new opportunities to design and analyze heterologous biosynthetic systems

Metabolic network Analysis and Pathway Prediction Server

Home Feedback About GuestUser

Job submission steps

1. Define job
2. Build network(s)
3. Enter job parameters
4. Review & submit job

Analysis:
Estimate evolution parameters
Show job summary

Results Tutorials
Saved Jobs Report bug

KEGG: Release 93.0/02-22, Feb 20

Data source
☒ KEGG ☐ Custom data
 MAPPS output is dependent on the accuracy of KEGG annotations

Dataset
☐ All pathways ☒ Select pathway (Parameter estimation works on a single pathway)
 Pathway set: All Pathways Search pathways...

ID	Pathway Name	Pathway Set
00525	Acarbose and validamycin biosynthesis	Biosynthesis of other secondary metabolites
01058	Acridone alkaloid biosynthesis	Biosynthesis of other secondary metabolites
00254	Aflatoxin biosynthesis	Biosynthesis of other secondary metabolites
00250	Alanine, aspartate and glutamate metabolism	Amino acid metabolism
00592	alpha-Linolenic acid metabolism	Lipid metabolism
00520	Amino sugar and nucleotide sugar metabolism	Carbohydrate metabolism

Network building mode
☒ Organism set(s) ☐ Phylogeny

Evolution model
☒ Independent edge model ☐ Neighbour dependent model ☐ Hybrid model

Organism set(s)

Start network
 Search Box Taxonomy View
 Search organism...
 Build network using: All reactions
☐ Perform *in silico* metabolic engineering
☐ Filter network using 'omics dataset

End network
 Search Box Taxonomy View
 Search organism...
 Build network using: All reactions
☐ Perform *in silico* metabolic engineering
☐ Filter network using 'omics dataset

Core edges
☒ Use default

Prohibited edges
☒ Use default

Stochastic network evolution parameters
 Runs: 1 Iteration: 3 Sampling interval: 1 Burn-in period: 0

Back Reset Next

Figure 2.31: MAPPS job submission interface for estimate evolution parameters analysis. MAPPS implements stochastic models of metabolic network evolution (Mithani et al., 2009b, 2010), which allows users to estimate evolution parameters (insertion rate, deletion rate, reaction dependencies) between two organism set(s) as well as over a phylogeny.

(Prather and Martin, 2008; Tomar and De, 2013). To facilitate the process of metabolic engineering, MAPPS provides a provision to perform *in silico* metabolic engineering by adding and/or removing reactions or enzymes while building metabolic networks (**Figure 2.32**). This option is available for all organism-based analyses described above involving

Metabolic network Analysis and Pathway Prediction Server

Home Feedback About GuestUser

Job submission steps

1. Define job
2. Build network(s)
3. Enter job parameters
4. Review & submit job

Analysis: Pathway prediction and comparison
Show job summary

Results Tutorials
Saved Jobs Report bug

KEGG: Release 93.0+02-22, Feb 20

Data source
KEGG Custom data
MAPPS output is dependent on the accuracy of KEGG annotations

Dataset
All pathways Select pathways

Network building mode
Organism set(s) Phylogeny

Organism set(s)
Organism set 1
Use Organism-based network Reference network Custom network (KEGG-based)
Search Box Taxonomy View
Escherichia coli K-12 MG1655 (ID:eco) Search organism...
Build network using: All reactions
Perform in silico metabolic engineering Engineering mode: Reactions Enzymes
Select pathways: Citrate cycle (TCA cycle) Search enzymes...
ID Name
1.1.1.37 malate dehydrogenase
1.1.1.42 isocitrate dehydrogenase (NADP+)
1.1.5.4 malate dehydrogenase (quinone)
1.2.4.1 pyruvate dehydrogenase (acetyl-transferring)
1.2.4.2 oxoglutarate dehydrogenase (succinyl-transferring)
1.2.7.1...
Filter network using 'omics' dataset
Perform comparative analysis
Add organism set

Figure 2.32: Performing *in silico* metabolic engineering in MAPPS. Users can add or remove one or more reaction/enzyme while building metabolic networks.

KEGG data and can be used to study the effects of *in silico* knockout and knock-in metabolic mutations on organisms' metabolic capabilities while performing different analyses. For example, addition of enzymes/reactions in a metabolic network enables users to determine the feasibility of engineering novel metabolic pathways in the network while removal of existing enzymes/reactions provides insight into the robustness of a metabolic network by detecting alternative routes between metabolites, and prediction of the potential effect of the knock-out on resulting metabolic pathways. To the best of our knowledge, no other tool provides the flexibility of simultaneously studying the effects of *in silico* knock-out (see **Section 2.3.2**) and knock-in metabolic (see **Section 2.3.3**) mutations. In addition, users can compare the results of metabolic engineering by running the analyses simultaneously on the original and modified networks, a feature not available in other tools.

2.2.6 Network filtering using ‘omics data

Another distinguishing feature that sets MAPPS apart from the currently available tools is the support for ‘omics data. Allowing users to compare metabolic annotations with other ‘omic datasets can provide greater insight into the metabolic capabilities of organisms. For example, incorporation of expression data into pathway prediction helps to identify enzymes that are co-expressed to give a functionally viable pathway, and in mapping functionally related genes to gene clusters. Similarly, metabolomic data mapping during pathway prediction can help in identifying pathways that are active under different experimental conditions and provide a detailed picture of what is going on inside a cell at a metabolic level.

MAPPS allows users to refine metabolic networks using ‘omics data at the network building step (**Figure 2.33**). To provide flexibility to users, MAPPS supports several public databases for ‘omics filtering. These include KEGG and NCBI identifiers of genes for transcriptomics data, UniProt (The UniProt Consortium, 2015) identifiers of enzymatic proteins for proteomics data, and KEGG, ChEBI (Hastings et al., 2013) and PubChem (S. Kim et al., 2016) identifiers for metabolites. The user can upload a tab-delimited file containing a list of unique identifiers for genes/proteins/metabolites from one of the supported public databases along with their expression/concentration values in addition to specifying a cut-off threshold, which is used to filter the metabolic networks. The nodes/edges that do not meet the required cut-off values are eliminated from the metabolic network. In the case of transcriptomic data, metabolic reactions (edges) are filtered using gene-enzyme-reaction mapping of provided genes. For proteomics data, the user provides a list of enzymatic proteins to filter reactions catalyzed by these enzymes. In the case of

a

Organism set(s)

Organism set 1

Use ☒ Organism-based network ☐ Reference network ☐ Custom network (KEGG-based)

Search Box Taxonomy View

Escherichia coli K-12 MG1655 (ID:eco) Search organism...

Build network using: All reactions

☐ Perform *in silico* metabolic engineering

☒ Filter network using 'omics dataset Dataset Type: ☒ Transcriptomics ☐ Proteomics ☐ Metabolomics

Dataset: Choose File No file chosen Gene database: KEGG No. of samples: 1 Count threshold: 0

☐ Perform comparative analysis

+ Add organism set

b

Organism set(s)

Organism set 1

Use ☒ Organism-based network ☐ Reference network ☐ Custom network (KEGG-based)

Search Box Taxonomy View

Escherichia coli K-12 MG1655 (ID:eco) Search organism...

Build network using: All reactions

☐ Perform *in silico* metabolic engineering

☒ Filter network using 'omics dataset Dataset Type: ☐ Transcriptomics ☒ Proteomics ☐ Metabolomics

Dataset: Choose File No file chosen Protein database: UniProt No. of samples: 1 Minimum threshold: 0

☐ Perform comparative analysis

+ Add organism set

c

Organism set(s)

Organism set 1

Use ☒ Organism-based network ☐ Reference network ☐ Custom network (KEGG-based)

Search Box Taxonomy View

Escherichia coli K-12 MG1655 (ID:eco) Search organism...

Build network using: All reactions

☐ Perform *in silico* metabolic engineering

☒ Filter network using 'omics dataset Dataset Type: ☐ Transcriptomics ☐ Proteomics ☒ Metabolomics

Dataset: Choose File No file chosen Database: KEGG No. of samples: 1 Minimum threshold: 0

☐ Perform comparative analysis

+ Add organism set

Figure 2.33: Network filtering using 'omics data. MAPPS allows users to refine metabolic networks using processed 'omics (transcriptomics/proteomics/metabolomics) data at the network building step. (a) For filtering based on transcriptomic data, a file containing a list of genes with KEGG or NCBI identifiers can be uploaded to filter the enzymes encoded by these genes in the metabolic network (b) For filtering based on proteomic data, a list of proteins with UniProt identifiers can be uploaded to filter enzymes of the metabolic network. (c) For filtering based on metabolomics data, metabolic networks can be filtered by uploading a file containing a list of metabolites with KEGG, PubChem or ChEBI identifiers. For (a)-(c), users must also specify a threshold value to consider the gene/protein/metabolite as being present or active in the network.

metabolomics data, reactions are filtered based on the presence/absence of metabolites acting as substrates in those reactions.

2.2.7 Pathway ranking

MAPPS reports all possible pathways between source and target metabolites that satisfy user constraints enabling discovery of previously unknown pathways. To help focus on biologically meaningful pathways, pathways can be ranked based on pathway length, the number of reversible reactions, pathway connectivity score, and a pathway score based on the scoring scheme introduced by Huang et al. (Huang et al., 2017).

The pathway connectivity score is calculated as follows. For each metabolite present in the underlying network, first, all reactions in which that metabolite is acting as a substrate are identified. Next, the total number of metabolites acting as products in these reactions is calculated by adding the number of product metabolites in individual reactions. This number is then normalized by dividing it with the maximum value for all the metabolites to get the metabolite connectivity score for that metabolite. The pathway connectivity score is then calculated as the sum of the log of metabolite connectivity scores of all intermediary metabolites involved in the pathway.

The scoring scheme introduced by Huang et al. (Huang et al., 2017) combines reaction thermodynamics information and structural similarity to calculate the similarity between any two metabolites. MAPPS computes a pathway score using the similarity scores between consecutive metabolites in a metabolic pathway. In this scheme, the score W_{ij} between any two consecutive metabolites v_i and v_j in a metabolic pathway is calculated as (Huang et al., 2017)

$$W_{ij} = \alpha \left(1 - \text{sim}(v_i, v_j) \right) + (1 - \alpha) (3200 + fe(r_{ij})) / 10000$$

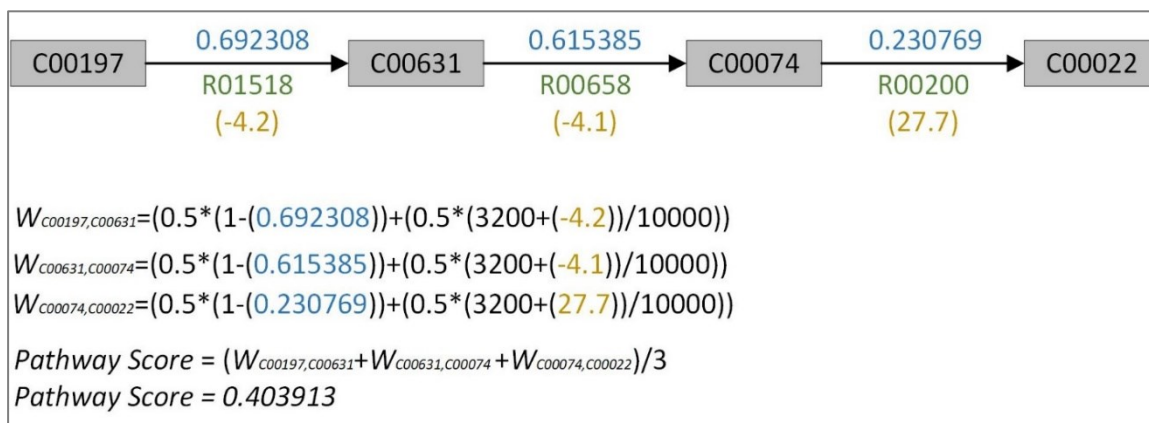


Figure 2.34: An example of pathway score calculation using reaction thermodynamics and structural similarity in MAPPS. Metabolite similarity score between two intermediate metabolites obtained using SIMCOMP2 (Hattori et al., 2010) is shown in blue, and the reaction energies downloaded from eQuilibrator (Flamholz et al., 2012) are shown in yellow.

where α is the proportional contribution of compound similarity and Gibbs free energy in the score (set to 0.5 in MAPPS), $\text{sim}(v_i, v_j)$ is the structural similarity between metabolites v_i and v_j (obtained using SIMCOMP2 (Hattori et al., 2010) in MAPPS, see **Section 2.2.1.1**), and $fe(r_{ij})$ is the Gibbs free energy of the reaction involving v_i and v_j (downloaded from eQuilibrator (Flamholz et al., 2012), see **Section 2.2.1.1**). MAPPS then calculates the pathway score by averaging the similarity scores across the length n of the pathway, as shown below.

$$Pathway\ Score = \sum W_{ij} / n$$

An example of pathway score calculation using this scheme is shown in **Figure 2.34**.

2.3 Applications of MAPPS

MAPPS is a versatile tool allowing a variety of analyses to be performed, ranging from simple pathway prediction to pathway-based network comparisons and phylogenetic analysis based on metabolic networks. Some important functionalities of MAPPS are demonstrated below by analyzing the data from published studies and comparing the

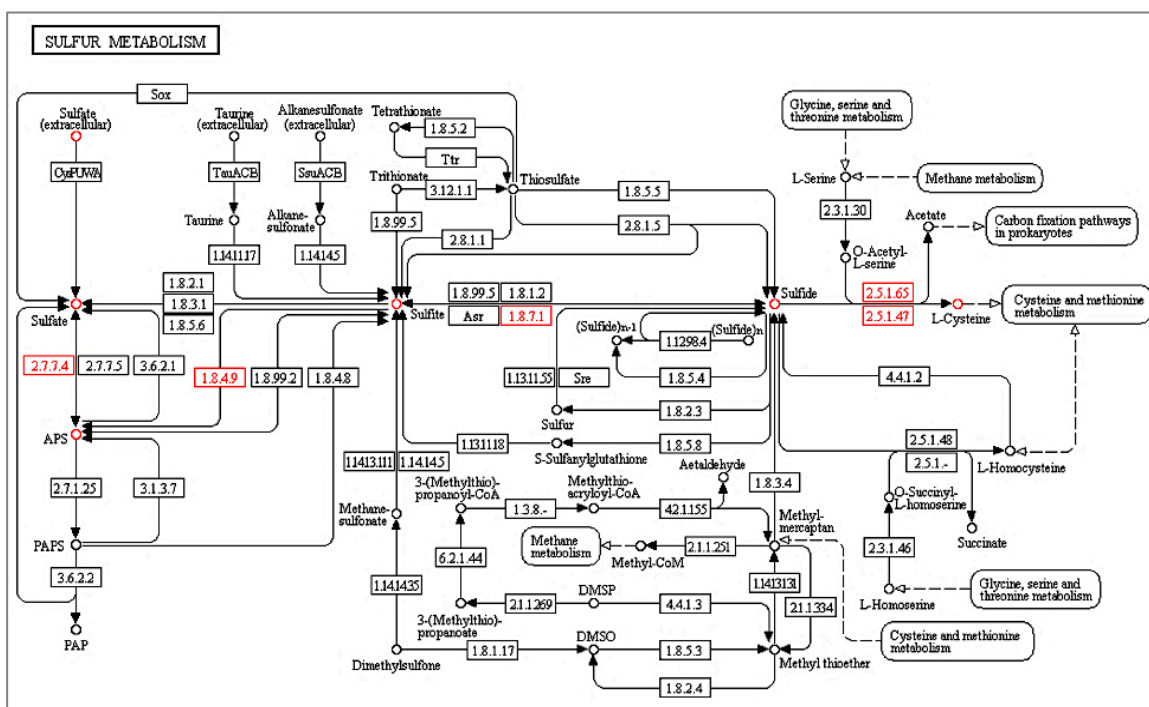


Figure 2.35: KEGG pathway map of sulfur metabolism. Red circles represent the participating metabolites of the reported pathway from sulfate to L-cysteine. Red rectangles are the reported enzymes for catalyzing the intermediate steps.

results obtained from some of the commonly used tools for metabolic pathway prediction and network analysis.

2.3.1 Predicting biologically meaningful metabolic pathways by tracing specific elements

In KEGG, the conversion of sulfate (KEGG ID: C00059) to L-cysteine (KEGG ID: C00097) corresponds to a four-step pathway in the sulfur metabolism map (KEGG map: 00920) (**Figure 2.35**). To demonstrate the application of MAPPS, abovementioned pathway in *Arabidopsis thaliana* starts when sulfur is taken up by roots and reduced to sulfide, then incorporated into activated O-Acetyl-L-serine to form cysteine (Nikiforova et al., 2004) (**Figure 2.36a**) and cysteine itself serves as a building stone for all further derived reduced sulfur-containing compounds such as methionine, proteins, glutathione, phytochelatin, biotin, thiamine, S-adenosylmethionine (SAM) and glucosinolates.

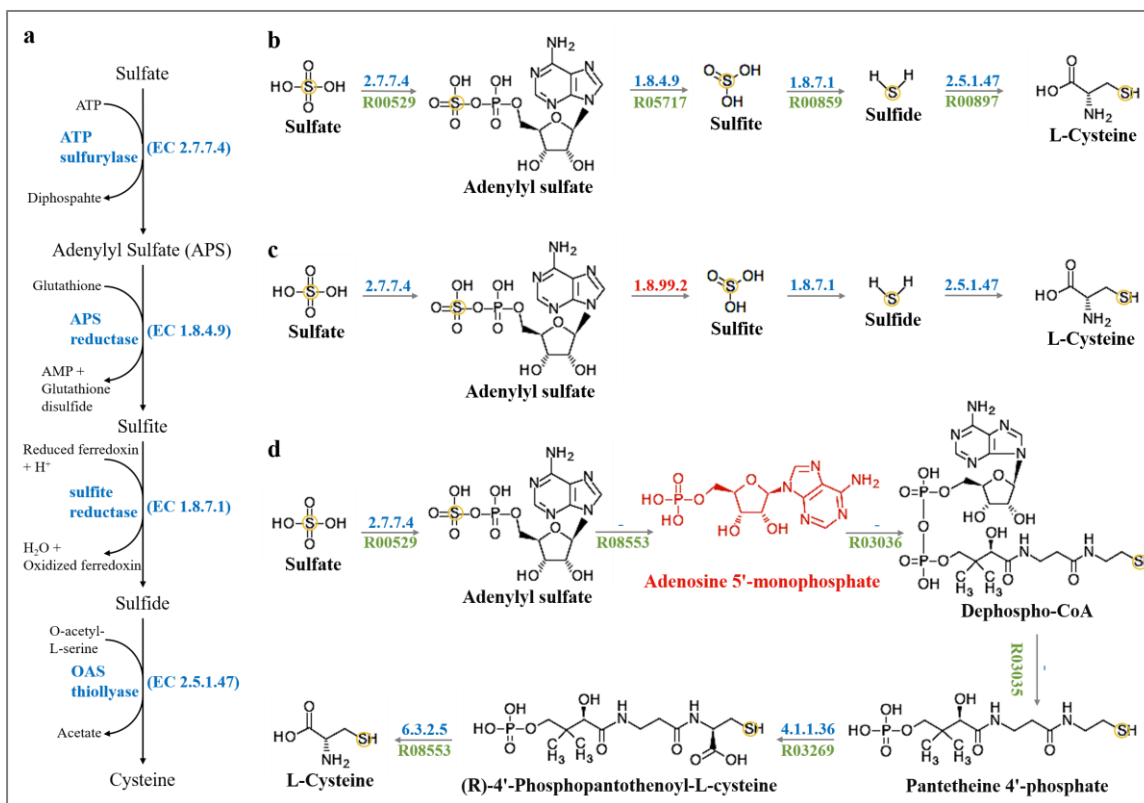


Figure 2.36: Sulfate assimilation pathway in *Arabidopsis thaliana*. (a) Schematic diagram of the sulfate assimilation pathway (adapted from (Koprivova and Kopriva, 2014)) (b) Pathway predicted by MAPPS from sulfate (KEGG ID: C00087) to L-cysteine (KEGG ID: C00097) corresponding to the experimentally validated pathway shown in (a). (c) FMM predicted pathway from sulfate to L-cysteine with correct intermediate metabolites but used an incorrect enzyme (EC 1.8.99.2) at the second step of the pathway (shown in red) instead of EC 1.8.4.9 for adenylyl-sulfate reductase, which is present in *A. thaliana*. (d) A seven-step pathway from sulfate to L-cysteine predicted by PHT containing non-sulfur containing ubiquitous metabolite, AMP (shown in red). Enzymes and their corresponding EC numbers are shown in blue whereas KEGG reaction ids are shown in green. Sulfur tracing is shown by the yellow circle.

Metabolic pathways are predicted from sulfate to L-cysteine using MAPPS, restricting the search to the sulfur metabolism map. MAPPS successfully computed the reported pathway (Figure 2.36b). As noted above, MAPPS allows users to define one or more elements as required or to be excluded in the pathway search (Section 2.2.4.1), which helps in tracing or avoiding a specific element in the reported metabolic pathways. To demonstrate the usability of tracing constituent elements, firstly pathways are computed from sulfate to L-cysteine across all KEGG pathway maps using metabolite pairing, which considers all

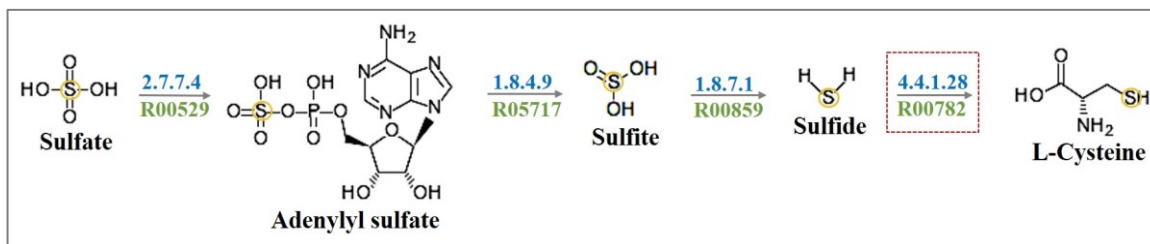


Figure 2.37: An alternate pathway from sulfate to L-cysteine predicted by MAPPs. Sulfur tracing is shown by the yellow circle. The alternate reaction compared to **Figure 2.36b** is shown in the red box.

possible pairs between source and target metabolites in a reaction without element tracing. MAPPs reported a total of six pathways from sulfate to L-cysteine including the pathway shown in **Figure 2.36b**. Out of these, four pathways, however, did not contain sulfur in one or more intermediary metabolites. Then, sulfur is designated as a required element and again pathway prediction was performed. This time MAPPs reported only two pathways with all sulfur-containing metabolites including the experimentally validated pathway described above. The other pathway differed by only one reaction/enzyme from the pathway shown in **Figure 2.36a** suggesting an alternate route to L-cysteine (**Figure 2.37**). Results of MAPPs were compared with the results of other pathway prediction tools including FMM (Chou et al., 2009), PHT (Rahman et al., 2005), PathComp (Ogata et al., 1998), MetaRoute (Blum and Kohlbacher, 2008), ATLAS (Hadadi et al., 2016) and MetQuest (Ravikrishnan et al., 2018). A significant limitation of the existing tools is that they do not allow the user to restrict the search space to some selected KEGG pathway(s), which results in irrelevant and out of context predictions. Besides, they do not allow tracing of a specific chemical element in the reported pathways thus resulting in biologically incorrect pathways. For example, FMM predicted 30 pathways from sulfate to L-cysteine. However, many of the predicted pathways did not pass through sulfur-containing metabolites (**Figure 2.38**). Although, it reported a pathway consisting of the same

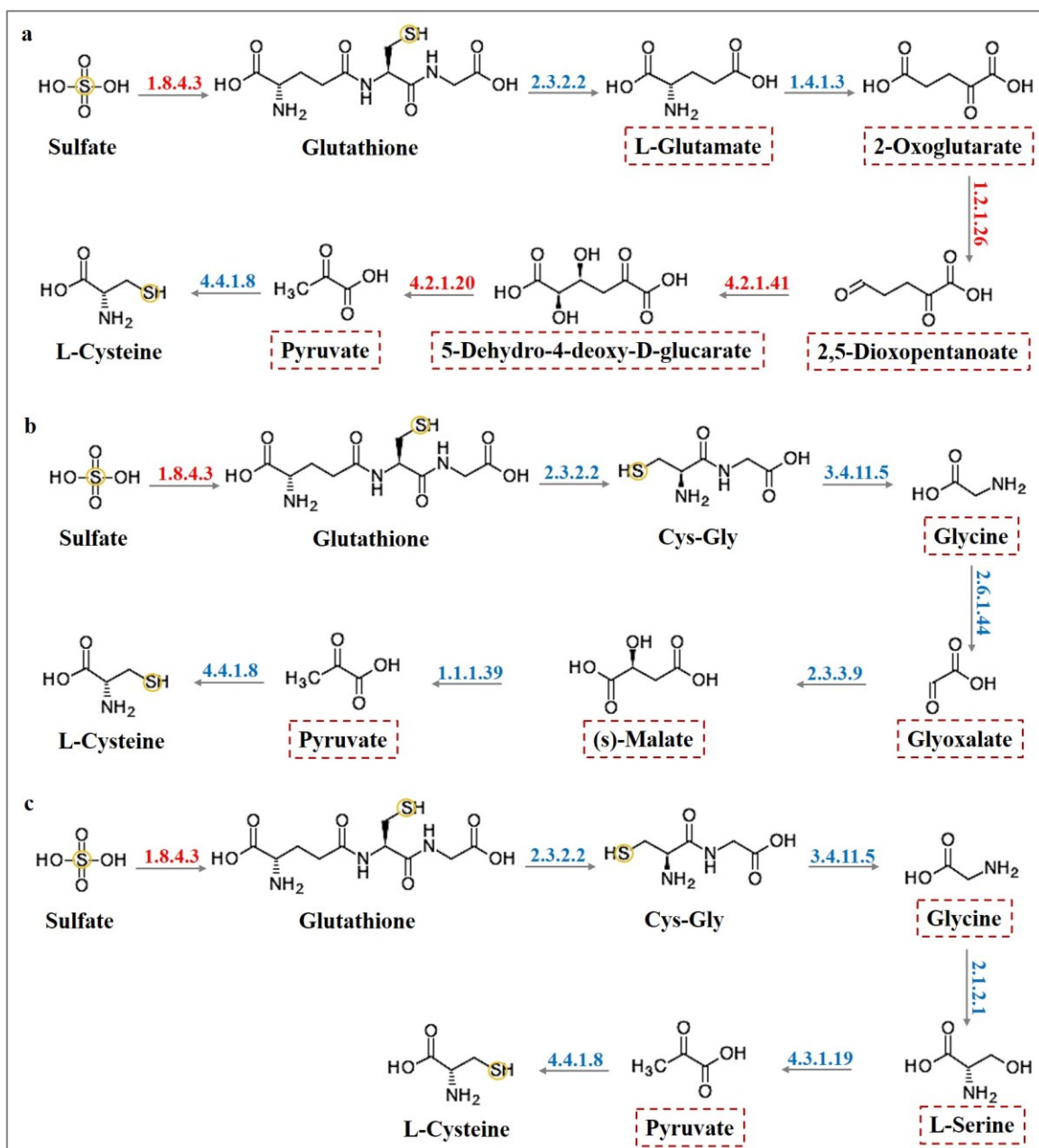


Figure 2.38: Examples of incorrect pathways predicted by FMM from sulfate to L-cysteine. FMM reported pathways utilizing non-native enzymes (red color) and non-sulfur containing intermediate metabolites (red boxes).

intermediary metabolites as the experimentally validated pathway it failed to assign the correct enzyme (adenylyl-sulfate reductase, EC 1.8.4.9) corresponding to genes *APR1*, *APR2* and *APR3* in *A. thaliana* at the second step of the pathway (Setya et al., 1996) (**Figure 2.36c**). PHT, on the other hand, reported only one seven-step pathway from sulfate

to L-cysteine, however this passed through adenosine monophosphate (AMP), a ubiquitous metabolite, rendering it biologically incorrect (**Figure 2.36d**). To minimize incorrect predictions, MAPPS is set by default to ignore ubiquitous metabolites such as AMP, ATP, and H₂O while predicting metabolic pathways (see **Section 2.2.4.1**) since including them during pathway prediction can lead to incorrect predictions (Ma and Zeng, 2003; Mithani et al., 2009a). PathComp, a tool from KEGG, predicted three pathways from sulfate to L-cysteine however none of them matched the experimentally validated pathway described above. MetaRoute allows users to restrict pathways based on atom tracing however in this case it failed to predict any pathway between the two metabolites. No pathway was found in ATLAS between sulfur and L-cysteine using KEGG reactions. Finally, MetQuest, a recently published tool, was used to predict pathways from sulfate to L-cysteine in *A. thaliana*. Since MetQuest is not linked to any database and requires an SBML file of the metabolic network, an SBML file of *A. thaliana* metabolic network was generated using the Network Enumeration option available in MAPPS (see **Section 2.2.4.5**). MetQuest also requires a seed metabolite set to guide the pathway search. For this, the list of ubiquitous metabolites (**Appendix B**) in addition to sulfate was added to the seed set and pathway prediction was performed. MetQuest predicted cyclic and branched pathways from various ubiquitous metabolites to L-cysteine but failed to predict any pathway from sulfate.

2.3.2 Studying the effects of *in silico* metabolic knock-out mutation

A distinguishing feature of MAPPS is that it can be used to perform *in silico* metabolic engineering to study the effects of knock-out and knock-in metabolic mutations on metabolic pathways (see **Section 2.2.5**). For knock-out mutations, a published study relating to galactose metabolism is used to demonstrate this feature. α -D-galactose is

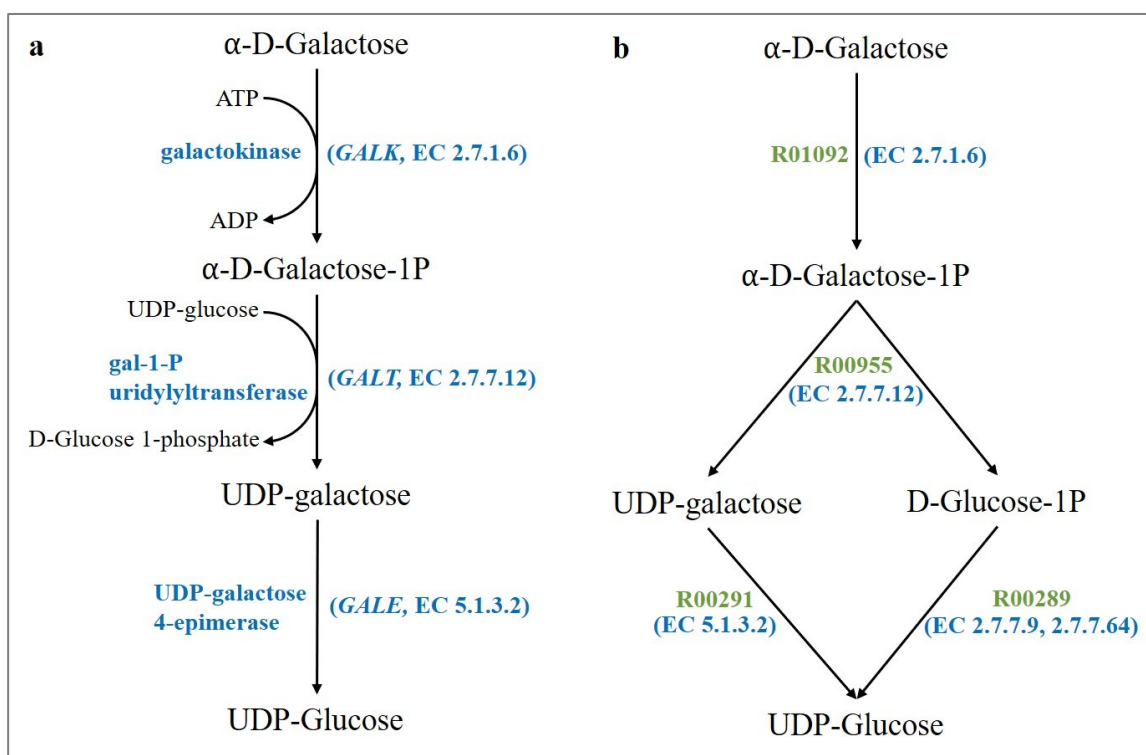


Figure 2.39: Leloir pathway in humans metabolizing α -D-galactose to produce UDP-glucose. (a) Schematic diagram showing the Leloir pathway for metabolizing α -D-galactose in humans. (b) Metabolic pathways from α -D-galactose to UDP-glucose predicted by MAPPS in the human metabolic network. In addition to reporting the Leloir pathway shown in (a), MAPPS also reports an alternate route to UDP-glucose via D-glucose-1P. *In silico* knock-out of 2.7.1.6 and/or 2.7.7.12 from the human metabolic network results in both pathways between α -D-galactose and UDP-glucose being eliminated whereas removal of 5.1.3.2 results in the Leloir pathway being eliminated. Enzymes and their corresponding EC numbers are shown in blue whereas KEGG reaction ids are shown in green.

metabolized in humans via a sequential pathway known as Leloir pathway (**Figure 2.39a**), where α -D-galactose is first phosphorylated by the enzyme galactokinase (GALK, EC 2.7.1.6) to produce α -D-galactose-1P, which is then converted into UDP-galactose using the enzyme galactose-1-P uridylyltransferase (GALT, EC 2.7.7.12), followed by interconversion of UDP-galactose to UDP-glucose through UDP-galactose 4'-epimerase (GALE, EC 5.1.3.2). Individuals with defects in any one of these enzymes are unable to properly metabolize milk sugar leading to Galactosemia, which is an inherited metabolic disorder (Fridovich-Keil, 2006). Pathway prediction between α -D-galactose (KEGG ID:

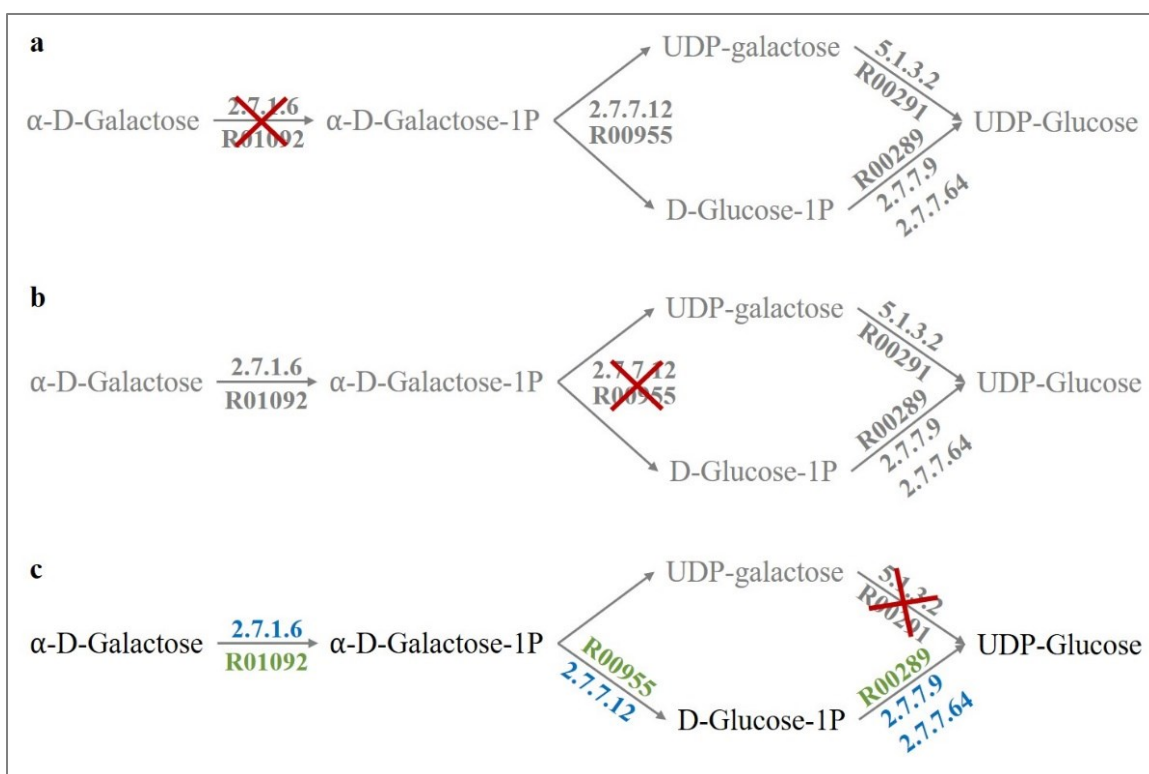


Figure 2.40: Effect of *in silico* knockouts on predicted pathways between α -D-galactose and UDP-glucose in human galactose metabolic network. The enzymes (and the corresponding reactions) removed from the network are shown by a red cross and pathway(s) eliminated as a result of the knock-out are shown in gray color.

C00984) and UDP-glucose (KEGG ID: C00029) using MAPPS on human galactose metabolic network (KEGG map: 00052) resulted in two pathways being reported including the Leloir pathway (**Figure 2.39b**). The alternate pathway only differs at the last step which uses α -D-glucose-1P to produce UDP-glucose using enzyme UTP-glucose-1-phosphate uridylyltransferase (UGP2, EC 2.7.7.9) instead of using UDP-galactose and forms a part of the UDP- α -D-glucose biosynthesis I pathway in MetaCyc (Caspi et al., 2016). Next, *in silico* metabolic engineering option (see **Section 2.2.5**) was used to remove the enzyme galactokinase from the human metabolic network and pathway prediction was performed again. No metabolic pathways were reported in this case (**Figure 2.40a**). A similar result was obtained when the enzyme galactose-1-P uridylyltransferase was removed (**Figure**

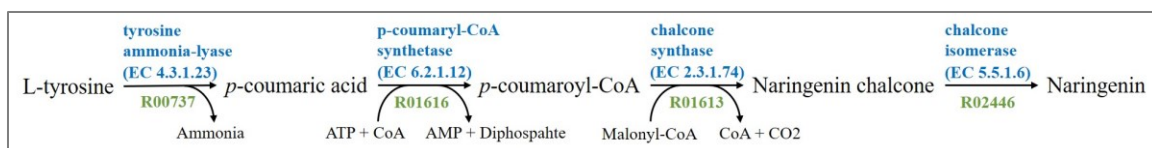


Figure 2.41: A heterologous pathway for flavonoid production in *Escherichia coli*. A heterologous biosynthetic pathway for producing flavonoid precursor naringenin from L-tyrosine in *Escherichia coli* (Santos et al., 2011) is shown. MAPPs reports this heterologous pathway in the modified *E. coli* metabolic network with enzymes tyrosine ammonia lyase (TAL, EC 4.3.1.23), 4-coumarate:CoA ligase (4CL, EC 6.2.1.12), chalcone synthase (CHS, EC 2.3.1.74) and chalcone isomerase (CHI, EC 5.5.1.6) added through *in silico* metabolic engineering. Enzymes and their corresponding EC numbers are shown in blue whereas KEGG reaction ids are shown in green.

2.40b) and only one pathway through α -D-glucose-1P was reported when UDP-galactose 4'-epimerase was removed from the human metabolic network (**Figure 2.40c**). PHT also allows users to predict pathways by excluding one or more enzymes during pathway prediction but, unlike MAPPs, does not allow modification of the underlying network. However, in this case, PHT failed to predict any pathway from α -D-galactose and UDP-glucose in the human metabolic network and, therefore, could not be used for studying the knockout of enzymes involved in Leloir pathway.

2.3.3 Designing heterologous pathways using *in silico* metabolic knock-in mutations

In silico metabolic engineering can be used to design heterologous biosynthetic pathways by incorporating foreign enzymes into a host. To demonstrate this, the *in silico* metabolic engineering option available in MAPPs (see **Section 2.2.5**) was used to reproduce a heterologous pathway for flavonoid production from L-tyrosine in *Escherichia coli* (Santos et al., 2011) (**Figure 2.41**). Four enzymes were added namely tyrosine ammonia lyase (TAL, EC 4.3.1.23), 4-coumarate:CoA ligase (4CL, EC 6.2.1.12), chalcone synthase (CHS, EC 2.3.1.74) and chalcone isomerase (CHI, EC 5.5.1.6) to the *E. coli* metabolic network and compared pathways from L-tyrosine (KEGG ID: C00082) to the main

Metabolic network Analysis and Pathway Prediction Server

Home Feedback About GuestUser

Job submission steps

1. Define job
2. Build network(s)
3. Enter job parameters
4. Review & submit job

Analysis:
Pathway prediction and comparison
Show job summary

Results Tutorials
Saved Jobs Report bug

KEGG: Release 93.0v/02-22, Feb 20

Data source
☒ KEGG ☐ Custom data
MAPPs output is dependent on the accuracy of KEGG annotations

Dataset
☒ All pathways ☐ Select pathways

Network building mode
☒ Organism set(s) ☐ Phylogeny

Organism set(s)

Organism set 1

Use ☒ Organism-based network ☐ Reference network ☐ Custom network (KEGG-based)

Search Box Taxonomy View

Escherichia coli K-12 MG1655 (ID:eco) Search organism...

Build network using: All reactions

☐ Perform *in silico* metabolic engineering
☐ Filter network using 'omics dataset

Organism set 2

Use ☒ Organism-based network ☐ Reference network ☐ Custom network (KEGG-based)

Search Box Taxonomy View

Escherichia coli K-12 MG1655 (ID:eco) Search organism...

Build network using: All reactions

☒ Perform *in silico* metabolic engineering
 Engineering mode: ☐ Reactions ☒ Enzymes

Select pathways: Flavonoid biosynthesis Search enzymes...

ID	Name
<input type="checkbox"/> 2.3.1.74	naringenin-chalcone synthase
<input type="checkbox"/> 2.4.1.185	flavanone 7-O-beta-glucosyltransferase
<input type="checkbox"/> 2.4.1.236	flavanone 7-O-glucoside 2"-O-beta-L-rhamnosyltransferase
<input type="checkbox"/> 2.4.1.357	phlorizin synthase
<input checked="" type="checkbox"/> 5.5.1.6	chalcone isomerase

Altered Reactions/Enzyme
 4.3.1.23 x 2.3.1.74 x 6.2.1.12 x 5.5.1.6 x

☐ Filter network using 'omics dataset

☒ Perform comparative analysis

Figure 2.42: Comparing *Escherichia coli* metabolic network with and without *in silico* knock-in mutations. Four enzymes, namely tyrosine ammonia lyase (TAL, EC 4.3.1.23), 4-coumarate:CoA ligase (4CL, EC 6.2.1.12), chalcone synthase (CHS, EC 2.3.1.74) and chalcone isomerase (CHI, EC 5.5.1.6) were inserted into the *E. coli* metabolic network to reproduce the heterologous metabolic pathway from L-tyrosine to naringenin (Santos et al., 2011).

flavonoid precursor naringenin (KEGG ID: C00509) in the original and modified *E. coli* network (**Figure 2.42**). While there was no pathway reported in the original network between L-tyrosine and naringenin, a pathway utilizing the newly added enzymes was reported in the modified *E. coli* network (**Figure 2.42**). MRE, which also predicts this heterologous pathway, however, takes a slightly different approach for predicting engineered pathways. It takes an organism as input and first predicts the pathway between a start metabolite and an end metabolite on the KEGG reference network. It then identifies

native and foreign enzymes present in the reported pathways. This makes the results pathway-specific and does not provide a complete picture at the network level. Moreover, MRE does not allow users to study the effects of knock-out mutations on metabolic pathways. MAPPS, on the other hand, provides a biologically intuitive platform by allowing users to modify the original network enabling users to study the effects of knock-out and/or knock-in mutations on the metabolic network. Besides, MAPPS provides an option for direct comparison between metabolic networks allowing users to study the effects of different *in silico* modifications on the predicted pathways from start metabolite(s) to end metabolite(s) in the underlying network (**Figure 2.42**).

2.3.4 Identification of potential drug targets

MAPPS allows users to identify potential drug targets in a metabolic network by identifying reactions which if removed from the metabolic network, will result in all pathways being eliminated between the specified source and target metabolites. For example, prostaglandin- endoperoxide synthase (EC 1.14.99.1) is a reported target of many anti-inflammatory drugs including aspirin and ibuprofen and catalyzes the conversion of arachidonic acid (KEGG ID: C00219) to prostaglandin H₂ (PGH₂, KEGG ID: C00427) via prostaglandin G₂ (PGH₂, KEGG ID: C05956) in two steps (Funk, 2001). PGH₂ is the precursor for all prostanoids including prostaglandins, thromboxanes and prostacyclins and is a key metabolite in arachidonic acid metabolism (**Figure 2.43**). MAPPS reports the two reactions (R00073 and R01590) catalyzed by the enzyme prostaglandin-endoperoxide synthase as potential drug targets for disrupting metabolic pathways from arachidonic acid to various prostanoids (**Figure 2.43**).

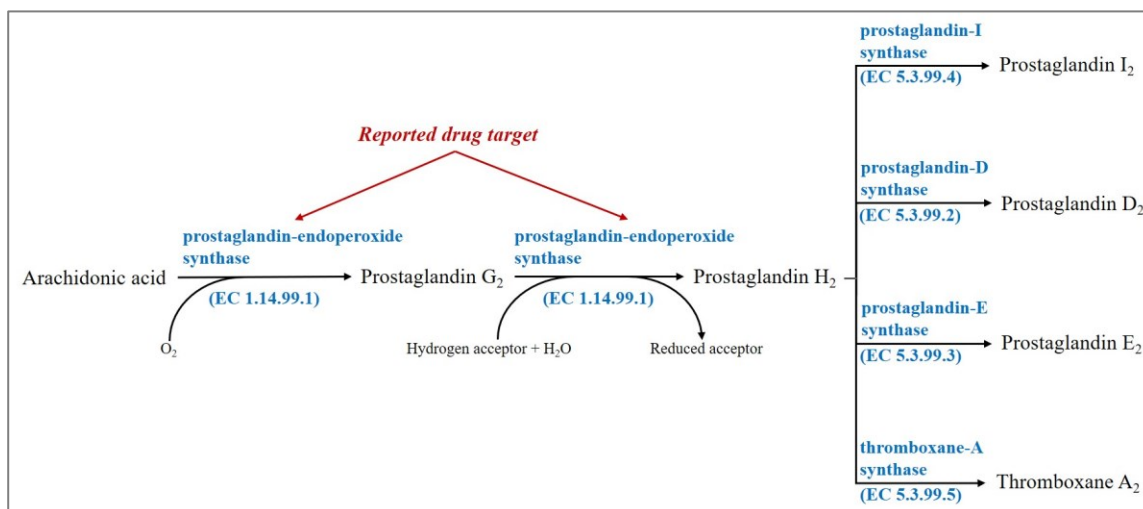


Figure 2.43: An application of potential drug target identification in MAPPS. The enzyme prostaglandin-endoperoxide synthase (EC 1.14.99.1) catalyzes two-step conversion of arachidonic acid to prostaglandin H₂, the precursor for all prostanoids including those shown here, and is a reported target of many anti-inflammatory drugs including aspirin and ibuprofen. MAPPS identified both reactions (R00073 and R01590) catalyzed by this enzyme as potential drug targets for disrupting metabolic pathways from arachidonic acid to various prostanoids. Enzymes and their EC numbers are shown in blue.

Another example demonstrating the efficacy of MAPPS in identifying potential drug targets relates to the phosphotransferase system (PTS) for transporting sugar into bacteria. PTS, which is crucial for the survival of bacterial growth, is specific to prokaryotes and thus can serve as a potential drug target (Ren et al., 2018). It has been reported that replication and survival of *Salmonella enterica* serovar Typhimurium (*S. Typhimurium*), which causes gastroenteritis and fatal typhoid, in mice depend on glucose and glycolysis (Bowden et al., 2009). To identify potential enzymes from the PTS which can be used as drug targets, the drug target identification module in MAPPS (see **Section 2.2.4.8**) was used to predict the metabolic pathways between D-glucose (KEGG ID: C00031) and pyruvate (KEGG ID: C00022) considering the whole carbohydrate metabolism.

MAPPS reported reaction R02738 which is catalyzed by Enzyme II^{Glc} (EC 2.7.1.199) as the potential drug target (**Figure 2.44**). Enzyme II^{Glc} is a key component in the PTS system and is involved in the transport of glucose across the membrane as well as its phosphorylation (Ren et al., 2018). To confirm the effect of enzyme removal, the metabolic pathways from glucose to pyruvate in the original *S. Typhimurium* metabolic network were compared against the one with Enzyme II^{Glc} removed using *in silico* metabolic engineering (see **Section 2.2.5**). While multiple pathways were reported in the unmodified network, no pathway was reported in the modified network suggesting that Enzyme II^{Glc} can indeed be used as a potential drug target to prevent *S. Typhimurium* infection in mammals.

2.3.5 Studying emergent pathways resulting from host-microbe interaction

MAPPS allows users to identify novel pathways emerging due to interaction between host and microbe enabling them to study the metabolic basis of the host-microbe interface. It is demonstrated by exploring a hypothesis relating to isoprenoid biosynthesis, discussed in detail elsewhere (Ahyong et al., 2019), that *Rickettsia parkeri*, a gram-negative obligate intracellular parasite that causes typhus and spotted fever in humans lacks necessary enzymes to produce isopentenyl pyrophosphate (IPP; KEGG ID: C00129), the central precursor molecule for producing isoprenoids, and its isomer dimethylallyl diphosphate (DMAPP; KEGG ID: C00235). Instead, *R. parkeri* uses the human mevalonate (MEV) pathway as the upstream source of IPP for its production of bactoprenols, which are essential building blocks for peptidoglycan and other cell wall polysaccharides, and ubiquinone, a coenzyme involved in electron transport chain (Heuston et al., 2012). To

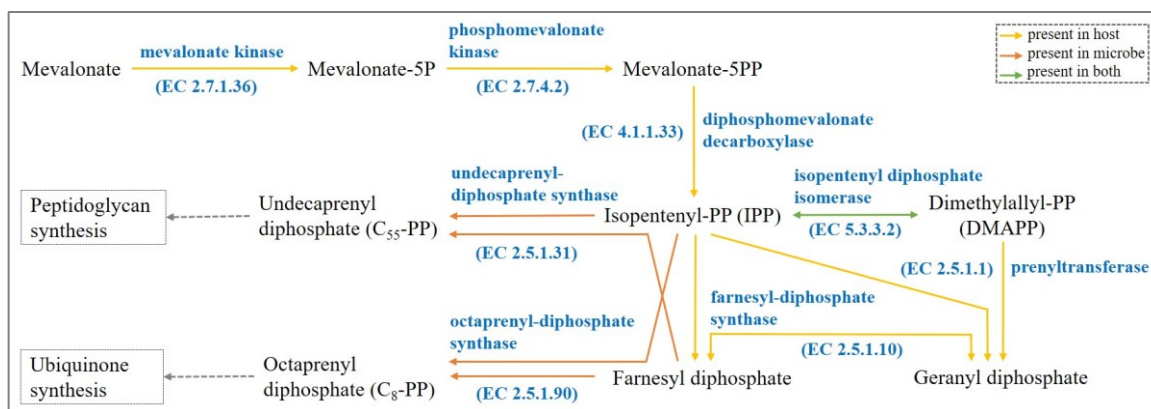


Figure 2.45: Prediction of emergent pathways arising due to the interaction between *Rickettsia parkeri* and humans. *R. parkeri* lacks upstream enzymes to produce isopentenyl pyrophosphate (IPP) from mevalonate (MEV) biosynthesis pathway. It uses human MEV pathway to produce Octaprenyl diphosphate (C₈-PP) and Undecaprenyl diphosphate (C₅₅-PP) which are precursors for ubiquinone synthesis and peptidoglycan synthesis respectively. MAPPS correctly predicted metabolic pathways from Mevalonate, via IPP produced by human metabolic enzymes, to C₈-PP and C₅₅-PP. Enzymes and their corresponding EC numbers are shown in blue, reaction present in both, host and microbe, is shown in green, while reactions present in only humans are shown in yellow and the microbe-specific reactions are shown in orange color.

investigate the metabolic interaction between *R. parkeri* and human in the isoprenoid biosynthetic pathway, the host-microbe interaction option available in MAPPS (see **Section 2.2.4.7**) was used by setting human and *R. parkeri* as host and microbe networks respectively, and predicted metabolic pathways from mevalonate (KEGG ID: C00418) to Octaprenyl diphosphate (C₈-PP; KEGG ID: C04146) and Undecaprenyl diphosphate (C₅₅-PP; KEGG ID: C04574), which are precursors for ubiquinone and peptidoglycan synthesis respectively in gram-negative bacteria (Ahyong et al., 2019).

While no pathways were reported in the host or microbe networks separately, MAPPS predicted multiple emergent pathways between mevalonate and the two precursor metabolites C₈-PP and C₅₅-PP in the combined network (**Figure 2.45**), which use the human MEV pathway for the upstream source of isoprene units for the synthesis of bacterial bactoprenols and ubiquinone. The reported pathways not only match the route suggested to be taken by *R. parkeri* for isoprenoid biosynthesis (Ahyong et al., 2019) but

rightly identify enzyme isopentenyl diphosphate isomerase (EC 5.3.3.2) catalyzing the reversible conversion of IPP to DMAPP (Berthelot et al., 2012) as the only enzyme to be present in both humans and *R. parkeri* in this pathway thus demonstrating the potential of MAPPS in studying pathway-based host-microbe interactions. To the best of our knowledge, no other tool allows the pathway-based analysis of host-microbe interactions.

2.4 Discussion

This chapter presented MAPPS, a web-based tool for metabolic pathway prediction and network analysis on groups of organisms as well as on a phylogeny. A number of case studies showing the efficacy of MAPPS in predicting metabolic pathways and analyzing metabolic networks were also presented. MAPPS uses data from KEGG in addition to allowing users to upload custom data, and can be used to compare metabolic networks of two or more sets of organisms, report all possible pathways between two or more metabolites, provide insights into the behavior of engineered metabolic networks, study host-pathogen interaction, identify potential drug targets, analyze the effects of availability/unavailability of metabolites on the reported metabolic pathways through *in silico* metabolic engineering, and build ancestral networks over a given phylogeny. It also provides an intuitive approach to answer biological questions focusing on metabolic capabilities of an organism as well as differences between organisms or the evolution of different species by allowing pathway-based comparisons of the metabolic network at an organism as well as at a phylogenetic level. Stochastic models of metabolic evolution (Mithani et al., 2010) are also incorporated in MAPPS enabling users to study the evolution of metabolic networks and to predict metabolic networks at the ancestral levels of a phylogeny. MAPPS also provides an interactive graphical user interface for job submission

and result visualization providing single-click access to relevant information in publicly available databases and explore the output by manipulating the parameters in addition to generating the output in standard SBML format, which provides the flexibility to upload the file to other software for further analysis. Importantly, MAPPS has an ‘omics pipeline to refine the pathway results using transcriptomic, proteomic or metabolomic data to provide a greater insight into the metabolic capabilities of organisms making it relevant in today’s post-genomic era.

Provision of filtering metabolic networks based on omics data is a distinguishing feature of MAPPS and is not offered by most of the available tools. A cell regulates its enzyme production and behavior depending on its requirements. This is achieved through several regulatory mechanisms which allow an organism to adapt to the environmental changes (Gonçalves et al., 2013; Rodriguez-Martinez et al., 2016). Filtering of metabolic networks based on ‘omics data allows user to study the effect of these changes on metabolic pathways, for example, by identifying the set of metabolites being affected by any internal or external stimuli (Barupal et al., 2012). Similarly, ‘omics data can help to identify genes or metabolites that are expressed under different conditions (for example healthy versus cancer). By identifying metabolic pathways active under each condition can provide insights into metabolic responses of an organism to different conditions. At present, MAPPS allows filtering of metabolic networks based on a user-defined cutoff. This can be further expanded by allowing user to identify differential metabolites or enzymes in the predicted metabolic pathways by mapping expression data on to the networks thereby reporting the flux passing through the computed pathways (Ganneru et al., 2019).

An interactive user interface is an essential requirement for any metabolic pathway prediction and network comparison tool for answering complex biological questions. Graphical results visualization interface of MAPPS enables user to manipulate the output by allowing them to filter predicted pathways, remove one or more metabolites/reactions to see its effect on the topology of predicted pathways or networks. Large outputs, however, hamper the speed of graphical results visualization in MAPPS since each manipulation of the output requires re-rendering of the results. This can be improved by utilizing state-of-the-art methods such as GPU accelerated rendering methods. Moreover, MAPPS provides an option to download the graphical output in PDF and PNG formats. Allowing users to download the graphical output in other image formats such as Scalable Vector Graphics (SVG) and Enhanced Metafile (EMF) which will help user to generate high-quality images for diverse applications.

Besides this, MAPPS currently uses KEGG as its primary data source with an option to upload custom data to build metabolic networks. Adding other data sources such as Reactome (Fabregat et al., 2016) and BioCyc (Caspi et al., 2016) would enhance the capabilities of MAPPS and make it more useful to the scientific community. In addition, MAPPS uses depth-first search of the metabolic graphs to compute pathways between metabolites. Consequently, jobs which employ exhaustive search by using metabolite pairing to establishing connection between metabolites or use all KEGG pathways as the underlying dataset require a long time to run due to an exponential increase in the search space. This can be sped up by using parallel programming and hybrid search algorithms to optimize pathway searches.

MAPPS predicts all theoretically possible metabolic routes between given source and target metabolites that match the criteria specified by the user through job parameters. While different type of constraints can be applied at enzyme, reaction, and metabolite levels which allow users to focus the pathway search on specific biological questions, tagging the experimentally validated pathways using public databases such as BioCyc can help users in filtering out false positive pathways. Similarly, by adding information about the co-expression and co-localization of metabolic enzymes into MAPPS will help users in predicting biologically intact pathway inside the cell as some of the enzymes catalyzing the reactions in the reported pathway may not be present in the same cellular compartment or may be expressed under different physiological conditions (Zecchin et al., 2015).

In summary, MAPPS provides a powerful resource for metabolic pathway prediction and comparison, specialized analyses such as drug target identification, *in silico* metabolic engineering by adding/removing metabolic reactions or enzymes, detection of metabolite-specific reactions, analyzing the effects of host-microbe interactions, and to study metabolic evolution using traditional as well as stochastic models. Omics filtering, interactive interface, provision for custom data, flexibility to apply constraints and alterations in the metabolic network and comparison of multiple networks are some of the distinguishing features in MAPPS making it an effective and useful resource for metabolic network analysis in the years to come.

3 Metabolic Diversity and Functional Specialization in *Pseudomonas*

The genus *Pseudomonas* is one of the most complex bacterial genera and is currently the genus of Gram-negative bacteria with the most significant number of species (Gomila et al., 2015). *Pseudomonas* are found in all of the major natural environments and show remarkable adaptability to their environments (Spiers et al., 2000). Members of the genus *Pseudomonas* show high metabolic and physiologic versatility, enabling the colonization of diverse terrestrial and aquatic habitats and are of great interest because of their importance in plant and human disease, and their growth potential in biotechnological applications (Silby et al., 2011). *Pseudomonas* primarily rely on the properties of their metabolic networks such as adaptation to extreme and diverse niches, pathogenic and non-pathogenic lifestyle, and production of essential compounds to achieve this adaptability (Oberhardt et al., 2008; Mailloux et al., 2011; Cabot et al., 2016; Botelho et al., 2019).

Different species of *Pseudomonas* vary in their ability to survive in different environments. Some are psychrophile, for example *Pseudomonas antarctica* and *Pseudomonas lurida* (Selvakumar et al., 2011; Lee et al., 2017), some are thermophile for example *Pseudomonas thermotolerans* (Manaia and Moore, 2002), some thrive in plant and soil environments like *Pseudomonas syringae* and *Pseudomonas fluorescens* (Rico and Preston, 2008; Couillerot et al., 2009) and some species, for example *Pseudomonas aeruginosa*, can colonize in diverse environments including humans (Remold et al., 2011). Adaptation to multiple niches depicts exceptional physiological and metabolic capabilities of this organism, and provides dominance against competitors (Frimmersdorf et al., 2010).

Pathogenic pseudomonads can cause severe disease in humans, insects, and plants (Oberhardt et al., 2010; Aditi et al., 2017) whereas some species have specialized to produce antifungal and antimicrobial compounds (Gross and Loper, 2009; Ramette et al., 2011; Calderón et al., 2015; Montes et al., 2016) or are involved in denitrification (Lalucat et al., 2006). For example, *P. aeruginosa* is also an opportunistic human pathogen, particularly associated with infections that are caused due to weakened immune system (Oberhardt et al., 2010; Subedi et al., 2018). Similarly, some pseudomonads live in commensal relationship with plants to fulfill their nutrient requirements from plant surfaces (Paulsen et al., 2005) while a group of pseudomonads including *Pseudomonas protegens*, *Pseudomonas chlororaphis* and *Pseudomonas orientalis* are classified as plant protecting bacteria which have shown to exhibit antagonistic properties against phytopathogens including fungi (Loewen et al., 2014; Calderón et al., 2015; Zengerer et al., 2018).

Over the years, various *Pseudomonas* strains have also been studied for their metabolic engineering potential to produce essential compounds and for different biotechnological applications. For example, *Pseudomonas putida* KT2440 was engineered to produce *para*-Hydroxy benzoic acid, a key component in the manufacturing of liquid crystal polymers for several industrial applications, from glucose (Yu et al., 2016). *P. putida* KT2440 is now commonly used for expressing heterologous genes to produce beneficial compounds due to its flexible lifestyle (Nelson et al., 2002; Nikel and de Lorenzo, 2018). Similarly, vanillin, an important compound of vanilla flavor, is produced from ferulic acid by altering the metabolism of a *Pseudomonas fluorescence* strain (Di Gioia et al., 2011) and metabolic network of *Pseudomonas denitrificans* has been

engineered for the production of 1,3-propanediol (1,3-PDO), an important intermediary used in the making of polymers, from glycerol (Zhou et al., 2019).

In recent years, several studies have provided insights into various traits such as pathogenicity, protein domain conservation and gene essentiality based on metabolic models of few *Pseudomonas* species (Koehorst et al., 2016; Subedi et al., 2018) and comparison of species classification based on phylogenetic analyses using 16S rRNA and different housekeeping genes (Gomila et al., 2015). In addition, catalogs of metabolic and genomic information have been developed to facilitate the researchers in understanding the conservation, differences, and functional annotations in *Pseudomonas* providing a framework to understand the specialization and diversity among different strains (Koehorst et al., 2016). These databases facilitated in the development and annotation of various metabolic models, and comparative analyses of *Pseudomonas* strains using these models (Tokic et al., 2020). This has further led to the construction of genome-scale metabolic networks of various *Pseudomonas* species, which are now widely used for the study of virulence (Bartell et al., 2017), metabolic changes in the pathogenic strains during disease progression (Oberhardt et al., 2010), identifying drug targets (Perumal et al., 2011) and metabolic rewiring for the production of secondary metabolites for industrial applications (Puchalka et al., 2008). Recently, reduction of growth rate and metabolic specialization have been used as a signature to study adaptive evolution of *Pseudomonas aeruginosa* when migrating from the environment to the airways of Cystic Fibrosis patients (La Rosa et al., 2018). Besides this, studies have also used the topology of metabolic networks to address various biological questions including identification of metabolic hubs (Perumal et al., 2009), evolutionary analysis of metabolic networks (Mithani et al., 2011),

cooperative and competitive metabolic interactions (Freilich et al., 2011), effect of pathogens on plant growth (Duan et al., 2013) and optimization of bacterial growth under different conditions and production of valuable compounds (Kampers et al., 2019; Bator et al., 2020). However, no study has systematically analyzed metabolic networks, compare primary metabolic pathways to assess the functional specialization of different *Pseudomonas* strains, and study the evolutionary relationships of this genus.

The diversity of pseudomonads and the availability of genome sequence and metabolic data for multiple strains provide an excellent opportunity to use comparative approaches to develop insight into the evolution of metabolic networks. This chapter investigates the metabolic diversity and functional specialization of nutrient assimilation pathways in *Pseudomonas*. MAPPS (see **Chapter 2**) was used to analyze metabolic networks in various species of *Pseudomonas* with diverse lifestyles. Comparisons of whole metabolic networks, as well as key metabolic pathways, were used to explore the metabolic diversity among the pseudomonads. The phylogenetic relationship between the strains was analyzed using the sequences for 16S rRNA, and four housekeeping genes via multilocus sequence analysis and compared with the groupings of different strains obtained based on their metabolic similarity.

3.1 *Pseudomonas* dataset

To explore the metabolic diversity and functional specialization in *Pseudomonas*, a total of 111 strains comprising of human pathogens, insect pathogens, plant pathogens, and non-pathogenic pseudomonads were analyzed in this study. **Table 3.1** shows the list of pseudomonads used in this analysis along with their KEGG codes, GenBank Ids, genome sizes, number of reactions in their metabolic network, and lifestyles and characteristics.

Table 3.1: *Pseudomonas* strains used for metabolic network analysis

Organism Name	KEGG Code	GenBank ID	Genome Size	No. of Reactions	Lifestyle/ Characteristic
<i>Candidatus Pseudomonas adelgestugas</i> (Adelges tsugae)	pade	CP026512	1,835,598	489	Endosymbiont
<i>Pseudomonas aeruginosa</i> B136-33	psg	CP004061	6,421,010	1,194	Human Pathogen
<i>Pseudomonas aeruginosa</i> c7447m	paec	CP006728	6,262,305	1,197	Human Pathogen
<i>Pseudomonas aeruginosa</i> DK2	pdk	CP003149	6,402,658	1,185	Human Pathogen
<i>Pseudomonas aeruginosa</i> LES431	pael	CP006937	6,550,070	1,192	Human Pathogen
<i>Pseudomonas aeruginosa</i> LESB58	pag	FM209186	6,601,757	1,194	Human Pathogen
<i>Pseudomonas aeruginosa</i> M18	paf	CP002496	6,327,754	1,192	Human Pathogen
<i>Pseudomonas aeruginosa</i> MTB-1	paem	CP006853	6,580,038	1,194	Human Pathogen
<i>Pseudomonas aeruginosa</i> NCGM 1900	paeb	AP014622	6,814,936	1,188	Human Pathogen
<i>Pseudomonas aeruginosa</i> NCGM2.S1	pnc	AP012280	6,764,661	1,195	Human Pathogen
<i>Pseudomonas aeruginosa</i> PA1	paep	CP004054	6,498,072	1,169	Human Pathogen
<i>Pseudomonas aeruginosa</i> PA1R	paer	CP004055	6,309,305	1,174	Human Pathogen
<i>Pseudomonas aeruginosa</i> PA38182	paeu	HG530068	7,586,152	1,200	Human Pathogen
<i>Pseudomonas aeruginosa</i> PA7	pap	CP000744	6,588,339	1,216	Human Pathogen
<i>Pseudomonas aeruginosa</i> PAO1	pae	AE004091	6,264,404	1,195	Human Pathogen
<i>Pseudomonas aeruginosa</i> PAO1-VE13	paev	CP006832	6,265,484	1,197	Human Pathogen
<i>Pseudomonas aeruginosa</i> PAO1-VE2	paei	CP006831	6,265,484	1,197	Human Pathogen
<i>Pseudomonas aeruginosa</i> PAO581	paeo	CP006705	6,043,974	1,186	Human Pathogen
<i>Pseudomonas aeruginosa</i> RP73	prp	CP006245	6,342,034	1,180	Human Pathogen
<i>Pseudomonas aeruginosa</i> SCV20265	paes	CP006931	6,725,183	1,191	Human Pathogen
<i>Pseudomonas aeruginosa</i> UCBPP-PA14	pau	CP000438	6,537,648	1,193	Human Pathogen
<i>Pseudomonas aeruginosa</i> YL84	paeg	CP007147	6,433,441	1,195	Human Pathogen
<i>Pseudomonas alcaligenes</i>	palc	CP014784	4,406,305	890	Human Pathogen
<i>Pseudomonas alkylphenolica</i>	palk	CP009048	5,764,622	1,141	Bioremediation
<i>Pseudomonas amygdali</i>	pamg	CP026558	6,109,228	1,105	Plant Pathogen
<i>Pseudomonas antarctica</i>	panr	CP015600	6,441,449	1,169	Psychrophile
<i>Pseudomonas avellanae</i>	pavl	CP026562	6,242,845	1,120	Plant Pathogen
<i>Pseudomonas azotoformans</i>	pazo	CP014546	6,859,618	1,201	Plant Pathogen
<i>Pseudomonas balearica</i>	pbm	CP007511	4,383,480	1,016	Denitrification
<i>Pseudomonas brassicacearum</i> DF41	pbc	CP007410	6,652,396	1,167	Soil bacterium

Continued from previous page					
<i>Pseudomonas brassicacearum</i> subsp. <i>brassicacearum</i> NFM421	pba	CP002585	6,843,248	1,208	Denitrification
<i>Pseudomonas chlororaphis</i> PA23	pch	CP008696	7,122,173	1,231	Biocontrol Activity
<i>Pseudomonas chlororaphis</i> PCL1606	pcz	CP011110	6,646,309	1,196	Biocontrol Activity
<i>Pseudomonas chlororaphis</i> subsp. <i>aurantiaca</i>	pcp	CP009290	6,702,062	1,197	Biocontrol Activity
<i>Pseudomonas cichorii</i>	pci	CP007039	5,986,012	1,064	Plant Pathogen
<i>Pseudomonas citronellolis</i>	pcq	CP014158	6,951,444	1,214	Biocontrol Activity
<i>Pseudomonas corrugata</i>	pcg	CP014262	6,124,363	1,160	Plant Pathogen
<i>Pseudomonas cremoricolorata</i>	psw	CP009455	4,780,403	1,021	Plant Pathogen
<i>Pseudomonas entomophila</i>	pen	CT573326	5,888,780	1,150	Insect Pathogen
<i>Pseudomonas fluorescens</i> A506	pfc	CP003041	5,962,570	1,158	Biocontrol Activity
<i>Pseudomonas fluorescens</i> F113	pfe	CP003150	6,845,832	1,223	Denitrification
<i>Pseudomonas fluorescens</i> Pf0-1	pfo	CP000094	6,438,405	1,182	Plant Protecting
<i>Pseudomonas fluorescens</i> SBW25	pfs	AM181176	6,722,539	1,191	Plant Protecting
<i>Pseudomonas fluorescens</i> UK4	pfn	CP008896	6,064,456	1,140	Plant Protecting
<i>Pseudomonas fragi</i>	pfz	CP013861	5,101,809	1,077	Psychrophile
<i>Pseudomonas frederiksbergensis</i>	pfk	CP018319	6,126,864	1,240	Denitrification
<i>Pseudomonas fulva</i>	pfv	CP002727	4,920,769	1,042	Non-Pathogen
<i>Pseudomonas knackmussii</i>	pkc	HG322950	6,162,905	1,195	Non-Pathogen
<i>Pseudomonas koreensis</i>	pkr	CP014947	6,301,761	1,134	Non-Pathogen
<i>Pseudomonas lundensis</i>	plq	CP017687	4,814,265	1,009	Non-Pathogen
<i>Pseudomonas lurida</i>	pfx	CP015639	6,175,426	1,185	Psychrophile
<i>Pseudomonas mandelii</i>	pman	CP005960	6,778,052	1,199	Psychrophile
<i>Pseudomonas mendocina</i> NK-01	pmk	CP002620	5,434,353	1,053	Denitrification
<i>Pseudomonas mendocina</i> ymp	pmy	CP000680	5,072,807	1,007	Opportunistic Pathogen
<i>Pseudomonas monteilii</i> SB3078	pmon	CP006978	6,000,087	1,161	Opportunistic Pathogen
<i>Pseudomonas monteilii</i> SB3101	pmot	CP006979	5,945,120	1,161	Opportunistic Pathogen
<i>Pseudomonas orientalis</i>	poi	CP018049	5,986,236	1,133	Plant Protecting
<i>Pseudomonas oryzihabitans</i>	por	CP013987	4,834,356	1,078	Animal Pathogen
<i>Pseudomonas parafulva</i>	ppv	CP009747	5,087,619	1,062	Plant Protecting
<i>Pseudomonas plecoglossicida</i>	ppj	CP010359	6,233,254	1,141	Animal Pathogen
<i>Pseudomonas poae</i>	ppz	CP004045	5,512,241	1,083	Plant Growth Promoting
<i>Pseudomonas protegens</i> Cab57	ppro	AP014522	6,827,892	1,211	Plant Protecting
<i>Pseudomonas protegens</i> CHA0	pprc	CP003190	6,867,980	1,221	Plant Protecting

Continued from previous page					
<i>Pseudomonas protegens</i> Pf-5	pfl	CP000076	7,074,893	1,222	Commensal
<i>Pseudomonas pseudoalcaligenes</i>	ppse	HG916826	4,686,340	995	Bioremediation
<i>Pseudomonas psychrotolerans</i>	ppsl	CP018758	5,271,920	1,195	Non-Pathogen
<i>Pseudomonas putida</i> BIRD-1	ppb	CP002290	5,731,541	1,131	Insect Pathogen
<i>Pseudomonas putida</i> DLL-E4	ppud	CP007620	6,484,062	1,122	Insect Pathogen
<i>Pseudomonas putida</i> DOT-T1E	ppx	CP003734	6,260,702	1,195	Insect Pathogen
<i>Pseudomonas putida</i> F1	ppf	CP000712	5,959,964	1,172	Insect Pathogen
<i>Pseudomonas putida</i> GB-1	ppg	CP000926	6,078,430	1,159	Insect Pathogen
<i>Pseudomonas putida</i> H8234	pput	CP005976	6,870,827	1,150	Insect Pathogen
<i>Pseudomonas putida</i> HB3267	ppuh	CP003738	5,875,750	1,132	Insect Pathogen
<i>Pseudomonas putida</i> KT2440	ppu	AE015451	6,181,873	1,131	Bioremediation
<i>Pseudomonas putida</i> NBRC 14164	ppun	AP013070	6,156,701	1,175	Opportunistic Human Pathogen
<i>Pseudomonas putida</i> ND6	ppi	CP003588	6,085,449	1,144	Insect Pathogen
<i>Pseudomonas putida</i> S16	ppt	CP002870	5,984,790	1,132	Insect Pathogen
<i>Pseudomonas putida</i> W619	ppw	CP000949	5,774,330	1,165	Insect Pathogen
<i>Pseudomonas resinovorans</i>	pre	AP013068	6,285,863	1,163	Bioremediation
<i>Pseudomonas rhizosphaerae</i>	prh	CP009533	4,688,635	1,058	Plant Protecting
<i>Pseudomonas savastanoi</i> pv. phaseolicola 1448A	psp	CP000058	5,928,787	1,096	Plant Pathogen
<i>Pseudomonas silesiensis</i>	psil	CP014870	6,823,539	1,279	Bioremediation
<i>Pseudomonas simiae</i> PCL1751	pfw	CP010896	6,143,950	1,158	Non-Pathogen
<i>Pseudomonas simiae</i> PICF7	pff	CP005975	6,136,735	1,154	Plant Pathogen
<i>Pseudomonas soli</i>	pmos	CP009365	6,247,860	1,091	Non-Pathogen
<i>Pseudomonas</i> sp. ATCC 13867	pdr	CP004143	5,696,307	1,142	Denitrification
<i>Pseudomonas</i> sp. CCOS 191	psec	LN847264	6,012,947	1,061	Phosphate solubilizing
<i>Pseudomonas</i> sp. MRSN12121	psem	CP010892	6,929,263	1,200	Opportunistic Human Pathogen
<i>Pseudomonas</i> sp. Os17	psos	AP014627	6,885,464	1,190	Biocontrol Activity
<i>Pseudomonas</i> sp. R2A2	psed	CP029772	4,559,447	1,013	Biocontrol Activity
<i>Pseudomonas</i> sp. StFLB209	pses	AP014637	6,332,373	1,127	Plant Pathogen
<i>Pseudomonas</i> sp. TCU-HL1	pset	CP015992	6,244,007	1,182	Non-Pathogen
<i>Pseudomonas</i> sp. TKP	psk	CP006852	7,012,672	1,216	Non-Pathogen
<i>Pseudomonas</i> sp. UW4	ppuu	CP003880	6,183,388	1,226	Plant Growth Promoting
<i>Pseudomonas</i> sp. VLB120	psv	CP003961	5,644,569	1,111	Bioremediation
<i>Pseudomonas stutzeri</i> 19SMN4	pstu	CP007509	4,725,662	1,111	Denitrification
<i>Pseudomonas stutzeri</i> 28a24	pstt	CP007441	4,731,359	1,040	Denitrification
<i>Pseudomonas stutzeri</i> A1501	psa	CP000304	4,567,418	1,048	Denitrification

Continued from previous page					
<i>Pseudomonas stutzeri</i> ATCC 17588	psz	CP002881	4,547,930	1,044	Denitrification
<i>Pseudomonas stutzeri</i> CCUG 29243	pse	CP003677	4,709,064	1,104	Denitrification
<i>Pseudomonas stutzeri</i> DSM 10701	psj	CP003725	4,174,118	985	Denitrification
<i>Pseudomonas stutzeri</i> DSM 4166	psr	CP002622	4,689,946	1,078	Denitrification
<i>Pseudomonas stutzeri</i> RCH2	psh	CP003071	4,575,057	1,044	Denitrification
<i>Pseudomonas synxantha</i> LBUM223	pfb	CP011117	6,690,033	1,149	Biocontrol Activity
<i>Pseudomonas syringae</i> CC1557	psyr	CP007014	5,758,024	1,086	Plant Pathogen
<i>Pseudomonas syringae</i> pv. <i>syringae</i> B728a	psb	CP000075	6,093,698	1,104	Plant pathogen
<i>Pseudomonas syringae</i> pv. <i>tomato</i> DC3000	pst	AE016853	6,397,126	1,119	Plant pathogen
<i>Pseudomonas trivialis</i>	ptv	CP011507	6,452,803	1,151	Commensalism
<i>Pseudomonas veronii</i>	pvr	CP018420	6,852,809	1,227	Bioremediation
<i>Pseudomonas versuta</i>	ppsy	CP012676	5,149,788	1,114	Psychrophile
<i>Pseudomonas yamanorum</i>	pym	CP012400	6,856,835	1,250	Psychrotolerant

The information about the lifestyle and characteristics of different *Pseudomonas* was collected from the KEGG PATHOGEN database (Kanehisa et al., 2017) and the available literature.

3.2 Comparison of metabolic networks in *Pseudomonas*

Comparison of metabolic networks in different *Pseudomonas* strains can reveal the extent of conservation and variability across different strains and provide insights into their pathogenic/non-pathogenic lifestyles and the niche in which they colonize. To systematically compare metabolic networks across 111 *Pseudomonas* strains, MAPPS was used to enumerate their metabolic networks for various KEGG pathway sets involved in central metabolism, including carbohydrate and amino acid metabolisms, and energy metabolism using ‘Network enumeration/comparison’ option (see **Section 2.2.4.5**). The comparisons were performed using custom scripts written in Python. The results are discussed in the subsequent sections.

3.2.1 Carbohydrate and amino acid metabolisms

Two major groups of metabolisms, namely carbohydrate and amino acid metabolisms, were selected for comparison. Carbohydrates are an essential source of energy that drives cellular reactions (McKee and McKee, 2016). Carbohydrate metabolism includes several essential pathways, such as glycolysis, gluconeogenesis, and citrate cycle, among others. The complete list of KEGG pathway maps included in the carbohydrate metabolism set is given in **Table 3.2**. The pathway maps involved in carbohydrate metabolism are linked to other parts of the metabolic network inside the cell via different compounds. For example, acetyl-CoA connects the carbohydrate metabolism to the metabolism of amino acids and other nutrients (McKee and McKee, 2016). Similarly, bacteria such as *Escherichia coli* or *Pseudomonas aeruginosa* can obtain the carbon skeletons for every amino acid, coenzyme, nucleotide, fatty acid, or other metabolic intermediates it needs for growth using glucose. Amino acid metabolism, on the other hand, includes the biosynthesis and biodegradation of amino acids and the assimilation of the amino group or carbon skeleton into other metabolic pathways. It also makes a significant contribution to the generation of metabolic energy through oxidative degradation. The fraction of metabolic energy obtained from amino acids, whether they are derived from dietary protein or tissue protein, varies significantly with the type of organism and with metabolic conditions (Lehninger, 2004). Thus, these two categories of metabolism provide a diverse range of metabolic reactions for comparison between different *Pseudomonas* strains.

To study the conservation of carbohydrate and amino acid metabolism across 111 *Pseudomonas* strains listed in **Table 3.1**, metabolic networks corresponding to KEGG pathways involved in carbohydrate and amino acid metabolisms (**Table 3.2**) for these

Table 3.2: List of pathway maps from KEGG relating to carbohydrate and amino acid metabolisms used for metabolic comparisons in *Pseudomonas*

KEGG ID	Pathway Map	Metabolism
MAP00010	Glycolysis / Gluconeogenesis	Carbohydrate metabolism
MAP00020	Citrate cycle (TCA cycle)	Carbohydrate metabolism
MAP00030	Pentose phosphate pathway	Carbohydrate metabolism
MAP00040	Pentose and glucuronate interconversions	Carbohydrate metabolism
MAP00051	Fructose and mannose metabolism	Carbohydrate metabolism
MAP00052	Galactose metabolism	Carbohydrate metabolism
MAP00053	Ascorbate and aldarate metabolism	Carbohydrate metabolism
MAP00220	Arginine biosynthesis	Amino acid metabolism
MAP00250	Alanine, aspartate and glutamate metabolism	Amino acid metabolism
MAP00260	Glycine, serine and threonine metabolism	Amino acid metabolism
MAP00270	Cysteine and methionine metabolism	Amino acid metabolism
MAP00280	Valine, leucine and isoleucine degradation	Amino acid metabolism
MAP00290	Valine, leucine and isoleucine biosynthesis	Amino acid metabolism
MAP00300	Lysine biosynthesis	Amino acid metabolism
MAP00310	Lysine degradation	Amino acid metabolism
MAP00330	Arginine and proline metabolism	Amino acid metabolism
MAP00340	Histidine metabolism	Amino acid metabolism
MAP00350	Tyrosine metabolism	Amino acid metabolism
MAP00360	Phenylalanine metabolism	Amino acid metabolism
MAP00380	Tryptophan metabolism	Amino acid metabolism
MAP00400	Phenylalanine, tyrosine and tryptophan biosynthesis	Amino acid metabolism
MAP00500	Starch and sucrose metabolism	Carbohydrate metabolism
MAP00520	Amino sugar and nucleotide sugar metabolism	Carbohydrate metabolism
MAP00562	Inositol phosphate metabolism	Carbohydrate metabolism
MAP00620	Pyruvate metabolism	Carbohydrate metabolism
MAP00630	Glyoxylate and dicarboxylate metabolism	Carbohydrate metabolism
MAP00640	Propanoate metabolism	Carbohydrate metabolism
MAP00650	Butanoate metabolism	Carbohydrate metabolism
MAP00660	C5-Branched dibasic acid metabolism	Carbohydrate metabolism

strains were enumerated. In addition, the KEGG reference network, which is a network containing all KEGG reactions across all domains of life, was also enumerated to be used as the reference for comparisons. The resulting reference network consisted of 1,445 reactions out of which 723 reactions (50%) were involved exclusively in the carbohydrate

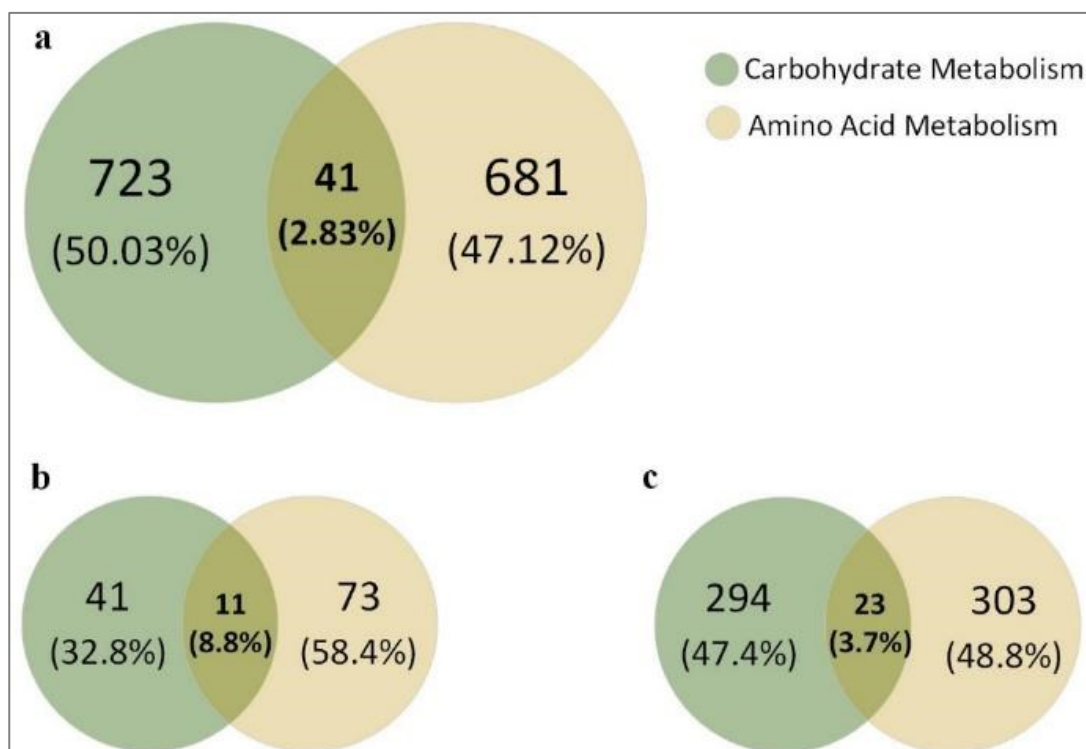


Figure 3.1: Number of reactions involved in carbohydrate and amino acid metabolisms predicted to be present in *Pseudomonas*. (a) Number of reactions in KEGG reference network (b) Number of reactions found to be conserved across 111 *Pseudomonas* strains (c) Number of reactions found to be variable across 111 *Pseudomonas* strains.

metabolism, 681 reactions (47.12%) were involved in the amino acid metabolism, and 41 reactions (2.83%) were present in both metabolisms (**Figure 3.1a**).

Comparison between the metabolic networks across the 111 *Pseudomonas* strains revealed that 125 reactions (8.6%) were conserved across all *Pseudomonas* strains, 620 reactions (42.9%) were present in one or more strains (variable reactions), and 700 reactions (48.4%) were completely absent in *Pseudomonas* but present in the KEGG reference network (**Figure 3.2**). The latter reactions were removed from subsequent analysis. The list of conserved and variable reactions is provided in **Appendix C** and **Appendix D**. Out of the 125 conserved reactions, 40 reactions (32.8%) belonged exclusively to carbohydrate metabolism, 74 reactions (58.4%) were involved in amino acid

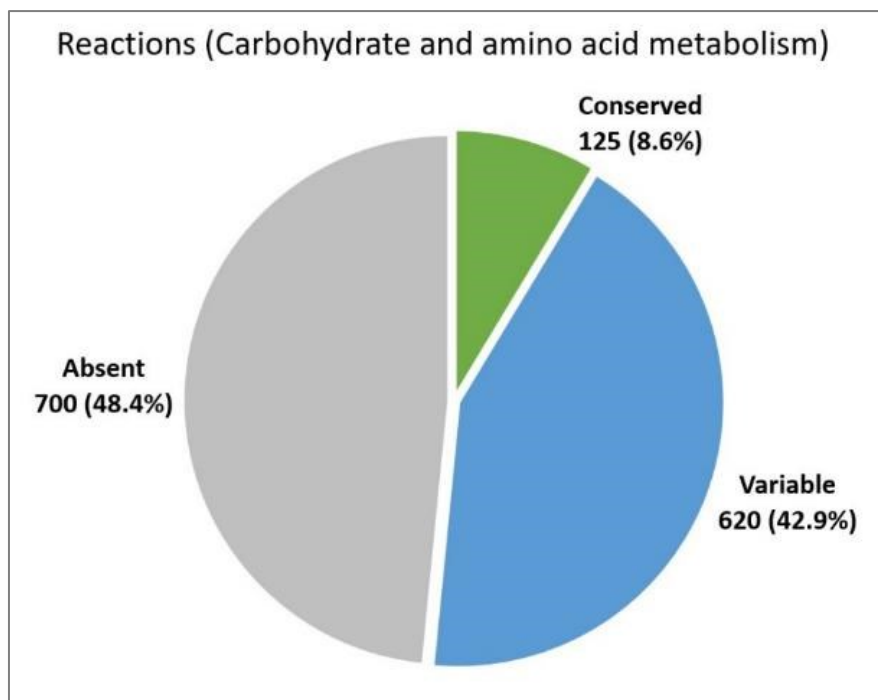


Figure 3.2: Distribution of KEGG reactions involved in carbohydrate and amino acid metabolism across different *Pseudomonas* strains. The figure shows the number of reactions which are conserved, variable and absent in 111 *Pseudomonas* strains.

metabolism, and 11 reactions (8.8%) were present in both carbohydrate and amino acid metabolism (**Figure 3.1b**). Similarly, out of 620 reactions which were found to be variable across pseudomonads, 294 reactions (47.4%) were solely involved in carbohydrate metabolism, 303 reactions (48.8%) were involved in amino acid metabolism, and 23 reactions (3.7%) were present in both carbohydrate and amino acid metabolism (**Figure 3.1c**).

Next, 'Network comparison' option available in MAPPS was used to identify the reactions present in one strain but absent in all other strains since these reactions are likely to provide unique traits to individual strains. A total of 26 reactions were found to be present in only one strain but absent from all other strains. These reactions along with their catalyzing enzymes and the KEGG pathway map(s) in which they are involved are listed

in **Table 3.3**. These uniquely present reactions reflect the distinct metabolic capabilities of these strains based on the KEGG annotations for carbohydrate and amino acid metabolism. For example, six reactions were predicted to be uniquely present in the insect pathogen *Pseudomonas entomophila* (KEGG organism code: pen). *P. entomophila* is a versatile pseudomonad found in different environments such as soil, water, and rhizosphere, and is capable of catabolizing various aromatic compounds and long-chain carbohydrates making it potentially useful for bioremediation and also shown pathogenicity against various insects (Dieppois et al., 2015). Similarly, the reaction R00262 catalyzed by glutamate mutase (EC 5.4.99.1), an enzyme involved in the degradation of L-Glutamate to Pyruvate (Buckel, 2001), is predicted to be uniquely present in *Pseudomonas poae* (KEGG organism code: ppz) suggesting a unique metabolic route for L-Glutamate assimilation in *P. poae*.

3.2.2 Energy metabolism

Energy metabolism is an essential component of cellular metabolism. It consists of KEGG maps of oxidative phosphorylation, which provides chemical energy by oxidizing nutrients, photosynthesis, carbon fixation pathways, methane, nitrogen, and sulfur metabolism (**Table 3.4**). *Pseudomonas* species inhabit a wide variety of habitats, ranging from the human body to soils, the rhizosphere, and the phyllosphere, and therefore utilize a diverse range of reactions in their pathways relating to energy metabolism (Udaondo et al., 2018; Xin et al., 2018; Gani et al., 2019). Metabolic networks relating to energy metabolism listed in **Table 3.4** were enumerated using MAPPS for the 111 *Pseudomonas* strains used in this study (**Table 3.1**) as well as for the KEGG reference network. The resulting reference network consisted of 227 reactions.

Table 3.3: Reactions involved in carbohydrate and amino acid metabolisms predicted to be present in only one pseudomonad but absent from all the others

Reaction Id	Reaction Equation	EC Number	KEGG Pathway map	Organism
R01063	D-Glyceraldehyde 3-phosphate + Orthophosphate + NADP+ \rightleftharpoons 3-Phospho-D-glyceroyl phosphate + NADPH + H+	1.2.1.13 1.2.1.59	Glycolysis / Gluconeogenesis	<i>Pseudomonas aeruginosa</i> PA38182
R02833	Chitosan + H ₂ O \rightleftharpoons D-Glucosaminide + Chitosan	3.2.1.132	Amino sugar and nucleotide sugar metabolism	<i>Pseudomonas alkylphenolica</i>
R10846	D-Galactaro-1,5-lactone \rightleftharpoons D-Galactaro-1,4-lactone	5.4.1.4	Ascorbate and aldarate metabolism	<i>Pseudomonas chlororaphis</i> PA23
R01525	D-Ribitol 5-phosphate + NADP+ \rightleftharpoons D-Ribulose 5-phosphate + NADPH + H+	1.1.1.137	Pentose and glucuronate interconversions	<i>Pseudomonas citronellolis</i>
R02921	CTP + D-Ribitol 5-phosphate \rightleftharpoons Diphosphate + CDP-ribitol	2.7.7.40	Pentose and glucuronate interconversions	<i>Pseudomonas citronellolis</i>
R00673	L-Tryptophan + H ₂ O \rightleftharpoons Indole + Pyruvate + Ammonia	4.1.99.1	Tryptophan metabolism	<i>Pseudomonas cremoricolorata</i>
R00308	1,3-beta-D-Glucan + H ₂ O \rightleftharpoons D-Glucose + 1,3-beta-D-Glucan	3.2.1.58	Starch and sucrose metabolism	<i>Pseudomonas entomophila</i>
R01974	Indolepyruvate \rightleftharpoons Indole-3-acetaldehyde + CO ₂	4.1.1.43 4.1.1.74	Tryptophan metabolism	<i>Pseudomonas entomophila</i>
R03115	1,3-beta-D-Glucan + H ₂ O \rightleftharpoons 1,3-beta-D-Glucan + alpha-D-Glucose	3.2.1.58	Starch and sucrose metabolism	<i>Pseudomonas entomophila</i>
R03168	Acetyl-CoA + N6-Hydroxy-L-lysine \rightleftharpoons CoA + N6-Acetyl-N6-hydroxy-L-lysine	2.3.1.102	Lysine degradation	<i>Pseudomonas entomophila</i>
R03629	Melatonin + [Reduced NADPH---hemoprotein reductase] + Oxygen \rightleftharpoons 6-Hydroxymelatonin + [Oxidized NADPH---hemoprotein reductase] + H ₂ O	1.14.14.1	Tryptophan metabolism	<i>Pseudomonas entomophila</i>
R03868	Maleylpyruvate \rightleftharpoons 3-Fumarylpyruvate	5.2.1.2 5.2.1.4	Tyrosine metabolism	<i>Pseudomonas entomophila</i>
R05133	Arbutin 6-phosphate + H ₂ O \rightleftharpoons Hydroquinone + beta-D-Glucose 6-phosphate	3.2.1.86	Glycolysis / Gluconeogenesis	<i>Pseudomonas fluorescens</i> A506

Continued from previous page				
R05134	Salicin 6-phosphate + H ₂ O \rightleftharpoons Salicyl alcohol + beta-D-Glucose 6-phosphate	3.2.1.86	Glycolysis / Gluconeogenesis	<i>Pseudomonas fluorescens</i> A506
R00650	S-Adenosyl-L-methionine + L-Homocysteine \rightleftharpoons S-Adenosyl-L-homocysteine + L-Methionine	2.1.1.10	Cysteine and methionine metabolism	<i>Pseudomonas fluorescens</i> Pf0-1
R02948	(S)-2-Acetolactate \rightleftharpoons (R)-Acetoin + CO ₂	4.1.1.5	Butanoate metabolism C5-Branched dibasic acid metabolism	<i>Pseudomonas lundensis</i>
R07613	LL-2,6-Diaminoheptanedioate + 2-Oxoglutarate \rightleftharpoons 2,3,4,5-Tetrahydrodipicolinate + L-Glutamate + H ₂ O	2.6.1.83	Lysine biosynthesis	<i>Pseudomonas orientalis</i>
R00262	L-threo-3-Methylaspartate \rightleftharpoons L-Glutamate	5.4.99.1	Glyoxylate and dicarboxylate metabolism	<i>Pseudomonas poae</i>
R01526	ATP + D-Ribulose \rightleftharpoons ADP + D-Ribulose 5-phosphate	2.7.1.47	Pentose and glucuronate interconversions	<i>Pseudomonas psychrotolerans</i>
R02429	D-Xylate \rightleftharpoons 2-Dehydro-3-deoxy-D-xylate + H ₂ O	4.2.1.82	Pentose and glucuronate interconversions	<i>Pseudomonas psychrotolerans</i>
R02754	5-Dehydro-4-deoxy-D-glucarate \rightleftharpoons Pyruvate + 2-Hydroxy-3-oxopropanoate	4.1.2.20	Ascorbate and aldarate metabolism	<i>Pseudomonas putida</i> H8234
R03277	2-Hydroxy-3-oxopropanoate + Pyruvate \rightleftharpoons 2-Dehydro-3-deoxy-D-glucarate	4.1.2.20	Ascorbate and aldarate metabolism	<i>Pseudomonas putida</i> H8234
R03028	Glutaconyl-CoA \rightleftharpoons Crotonyl-CoA + CO ₂	1.3.8.6	Butanoate metabolism	<i>Pseudomonas</i> sp. R2A2
R09293	(2S)-Methylsuccinyl-CoA + Electron-transferring flavoprotein \rightleftharpoons 2-Methylfumaryl-CoA + Reduced electron-transferring flavoprotein	1.3.8.12	Glyoxylate and dicarboxylate metabolism	<i>Pseudomonas</i> sp. TCU-HL1
R02755	meso-2,6-Diaminoheptanedioate + NADP ⁺ + H ₂ O \rightleftharpoons L-2-Amino-6-oxoheptanedioate + Ammonia + NADPH + H ⁺	1.4.1.16	Lysine biosynthesis	<i>Pseudomonas stutzeri</i> DSM 10701
R02749	2-Deoxy-D-ribose 1-phosphate \rightleftharpoons 2-Deoxy-D-ribose 5-phosphate	5.4.2.7	Pentose phosphate pathway	<i>Pseudomonas yamanorum</i>

Table 3.4: List of pathway maps from KEGG relating to energy metabolism used for metabolic comparisons in *Pseudomonas*

KEGG ID	Pathway Map
MAP00190	Oxidative phosphorylation
MAP00195	Photosynthesis
MAP00196	Photosynthesis - antenna proteins
MAP00680	Methane metabolism
MAP00710	Carbon fixation in photosynthetic organisms
MAP00720	Carbon fixation pathways in prokaryotes
MAP00910	Nitrogen metabolism
MAP00920	Sulfur metabolism

Comparison between these metabolic networks revealed that 33 reactions (14.5%) were conserved in all *Pseudomonas* strains, 77 reactions (33.9%) were variable, and 117 reactions (51.5%) were absent in all the *Pseudomonas* strains but present in the KEGG reference network (**Figure 3.3**). Like before, the reactions absent from all strains were removed from subsequent analysis. Lists of conserved and variable reactions are given in **Appendix E** and **Appendix F**. The all-but-one comparison identified six reactions to be present in only one strain but absent from all the other strains (**Table 3.5**). Three KEGG reactions related to carbon fixation and methane metabolism were predicted to be only present in the *Pseudomonas* sp. TCU-HL1. Similarly, R03025 catalyzed by coenzyme F₄₂₀ hydrogenase (EC 1.12.98.1) was predicted to be only present in the metabolic network of *P. balearica*, an environmental bacterium that has been mostly isolated from polluted environments (Bennasar et al., 1996). Besides this, R09513 catalyzed by the enzyme methanesulfonate monooxygenase (NADH) (EC 1.14.13.111) was predicted to be uniquely present in *P. pseudoalcaligenes*, an aerobic soil bacterium first isolated from swimming pool water (Monias, 1928). Methanesulfonate is a highly stable compound and is used by diverse aerobic bacteria as a source of sulfur for growth (Seitz et al., 1993).

Table 3.5: Reactions involved in from energy metabolism predicted to be present in only one pseudomonad but absent from the others

Reaction Id	Reaction Equation	EC Number	KEGG Pathway map	Organism
R01063	D-Glyceraldehyde 3-phosphate + Orthophosphate + NADP+ <=> 3-Phospho-D-glyceroyl phosphate + NADPH + H+	1.2.1.13 1.2.1.59	Carbon fixation in photosynthetic organisms	<i>Pseudomonas aeruginosa</i> PA38182
R03025	Coenzyme F420 + Hydrogen <=> Reduced coenzyme F420	1.12.98.1	Methane metabolism	<i>Pseudomonas balearica</i>
R09513	Methanesulfonic acid + NADH + H+ + Oxygen <=> Formaldehyde + NAD+ + Sulfite + H2O R10150	1.14.13.111	Sulfur metabolism	<i>Pseudomonas pseudoalcaligene</i>
R02560	Trimethylamine + 2 Ferricytochrome c + H2O <=> Trimethylamine N-oxide + 2 Ferrocycytochrome c + 2 H+	1.7.2.3	Methane metabolism	<i>Pseudomonas</i> sp. TCU-HL1
R09282	3-Methylfumaryl-CoA + H2O <=> (3S)-Citramalyl-CoA	4.2.1.153	Carbon fixation pathways in prokaryotes	<i>Pseudomonas</i> sp. TCU-HL1
R09283	2-Methylfumaryl-CoA <=> 3-Methylfumaryl-CoA	5.4.1.3	Carbon fixation pathways in prokaryotes	<i>Pseudomonas</i> sp. TCU-HL1

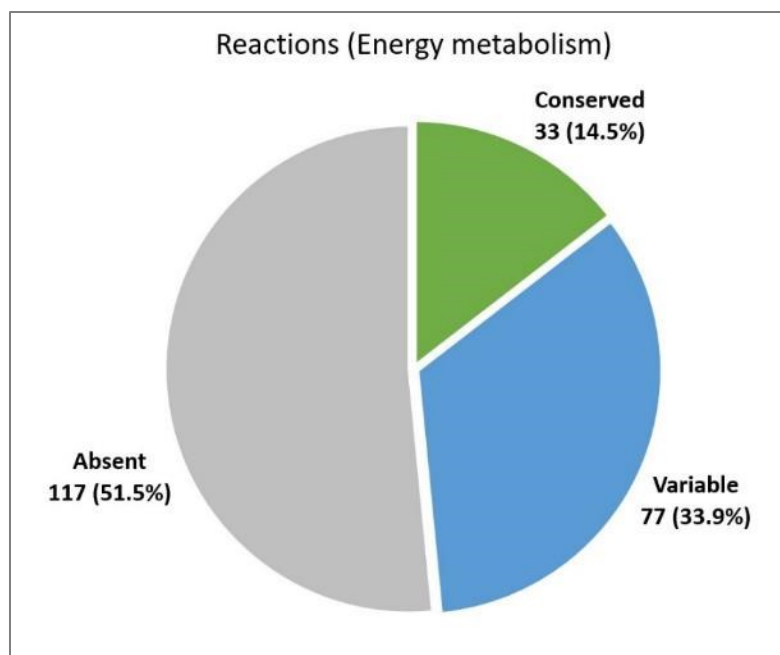


Figure 3.3: Distribution of KEGG reactions involved in energy metabolism across different *Pseudomonas* strains. The figure shows the number of reactions which are conserved, variable and absent in 111 *Pseudomonas* strains.

3.3 Deciphering the patterns of amino acid assimilation pathways in *Pseudomonas*

Analysis of metabolic pathways plays a vital role in deciphering the evolutionary relationships and functional specializations of closely related organisms (Yamanishi et al., 2007; Mithani et al., 2011). With the improvement in sequencing technologies, the number of complete genome sequences of *Pseudomonas* strains in public databases is rapidly growing. This has provided an opportunity to enhance our understanding of metabolic capabilities and diversity in *Pseudomonas*, and to study how metabolic pathways have evolved across different *Pseudomonas* lineages. For example, it has been shown that the genes encoding the carbohydrate catabolic pathways in *Pseudomonas* are organized in operons that are under the control of different regulators that respond differentially to distinct pathway intermediates (Udaondo et al., 2018). Similarly, studies focusing on the

Table 3.6: Number of distinct pathways of various lengths predicted from amino acids to ammonia across all *Pseudomonas* strains

KEGG Code	Amino Acid	Pathway Length						Total Pathways
		1	2	3	4	5	6	
C00025	L-Glutamate	4	-	5	-	-	-	9
C00037	Glycine	-	-	-	-	-	-	0
C00041	L-Alanine	-	-	-	-	-	118	118
C00047	L-Lysine	-	-	-	-	-	-	0
C00049	L-Aspartate	-	-	-	-	-	-	0
C00062	L-Arginine	-	-	-	-	60	-	60
C00064	L-Glutamine	-	6	-	-	-	-	6
C00065	L-Serine	-	-	-	-	-	-	0
C00073	L-Methionine	-	-	-	-	-	-	0
C00078	L-Tryptophan	-	-	-	-	-	-	0
C00079	L-Phenylalanine	-	-	-	-	-	-	0
C00082	L-Tyrosine	-	-	-	-	-	-	0
C00097	L-Cysteine	-	-	-	-	-	30	30
C00123	L-Leucine	-	-	-	-	-	-	0
C00135	L-Histidine	-	-	-	-	-	-	0
C00148	L-Proline	-	-	38	2	40	4	84
C00152	L-Asparagine	-	-	-	-	-	-	0
C00183	L-Valine	-	-	-	-	-	-	0
C00188	L-Threonine	-	-	-	-	-	-	0
C00263	L-Homoserine	-	-	-	-	-	-	0
C00334	4-Aminobutanoate	-	-	-	-	-	-	0
C00407	L-Isoleucine	-	-	-	-	-	-	0
Total Pathways		4	6	43	2	100	152	307

evolution of metabolic networks have identified genes and pathways which play essential roles in antibiotic resistance and adaptation to different niches (Cabot et al., 2016; Botelho et al., 2019).

To study how different pseudomonads use various amino acids as nitrogen and carbon sources and how these pathways are conserved across 111 *Pseudomonas* strains (Table 3.1), metabolic pathways were predicted using the pathway prediction option in MAPPS (see Section 2.2.4.1) from 22 amino acids listed in Table 3.6 to ammonia and the

TCA cycle intermediates (2-Oxoglutarate, Oxaloacetate, Succinyl-CoA, and Fumarate). To avoid false positives, the search was restricted to carbohydrate and the amino acid metabolism datasets, and the connections between reactions were established using the KEGG RCLASS option. Moreover, minimum and maximum pathway lengths were set to one and six reactions respectively, and the default list of ignored metabolites provided in **Appendix B** was used to filter out ubiquitous metabolites. The results for the pathway prediction are discussed below.

3.3.1 Pathway prediction from amino acids to ammonia

To study the assimilation pathways in *Pseudomonas*, metabolic pathways of length one to six reactions were predicted from 22 different amino acids to ammonia in 111 genome-sequenced pseudomonads listed in **Table 3.1** using MAPPS. A total of 307 distinct nitrogen assimilation pathways originating from six out of twenty-two amino acids were found across 111 *Pseudomonas* strains (**Table 3.6** and **Figure 3.4**). These include L-Alanine, L-Arginine, L-Cysteine, L-Glutamate, L-Glutamine and L-Proline. No pathways were reported from the remaining sixteen amino acids. The numbers of pathways of various lengths originating from different amino acids in each strain are listed in **Appendix G**. The shortest route to ammonia consisted of a single step and originated from L-Glutamate (**Table 3.6**). The second shortest route consisting of two reactions originated from L-Glutamine with six distinct pathways reported across different pseudomonads. The longest pathways consisting of six reactions originated from three amino acids including L-Alanine (118 pathways), L-Cysteine (30 pathways) and L-Proline (4 pathways).

When considering individual strains, *Pseudomonas putida* H8234 was reported to have the highest number of pathways (200 pathways) starting from L-Alanine (118

pathways), L-Cysteine (30 pathways), L-Proline (42 pathways), L-Glutamate (7 pathways) and L-Glutamine (3 pathways). The lowest number of nitrogen assimilation pathways (7 pathways) were found in the metabolic network of *Candidatus Pseudomonas adelgestugas*. Six of these were three-step pathways starting from L-Proline, and one single-step pathway from L-Glutamate. This is in agreement with the limited amino acid utilization profile of this pseudomonad (Weglarz et al., 2018). Seventy-one strains, most of which are pathogens including various strains of *Pseudomonas aeruginosa*, *Pseudomonas simiae*, *Pseudomonas putida*, *Pseudomonas syringae*, *Pseudomonas monteilii*, and *Pseudomonas mendocina* were found to have 52 distinct pathways to ammonia out of which 42 started from L-Proline, seven from L-Glutamate and three from L-Glutamine (**Appendix G**).

Next, pathway-based comparison was performed using custom scripts to identify metabolic pathways from various amino acids to ammonia which were present in only one strain but absent from all the others. Comparison results suggested the presence of unique metabolic pathways in only two *Pseudomonas* strains including *P. putida* H8234 (from L-Alanine and L-Cysteine) and *P. lurida* (from L-Proline) (**Appendix H**). *P. putida* H8234 was found to have the highest number of unique nitrogen assimilation pathways that were missing from all other *Pseudomonas* strains. These include 118 six-step pathways from L-Alanine and 30 pathways of length six from L-Cysteine. In the case of *P. lurida*, a psychrophile, only one pathway from L-Proline consisting of four reactions was found to be uniquely present.

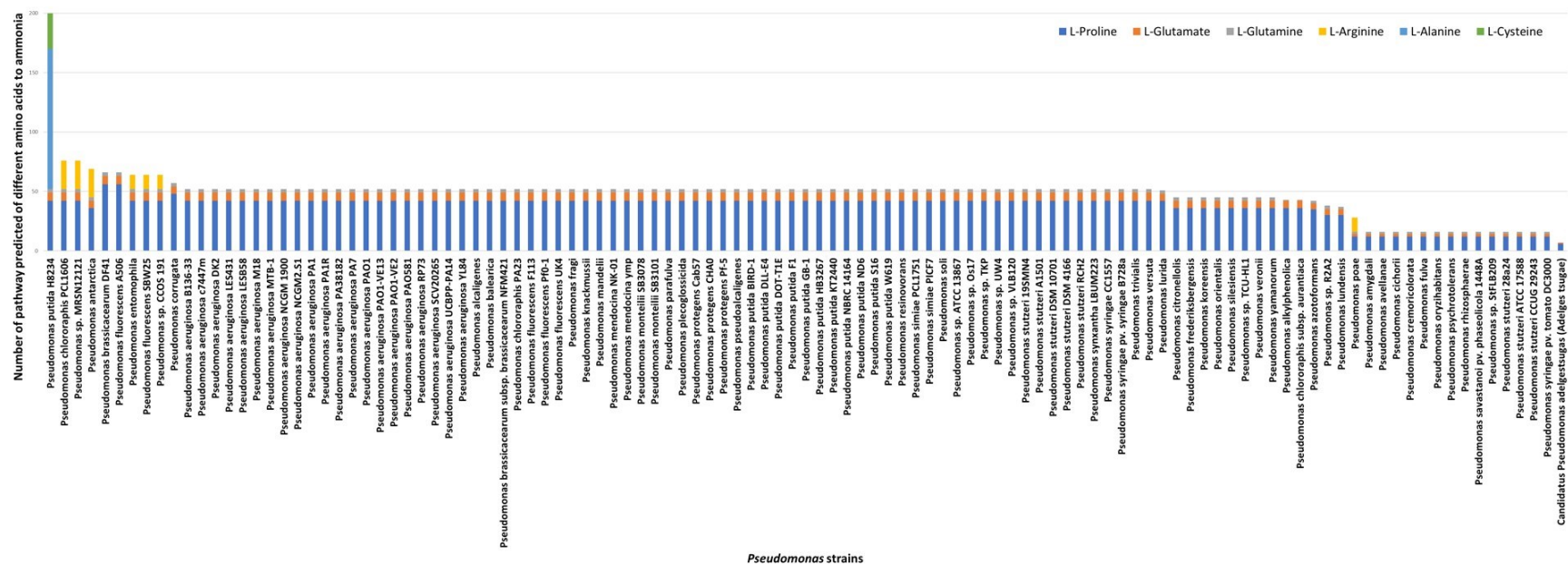


Figure 3.4: Number of pathways predicted in different pseudomonads from amino acids to ammonia. Pathways of length two to six reactions were predicted from 22 amino acids to ammonia restricting the search to carbohydrate and amino acid metabolism dataset.

Table 3.7: Number of distinct pathways of various lengths predicted from amino acids to the TCA cycle intermediate

KEGG Code	Amino Acid	Pathway Length						Total Pathways
		1	2	3	4	5	6	
C00025	L-Glutamate	40	0	0	0	0	0	40
C00037	Glycine	0	0	4	7	36	29	76
C00041	L-Alanine	0	7	64	4	69	128	272
C00047	L-Lysine	0	0	0	0	0	0	0
C00049	L-Aspartate	2	0	0	0	0	0	2
C00062	L-Arginine	0	0	0	0	468	0	468
C00064	L-Glutamine	0	40	0	123	0	0	163
C00065	L-Serine	0	0	0	0	0	0	0
C00073	L-Methionine	0	0	0	0	3	21	24
C00078	L-Tryptophan	0	0	0	3	28	32	63
C00079	L-Phenylalanine	0	0	0	0	0	0	0
C00082	L-Tyrosine	0	0	0	0	0	0	0
C00097	L-Cysteine	0	5	33	1	35	42	116
C00123	L-Leucine	0	0	0	0	0	0	0
C00135	L-Histidine	0	0	0	0	0	0	0
C00148	L-Proline	0	0	306	22	0	0	328
C00152	L-Asparagine	0	2	0	0	0	0	2
C00183	L-Valine	0	0	0	0	0	0	0
C00188	L-Threonine	0	0	0	0	0	0	0
C00263	L-Homoserine	0	0	0	0	3	18	21
C00334	4-Aminobutanoate	0	0	0	0	0	0	0
C00407	L-Isoleucine	0	0	0	0	0	0	0
Total Pathways		42	54	407	160	642	270	1575

3.3.2 Pathway prediction from amino acids to TCA cycle intermediates

To study carbon assimilation pathways in *Pseudomonas*, metabolic pathways were predicted from 22 amino acids (**Table 3.7**) to the four TCA cycle intermediates including 2-Oxoglutarate (KEGG ID: C00026), Oxaloacetate (KEGG ID: C00036), Succinyl-CoA (KEGG ID: C00091) and Fumarate (KEGG ID: C00122). These four metabolites directly link the KEGG map of the TCA Cycle with other KEGG pathway maps (**Figure 3.5**). A total of 1,575 distinct pathways of length one to six reactions were found from twelve out

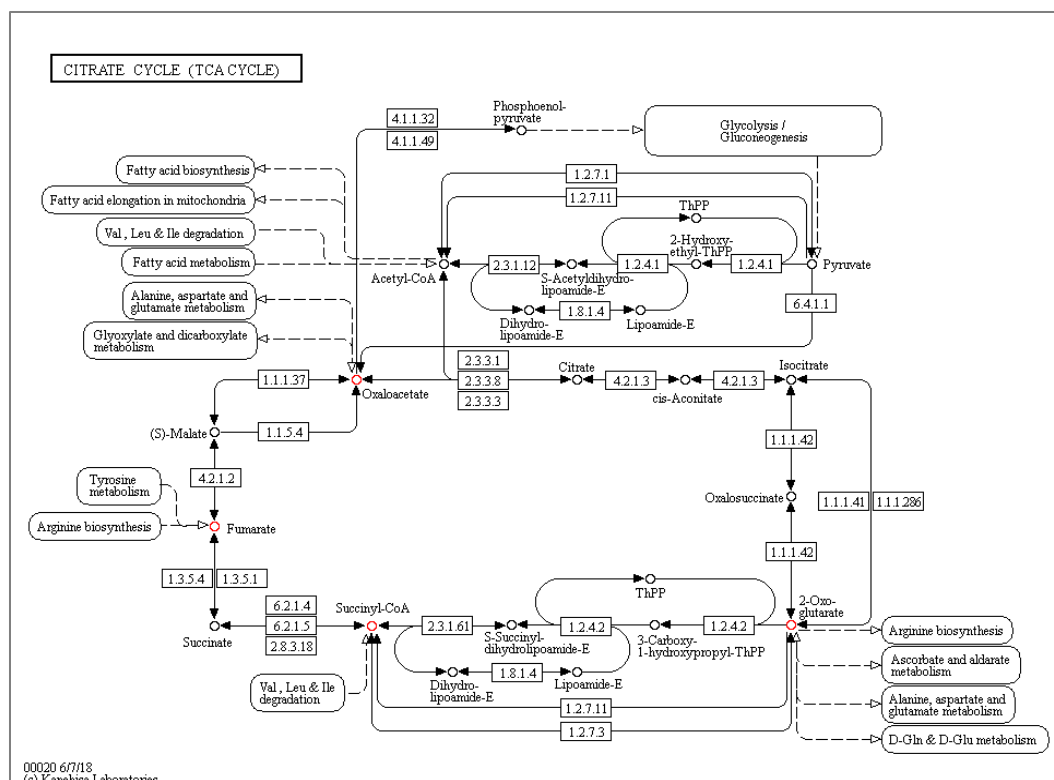


Figure 3.5: KEGG map of Tricarboxylic Acid (TCA) cycle. TCA cycle intermediates used as target metabolites for pathway prediction are highlighted as red circles.

of twenty-two amino acids across 111 *Pseudomonas* strains (**Table 3.7** and **Figure 3.6**). No pathways were found for the remaining ten amino acids. The numbers of pathways of various lengths originating from different amino acids in each strain are listed in **Appendix I**. Unlike nitrogen assimilation pathways where the shortest pathway consisting of a single reaction originated from only L-Glutamate, the shortest routes consisting to the four TCA cycle intermediates originated from L-Aspartate in addition to L-Glutamate. Similarly, two-step pathways to the TCA cycle intermediates originated four amino acids compared to only one (L-Glutamine) in the case of ammonia. These include L-Alanine, L-Asparagine, L-Cysteine, and L-Glutamine. Between these six amino acids, a total of 42 one-step and 54 two-step pathways were found across all *Pseudomonas* strains (**Table 3.7**).

When all pathways of length between one to six reactions were taken into account, *P. antarctica* PAMC 27494, a non-pathogenic psychrophile capable of producing antimicrobial compounds (Lee et al., 2017), was found to have the highest number of metabolic routes (540 pathways) from various amino acids to the TCA cycle intermediate. These include two hundred and twenty-eight pathways from L-Arginine, twelve pathways from L-Proline, six pathways from L-Glutamine, and one pathway from L-Asparagine. *P. chlororaphis* PCL1606, also a non-pathogen and capable of producing antifungal compounds (Calderón et al., 2015), and *Pseudomonas sp.* MRSN12121, a clinical isolate, were predicted to have 528 pathways of varying lengths to the TCA Cycle intermediates from eleven amino acids including Glycine, L-Alanine, L-Arginine, L-Asparagine, L-Aspartate, L-Cysteine, L-Glutamate, L-Glutamine, L-Homoserine, L-Methionine, L-Proline and L-Tryptophan. As in the case of nitrogen assimilation pathways, *Ca. Pseudomonas adelgestugas* had the lowest number of metabolic routes to the TCA cycle intermediates (forty-two pathways) which originated from L-Glutamate (six pathways) and L-Proline (36 pathways) further strengthening the hypothesis about its limited metabolic capability (Weglarz et al., 2018).

Pathway comparison using custom scripts suggested the presence of several unique pathways from L-Alanine, L-Glutamate, L-Glutamine, L-Tryptophan, L-Cysteine, L-Proline, L-Arginine, L-Homoserine and Glycine in various pseudomonads. The number of pathways for each amino acid is listed in **Appendix J**. The highest number of unique pathways were found to be present in *P. antarctica* with twenty-four five-step pathways originating from L-Arginine. Similarly, twelve unique pathways were predicted in *Pseudomonas rhizosphaerae* including two three-step pathways from L-Alanine, two six-step pathways from L-Tryptophan and eight pathways consisting of four to five reactions from Glycine. *P. putida* H8234, which also contained unique pathways to ammonia (see **Section 3.3.1**), was found to have ten unique pathways to the TCA Cycle intermediates including eight four- and six-step pathways from L-Alanine, and a four-step and a six-step pathway from L-Cysteine.

3.4 16S rRNA and Multilocus Sequence Analysis

16S rRNA sequences have been widely used for deciphering the relationship between different species and strains as well as for the allocation of newly found strains in the genus or family (Anzai et al., 2000). While comparisons based on 16S rRNA sequence data provide an invaluable framework for phylogenetic studies, they sometimes do not allow sufficient resolution to distinguish between closely related species. Alternatively, Multilocus Sequence Analysis (MLSA), has been used for generating a high-resolution phylogenetic relationship between different species (Glaeser and Kämpfer, 2015). In MLSA studies, the sequences of housekeeping genes coding for proteins with conserved functions are used to generate phylogenetic trees and subsequently deduce phylogenies. Compared to single gene-based phylogenies, MLSA gives the phylogenetic resolution needed for species delineation, especially for closely related

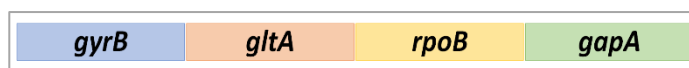


Figure 3.7: A schematic diagram showing concatenation of housekeeping gene sequences for MLSA. Gene sequences for four housekeeping genes (*gyrB*, *gltA*, *rpoB* and *gapA*) for 111 *Pseudomonas* strains and *Escherichia coli* were downloaded from KEGG GENES database and NCBI GenBank database.

species, and avoids possible misleading results that attribute to sequences based on single genes that may be affected by recombination (Didelot and Falush, 2007). An essential advantage of the MLSA method is that it provides information on the nucleotide divergence of internal fragments of housekeeping genes.

When applied to *Pseudomonas*, both 16S rRNA sequences and MLSA can provide useful insights into the phylogenetic relationship between different strains and provide clues about their evolution. To this end, phylogenetic trees for 111 *Pseudomonas* strains listed in **Table 3.1** were generated using both approaches, including 16S rRNA sequences and MLSA using four housekeeping genes (*gyrB*, *gltA*, *rpoB*, and *gapA*) (**Figure 3.7**). Previous studies have shown that these genes have greater evolutionary divergence for the bacteria especially closely related-species, hence can be used for the identification and classification of strains (Fukushima et al., 2002; Sarkar and Guttman, 2004; Sakamoto and Ohkuma, 2011; Gomila et al., 2015; Flores et al., 2018). *Escherichia coli* used as an outgroup in both phylogenetic trees.

First, gene sequences were downloaded from the KEGG GENES database (Kanehisa et al., 2016) and the NCBI GenBank database (Clark et al., 2016) for 111 *Pseudomonas* strains and *E. coli*. Two separate FASTA files were created. The first file contained the sequences of the 16S rRNA gene for all the strains, whereas the second file contained the concatenated sequence of four housekeeping genes of all the strains. Next, CLUSTAL OMEGA (Madeira et al., 2019) was used to generate multiple sequence alignments, and the resulting alignments were downloaded in PHYLIP format

to generate the phylogenetic trees using the PHYLIP package. Distance matrices were calculated using the F84 method (Felsenstein, 1996), and the dendrogram was generated using the Fitch method (Fitch and Margoliash, 1967). Resulting trees were visualized in MEGAX (Kumar et al., 2018) software and are shown in **Figure 3.8** and **Figure 3.9** for the 16S rRNA sequence and MLSA, respectively.

While both phylogenetic trees captured the overall phylogenetic relationship between different *Pseudomonas* species (Gomila et al., 2015), there were some subtle differences in the trees generated using 16S rRNA sequences (henceforth called ‘16S tree’) and MLSA approaches (henceforth called ‘MLSA tree’). For example, both methods have correctly grouped all the *Pseudomonas putida*, an insect and human pathogen, strains grouped including *P. putida* HB3267, which has high number of unique carbon and nitrogen assimilation pathways (see **Section 3.3**). *P. monteilii*, which is closely related to *P. putida* (Aditi et al., 2017), grouped together with various strains of *P. putida* in both the trees, as previously reported (Gomila et al., 2015). Interestingly, the MLSA tree suggests a smaller evolutionary distance between *P. putida* HB3267 and other *P. putida* strains compared to the 16S tree. This is most likely due to the fact that while both the 16S rRNA sequence and MLSA methods are devised for species classification, the protein-coding genes tend to evolve at a slower but constant rate compared to non-protein-coding genes and thus have a better resolution power (Glaeser and Kämpfer, 2015). *P. entomophila*, which is also an insect pathogen like *P. putida* and is found in the soil was grouped closer to *P. putida* strains by both the methods but had a greater evolutionary distance from other *P. putida* strains compared to *P. putida* HB3267. This was further confirmed by the metabolic similarity analysis which grouped *P. entomophila* with the strains of *P. putida* by only one of the three linkage methods (see **Section 3.5**).

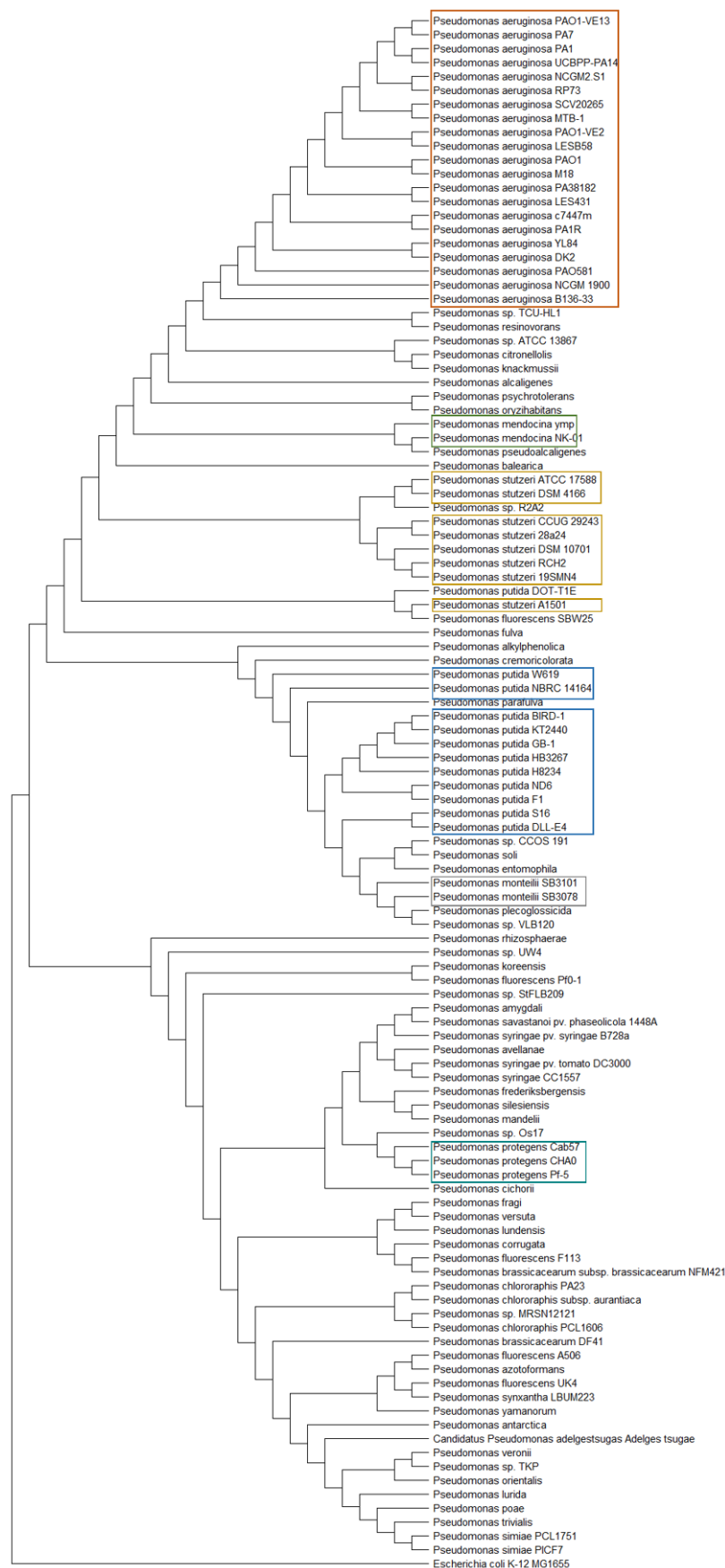


Figure 3.8: Phylogenetic tree of the 111 pseudomonads constructed using 16S rRNA sequences. Distance matrices were calculated by the F84 method (Felsenstein, 1996) and dendrogram was generated by Fitch method (Fitch and Margoliash, 1967). *Escherichia coli* was used as the outgroup.

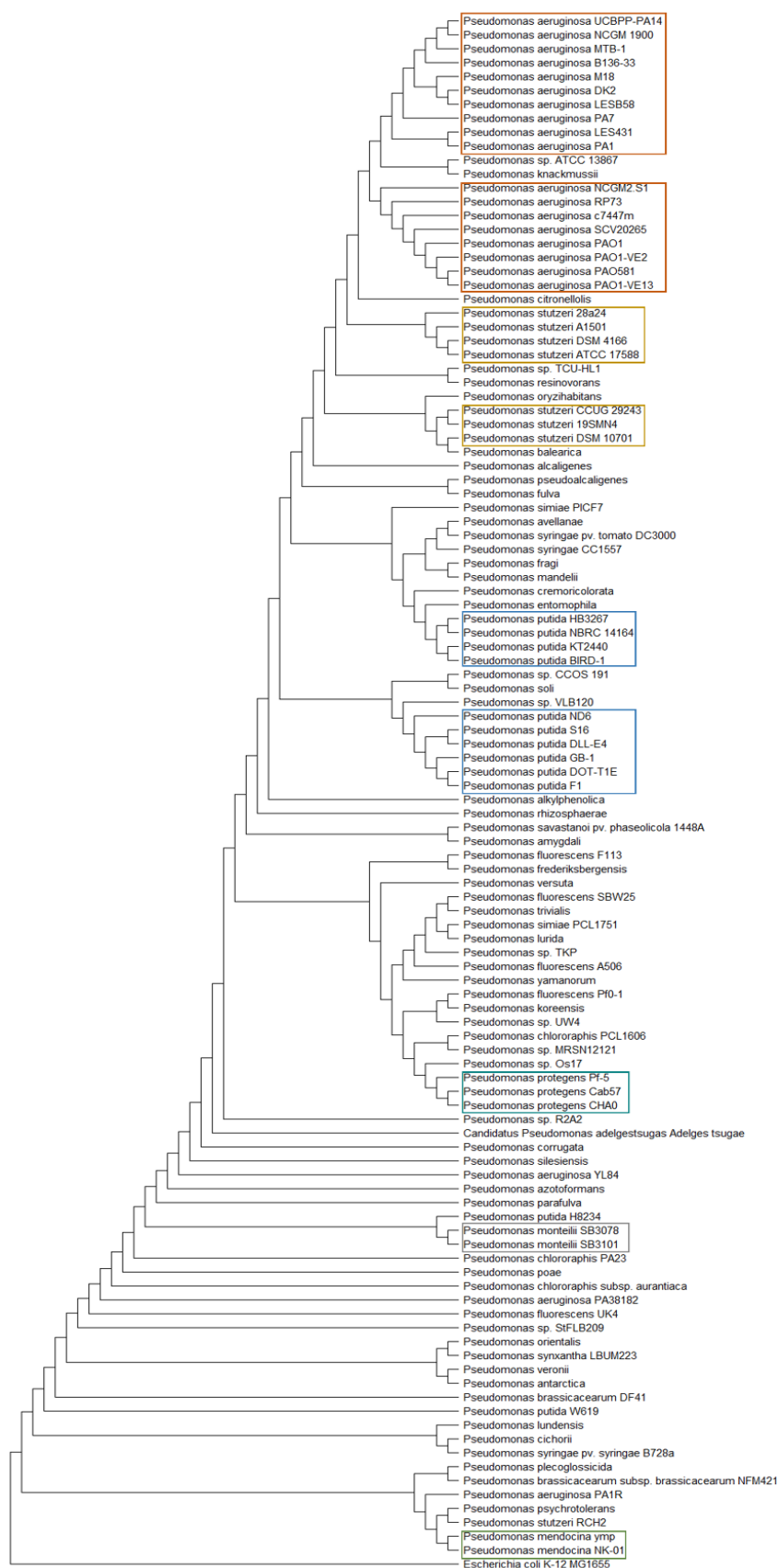


Figure 3.9: Phylogenetic tree of the 111 pseudomonads constructed using MLSA. MLSA was performed using four housekeeping genes (*gyrB*, *gltA*, *rpoB* and *gapA*). Distance matrices were calculated by the F84 method (Felsenstein, 1996) and dendrogram was generated by Fitch method (Fitch and Margoliash, 1967). *Escherichia coli* was used as the outgroup.

P. aeruginosa, an opportunistic human pathogen, is the leading cause of lung infections patients with cystic fibrosis (Subedi et al., 2018). All of the twenty-one *P. aeruginosa* strains listed in **Table 3.1** were grouped in the 16S tree (**Figure 3.8**). Similar result was obtained when all the *Pseudomonas* strains were clustered based on their metabolic similarity (see **Section 3.5**). The MLSA tree, on the other hand, grouped eighteen out of twenty-one strains together. The tree strains which did not group with the other *P. aeruginosa* strains in the MLSA tree included *Pseudomonas aeruginosa* PA38182, *Pseudomonas aeruginosa* YL84 and *Pseudomonas aeruginosa* PA1R.

In the case of plant-related *Pseudomonas*, pathogens *Pseudomonas syringae* pv. tomato DC3000, *Pseudomonas syringae* pv. *syringae* B728a (KEGG organism code: psb), *Pseudomonas syringae* CC1557, *Pseudomonas savastanoi* pv. *phaseolicola* 1448A, *Pseudomonas avellanae* and *Pseudomonas cichorii* were placed closer to each other in the 16S tree compared to MLSA tree. Interestingly, 16S tree put the two strains of *P. syringae*, which is *P. syringae* pv. tomato DC3000 and *P. syringae* CC1557, together compared to MLSA tree, which grouped *P. syringae* pv. tomato DC3000 with another plant pathogen *P. avellanae*.

Pseudomonas stutzeri is involved in soil denitrification and also identified as an opportunistic pathogen (Lalucat et al., 2006). Five out of eight *Pseudomonas stutzeri* strains including *Pseudomonas stutzeri* 19SMN4, *Pseudomonas stutzeri* RCH2, *Pseudomonas stutzeri* DSM 10701, *Pseudomonas stutzeri* 28a24 and *Pseudomonas stutzeri* CCUG 29243 were placed together while other three strains *Pseudomonas stutzeri* ATCC 17588, *Pseudomonas stutzeri* DSM 4166 and *Pseudomonas stutzeri* A1501 were placed in the neighboring clade in the 16S tree compared to MLSA tree which puts together only three of the five strains grouped together by 16S tree, *P. stutzeri* 19SMN4, *P. stutzeri* DSM 10701 and *P. stutzeri* CCUG 29243 with same

ancestral node while putting the strain *P. stutzeri* RCH2 farther apart. *P. stutzeri* 28a24 was placed in the clade containing the remaining three strains of *P. stutzeri*. *P. mendocina*, which is also an opportunistic human pathogen, although the infections are extremely rare (Gani et al., 2019) were placed closer to *P. aeruginosa* in the 16S tree but grouped with *P. stutzeri* RCH2 in the MLSA tree.

P. fluorescens species complex has great diversity and is divided into taxonomic subclades based on genomic differences, which also correlate with their functional diversity (Garrido-Sanz et al., 2016). *Pseudomonas fluorescens* Pf0-1 was grouped with *Pseudomonas koreensis*, which is also a non-pathogen, in both 16S and MLSA trees. Similarity between *P. fluorescens* Pf0-1 with *P. koreensis* has been reported by previous studies (Gomila et al., 2015; Garrido-Sanz et al., 2016). Similarly, *Pseudomonas fluorescens* F113 was found to group with *Pseudomonas brassicacearum* subsp. *brassicacearum* NFM421 and *Pseudomonas corrugata* in the 16S trees, which is in agreement with a previous study (Garrido-Sanz et al., 2016). *Pseudomonas fluorescens* A506 (KEGG organism code: pfc), a strain belonging to *fluorescens* subgroup (Garrido-Sanz et al., 2016) was placed close to *Pseudomonas azotoformans* and *Pseudomonas synxantha*, also members of *fluorescens* subgroup (Gomila et al., 2015) in the 16S tree while the MLSA tree puts it in the clade containing *Pseudomonas fluoresces* SBW25, *Pseudomonas trivialis* and *P. lurida*, which are also part of *fluorescens* subgroup (Gomila et al., 2015).

Besides this, various plant protecting pseudomonads showing biocontrol activity against plant-associated pathogens including fungi, were found grouped close to each other. *P. protegens* which were previously classified as *P. fluorescens* (Ramette et al., 2011) grouped together in both the trees while various strains of *Pseudomonas chlororaphis* including *P. chlororaphis* PA23, *P. chlororaphis* subsp. *aurantiaca*, and

P. chlororaphis PCL1606 (Jiao et al., 2013; Loewen et al., 2014; Calderón et al., 2015) grouped together in the 16S tree but not in the MLSA tree. *Pseudomonas veronii* and *P. orientalis* which have antagonistic properties against phytopathogens (Montes et al., 2016; Zengerer et al., 2018), were also placed closely to each other in both the trees.

3.5 Metabolic Similarity Analysis

Metabolic Similarity Analysis (MSA) is a distance-based hierarchical clustering approach, which can be used to group different strains of *Pseudomonas* based on their similarity at a metabolic level. To this end, metabolic similarity analysis (henceforth called ‘MSA’) function of MAPPS (see **Section 2.2.4.6**) was used to identify the clusters of pseudomonads (**Table 3.1**) based on their similarity at enzyme and reaction levels. KEGG maps relating to carbohydrate and amino acid metabolism (**Table 3.2**) were used to build the metabolic networks MSA was performed on the resulting networks using three linkage modes including single linkage, maximum linkage, and average linkage in MAPPS. Output generated in Newick format was visualized in Interactive Tree of Life (iTOL) (Letunic and Bork, 2019). The results for MSA are discussed below.

In single linkage clustering, which groups the data based on the closest pair of elements in the clusters, nineteen out of twenty-one strains of *P. aeruginosa* were grouped together irrespective of whether reaction or enzyme data was used for clustering (**Figure 3.10** and **Figure 3.11**). *Pseudomonas aeruginosa* PA1 and *Pseudomonas aeruginosa* PA1R did not cluster together with other *P. aeruginosa* strains using the enzyme data (**Figure 3.10**). Reaction-based clustering, on the other hand, did not put *P. aeruginosa* PA1 and *Pseudomonas aeruginosa* DK2 with other *P. aeruginosa* strains (**Figure 3.11**). Average linkage clustering, which groups the data

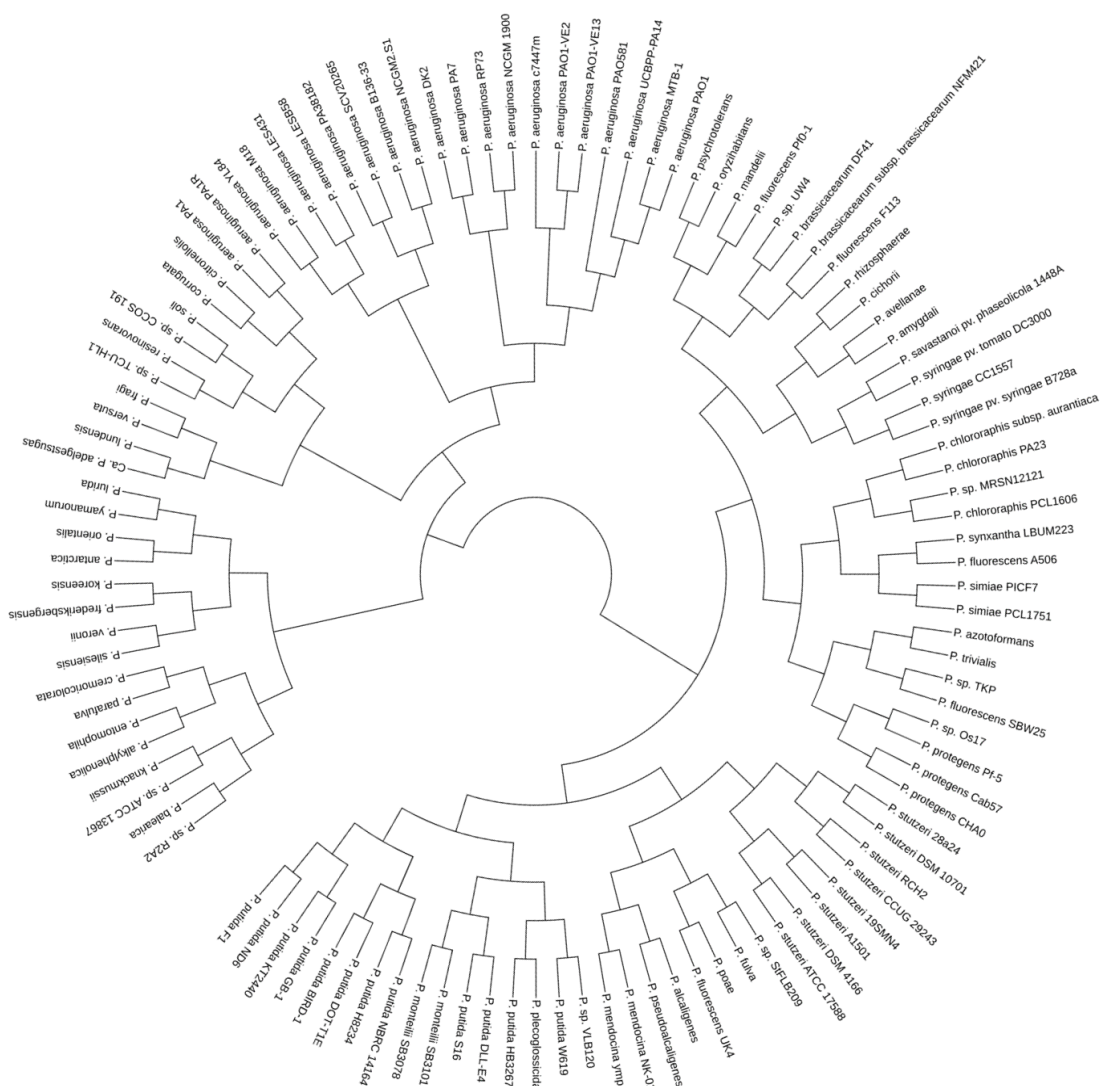


Figure 3.10: Enzyme-based metabolic similarity analysis of *Pseudomonas* using single linkage clustering. The analysis was performed using MAPPS and restricted to pathway maps relating to carbohydrate and amino acid metabolisms listed in **Table 3.2**.

based on average distances between the clusters, put all of the twenty-one strains of *P. aeruginosa* together using both enzyme and reaction data (**Figure 3.12** and **Figure 3.13**). A similar result was obtained when the phylogenetic tree was generated using 16S rRNA sequences (**Figure 3.8**). Like single linkage, maximum linkage could not group all *P. aeruginosa* strains together irrespective of whether enzyme or reaction data was used (**Figure 3.14** and **Figure 3.15**). It only put fifteen strains of *P. aeruginosa* together while clustering remaining seven strains in a cluster containing strains of *P. putida* and *P. monteilii* when using enzyme data (**Figure 3.14**). Similarly, clustering

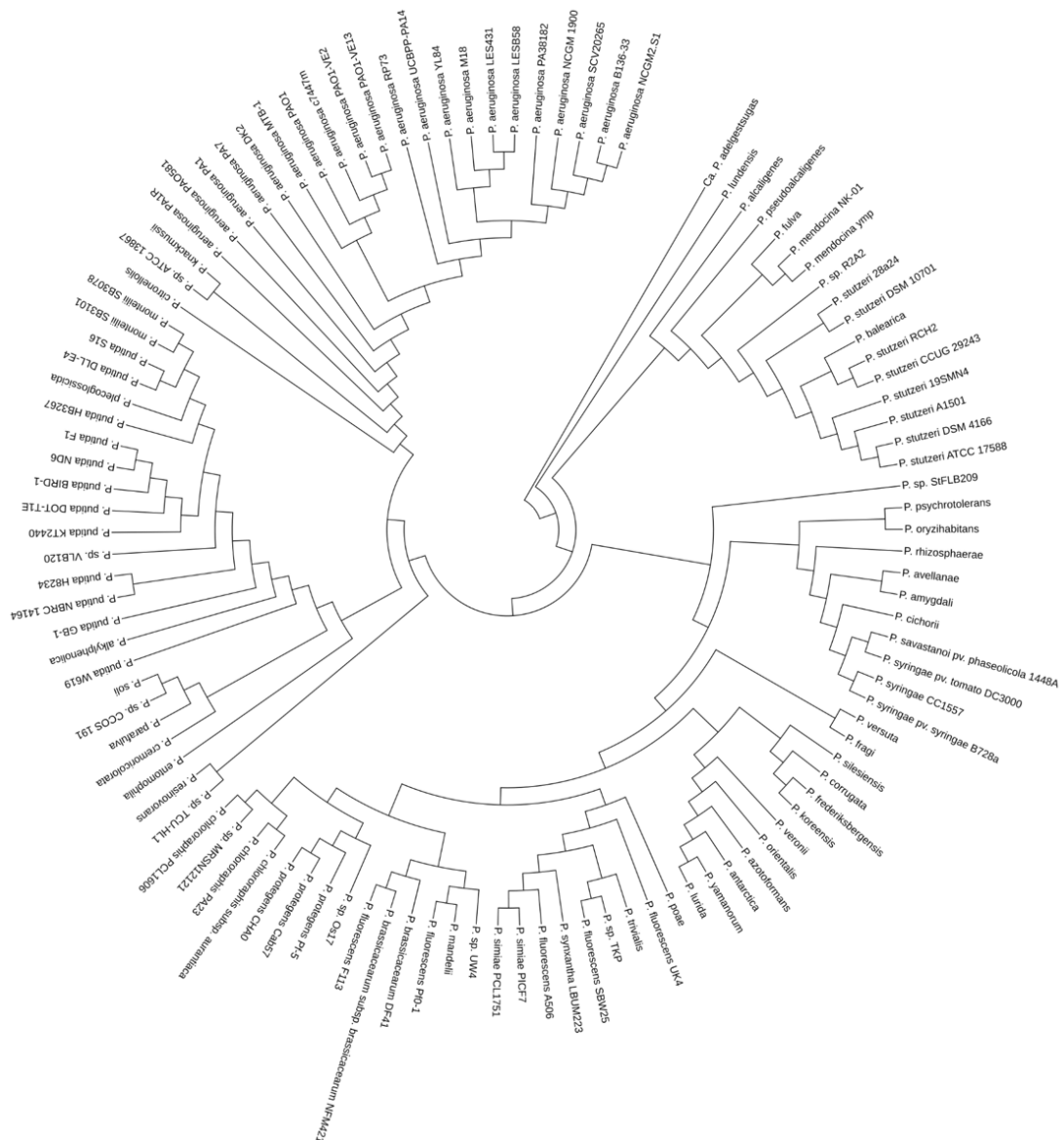


Figure 3.12: Enzyme-based metabolic similarity analysis of *Pseudomonas* using average linkage clustering. The analysis was performed using MAPPS and restricted to pathway maps relating to carbohydrate and amino acid metabolisms listed in **Table 3.2**.

however placed farther apart from *P. putida* group by single linkage (**Figure 3.10** and **Figure 3.11**) and maximum linkage methods (**Figure 3.14** and **Figure 3.15**) but clustered together with *P. putida* using average linkage method (**Figure 3.12** and **Figure 3.13**) as well as 16S and MLSA tree. As noted in **Section 3.4**, *P. entomophila* was found to have a greater evolutionary distance from other *P. putida* strains which might explain the inconsistency in the clustering results.

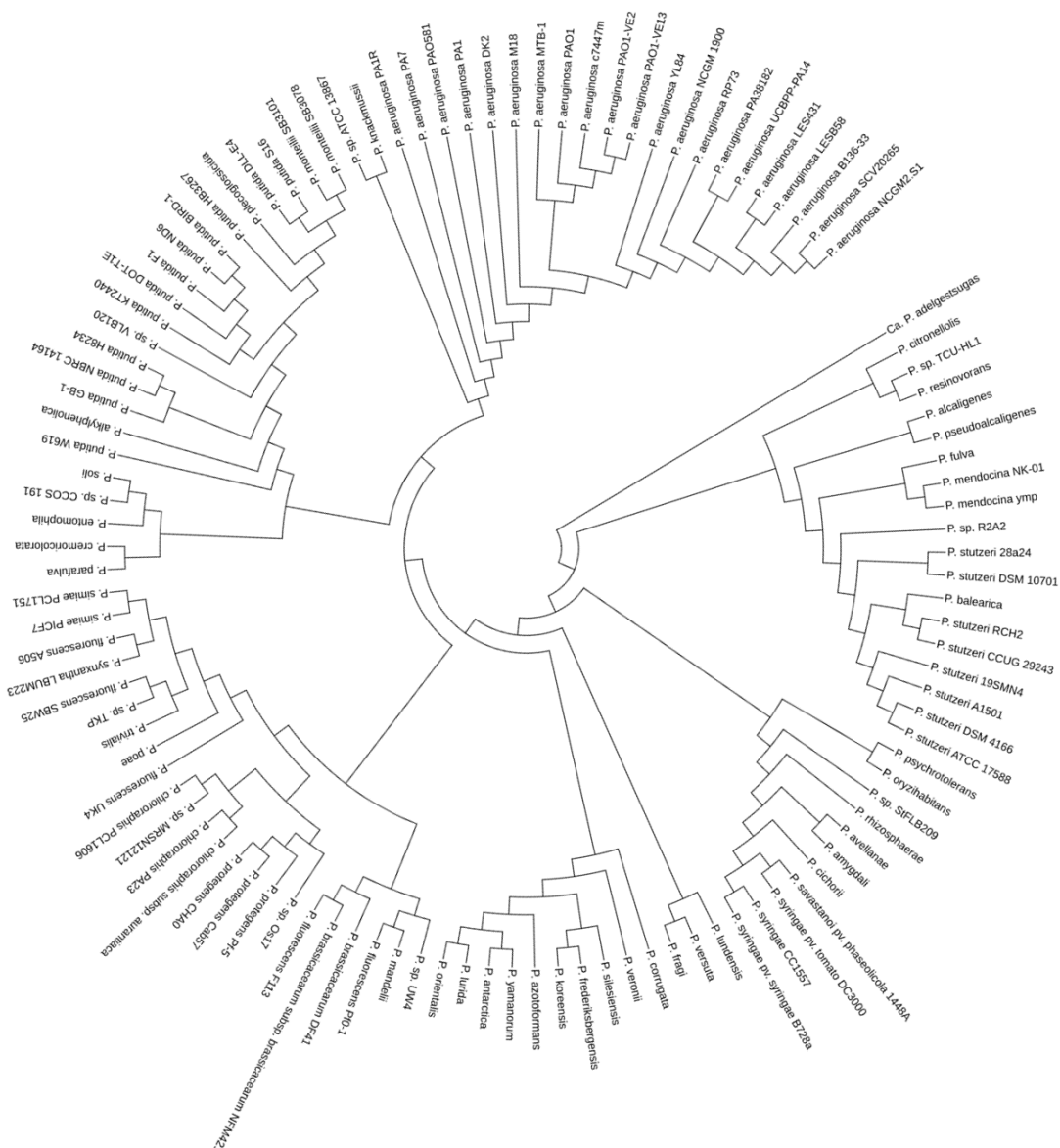


Figure 3.13: Reaction-based metabolic similarity analysis of *Pseudomonas* using average linkage clustering. The analysis was performed using MAPPS and restricted to pathway maps relating to carbohydrate and amino acid metabolisms listed in **Table 3.2**.

All three strains of *P. protegens* were placed together in the enzyme- and reaction-based clustering using single linkage (**Figure 3.10** and **Figure 3.11**), average linkage (**Figure 3.12** and **Figure 3.13**), and maximum linkage (**Figure 3.14** and **Figure 3.15**). Similar results were obtained for the 16S tree (**Figure 3.8**) and the MLSA tree (**Figure 3.9**). Besides this, *P. syringae* pv. tomato DC3000, *P. syringae* pv. syringae B728a, *P. syringae* CC1557 and *P. savastanoi* pv. phaseolicola 1448A were grouped together in all MSA results as well as in the 16S tree (**Figure 3.8**). Moreover, all eight

citronellolis by enzyme-based single linkage method (**Figure 3.10**). Both enzyme and reaction-based maximum-linkage MSA trees placed *P. corrugata* with *P. brassicacearum* DF41 (**Figure 3.14** and **Figure 3.15**). Enzyme- and reaction-based average linkage clustering, on the other hand, placed *P. corrugata* closer to *Pseudomonas frederiksbergensis* (**Figure 3.12** and **Figure 3.13**). All variants of MSA grouped *P. frederiksbergensis* with *P. koreensis*.

Finally, various cold-adapted pseudomonads including *P. antarctica*, *P. lurida*, *P. orientalis* and *Pseudomonas yamanorum* (Selvakumar et al., 2011; Arnau et al., 2015; Lee et al., 2017), which were placed closer to each other by both 16S (**Figure 3.8**) and MLSA trees (**Figure 3.9**), grouped together in all variants of MSA except reaction-based maximum linkage clustering. The reaction-based maximum linkage clustering divided them into two groups (**Figure 3.15**) with *P. antarctica* and *P. yamanorum* placed together in a cluster and *P. lurida* and *P. orientalis* put together in a separate cluster.

3.6 Discussion

Pseudomonas is a very diverse group of bacteria, including pathogenic strains that can cause diseases in humans, animals, and plants. Internal and external stimuli steer metabolic specialization for a specific lifestyle in these bacteria (Preston et al., 1998; Spiers et al., 2000; Remold et al., 2011; Silby et al., 2011). In the past five decades, a large number of new strains have been isolated and characterized by various experimental methods. Advancement of technology has also played a vital role in expediting the discoveries and classification of new strains. Hundreds of genome sequences of *Pseudomonas* strains have become available in the public databases (Clark et al., 2016). This has paved the way to develop specialized databases containing

metabolic, genomic and taxonomic data relating to various pseudomonads (Winsor et al., 2016; Kanehisa et al., 2019).

In this chapter, information about the metabolic networks of *Pseudomonas* available in KEGG was used for studying the metabolic diversity and functional specialization in *Pseudomonas*. Analyses were performed on a subset of metabolism, focusing on carbohydrate, amino acid and energy metabolisms. A comparison of metabolic reactions has shown that some of the KEGG reactions were predicted to be present in only one strain, including insect pathogen *P. entomophila*, *P. putida* H8234, and an extremophile, *P. psychrotolerans*, but absent from the others (see **Table 3.3** and **Table 3.5**). Overall, higher number of metabolic reactions were predicted to be conserved in energy metabolism (14.5%) as compared to carbohydrate and amino acid metabolism (8.6%). A higher variability in the reactions involved in carbohydrate metabolism (see **Section 3.2.1**) is in agreement with the finding that consumptions of sugar in *Pseudomonas* is limited and is highly dependent on their habitat and lifestyle (Udaondo et al., 2018). Including reactions from pathways involved in the metabolism of glycans, lipids and secondary amino acids in the comparison can provide further insights into the metabolic diversity and functional specialization of different *Pseudomonas* strains.

To further gain insight into the metabolic diversity of various pseudomonads, metabolic pathways were predicted from amino acids to ammonia and four TCA Cycle intermediates, oxaloacetate, fumarate, succinyl-CoA, and 2-oxoglutarate for each pseudomonad (see **Section 3.3**). Connections between metabolites were established using KEGG RCLASS when predicting metabolic pathways. This, however, resulted in a few pathways being missed due to the unavailability of reactant pairs to generate connections between intermediary metabolites. For example, MAPPS predicted

metabolic pathways from L-Arginine to ammonia, all of which were five reactions long, in only seven strains, including *P. antarctica*, *P. chlororaphis* PCL1606, *P. entomophila*, *P. fluorescens* SBW25, *P. poae*, *Pseudomonas sp.* CCOS 191 and *Pseudomonas sp.* MRSN12121. This is, however, in contradiction to the reported literature because L-Arginine deiminase (EC 3.5.3.6) can produce ammonia from L-arginine in one step (Lu et al., 2006). Similarly, for some amino acids including Glycine, L-Serine, L-Tryptophan, L-Threonine no pathways of length up to six reactions were found. Longer pathways of length between seven to ten reactions could be computed for these metabolites. The literature, however, suggests presence of shorter pathways for these amino acids (Mithani et al., 2011). These results highlight the limitations of using only main reactant pairs of a reaction for establishing connections between metabolites. Using metabolite pairing, on the other hand, which pairs each source metabolite in a reaction with each target metabolite, tends to produce a large number of false positive pathways (Mithani et al., 2009a).

Furthermore, phylogenetic relationship based on 16S rRNA sequence and four housekeeping genes including *gyrB*, *gltA*, *rpoB*, and *gapA* were generated for the 111 *Pseudomonas* strains (**Section 3.4**), and used as reference to compare the clustering of various pseudomonads generated using reaction and enzyme data via metabolic similarity analysis. The metabolic similarity analysis was performed on the dataset relating to carbohydrate and amino acid metabolisms with three different methods including single linkage, average linkage, and maximum linkage used for joining the clusters. Out of these three methods, analysis performed using average linkage successfully captured the relationship between pseudomonads, including large groups such as *P. aeruginosa*. Interestingly, strains with similar lifestyle, for example those belonging to *P. putida* and *P. stutzeri*, were placed together by all three methods

suggesting the reactions involved in carbohydrate and amino acid metabolisms were able to differentiate the pseudomonads based on their lifestyle. As discussed above, variability in carbohydrate and amino acid metabolism of *Pseudomonas* depicts their diverse metabolic characteristics required for adaptation to different niches. Clustering result shows that metabolic data can be successfully used to identify the relationship between large groups of bacteria. The metabolic similarity analysis results presented in this chapter were restricted to reactions and enzymes involved in carbohydrate and amino acid metabolism. This can be further expanded to include reactions and enzymes involved in energy and secondary metabolism to allow further understanding of metabolic variability and conservation in various *Pseudomonas* strains exhibiting specialized metabolic profiles.

4 Differential Metabolic Analysis of Mango between Immature and Mature Stages

Mangifera indica (Mango) belongs to the genus *Mangifera* of the family Anacardiaceae, and is popularly known as “The king of fruits” (Tharanathan et al., 2006). The mango fruit is climacteric (Lakshminarayana, 1973), with varying thickness and length depending on the cultivar. Mango, by production, is the third largest tropical fruit crop in the world behind bananas and pineapple (Bally and Dillon, 2018), with over 1,000 varieties around the world (Mukherjee, 1953; Litz, 2009). In Pakistan, mango is the second-largest fruit crop after citrus (Nazish et al., 2017). Pakistan is also an important exporter of mangoes with production centered in two regions, Punjab and Sindh (Akhtar et al., 2009).

The development of mango fruit involves cell division, and cell expansion is responsible for the increase in fruit size (Seymour et al., 2013). Growth of soft and edible fruit with desirable quality attributes depends on the biochemical and physiological changes which are associated with highly coordinated, genetically programmed, and an irreversible process of fruit development and ripening (Giovannoni, 2001). The fruit development process involves phytohormone activities (Gillaspy et al., 1993), after fertilization, the auxin upregulates the biosynthesis of gibberellin, which initiate the gibberellin signaling in the ovules, consequently stimulate fruit growth (Seymour et al., 2013). Other phytohormones like ethylene, cytokinin, abscisic acid (ABA), jasmonates, and brassinosteroids also play significant roles in fruit development and ripening, and polyamines are also found to be growth

regulators of fruit quality (Srivastava and Handa, 2005; Handa et al., 2012). Biological pathways drive the morphological changes in the fruit, such as exocarp of mango fruit develops into protective skin of green color; during maturity, the color changes to pale green or yellow. The edible fleshy region of the fruit is mesocarp with a flavor ranging from turpentine to sweet, variation in the flavor is also controlled by the underlying metabolic pathways. During development, the production of different compounds in the mango starts to change, such as chlorophyll starts disappearing, and the number of anthocyanins and carotenoids increases during maturity (Tharanathan et al., 2006). Several studies have explored the transcriptional dynamics of various genes (Pandit et al., 2010), the composition of volatile compounds (Pandit et al., 2009), changes in different compounds like sugars and pectin (Tandon and Kalra, 1983, 1984) and morphological variation (Bally, 1999) happens during the development of mango.

At a genomic level, mango is a diploid species containing 20 pairs of chromosomes ($2x = 40$). With a genome size of approximately 400 Mb, it is almost 3.3 times the size of the model plant *Arabidopsis thaliana* but has a relatively smaller genome compared to other important fruits (banana: 600 Mb, grapes: 500 Mb, apples: 750 Mb) (Feuillet et al., 2011; Ravishankar et al., 2011). Studies on mango leaf transcriptome (Azim et al., 2014) and mango fruit peel transcriptome (Luria et al., 2014) have generated 30,509 and 57,444 contigs, respectively, through *de novo* assembly. Recently, a high-quality assembly of mango was published, which reported a 392 Mb genome of Indian Mango cultivar ‘Alphonso’ containing 41,251 protein-coding genes (Wang et al., 2020).

Recent advances in high-throughput DNA sequencing technologies and associated computational techniques are increasingly enabling genome sequence analysis and comparisons on a genome-wide basis (Heather and Chain, 2016; Stark et

al., 2019). When applied to RNA data, this is referred to as RNA-sequencing or RNA-seq. High-throughput RNA-seq analysis has emerged as an efficient and cost-effective approach for transcriptome profiling due to increases in throughput of next-generation sequencing and third-generation sequencing technologies (Costa-Silva et al., 2017; Stark et al., 2019). In RNA-seq experiments, the primary interest is differential gene expression analysis between case and control groups or at different time courses (Spies and Ciaudo, 2015; Sahraeian et al., 2017) in addition to identifying splice variants and differential alternative splicing (Marco-Puche et al., 2019). RNA-seq based differential gene expression analysis provides a platform to compare the expression levels of two or more groups such as different environmental conditions, physiological conditions, organs/tissues or developmental stages (Boscari et al., 2013; Klepikova et al., 2016; Schiano et al., 2017). For this, RNA-seq data is typically aligned to the reference genome for model organisms or to the transcriptome sequences reconstructed using *de novo* assembly strategies for organisms without reference sequences to quantify gene expressions (Jung et al., 2019) and the number of mapped reads are used to estimate and compare the relative expression level of genes. Subsequently, statistical methods are applied to test the significance of differences in gene expression between groups (Zhang et al., 2014).

When mapped onto metabolic networks, the RNA-seq data can provide valuable information about the metabolic pathways that are active under different conditions. In the last decade, several studies have shown that mapping differentially expressed genes on the metabolic network can give insight to the critical metabolic pathways responsible for the intermediary regulation of a specific phenotype or trait. For example, studies have demonstrated how association of novel genes with secondary metabolic pathways can be established through mapping of RNA-seq data (Shi et al., 2011), how important

regulatory patterns of metabolic pathways are inferred from differential expression data (Haynes et al., 2013), and how genes and metabolic pathways related to known phenotypic and physiological effects are evaluated by using RNA-seq data and metabolic pathways (Glagoleva et al., 2017). Studies have also observed higher level of consistency that exists in the expression of genes associated with similar metabolic pathways compared to randomly selected set of genes (Huang et al., 2006; J. L. Deng et al., 2019) adding functional context to observed gene expression patterns and laying grounds for further exploration at other ‘omics levels.

This chapter explores the regulation of metabolism during mango fruit development by identifying genes differentially expressed between immature and mature stages in two cultivars of mango, ‘Sindhri’ and ‘Kala Chaunsa’, using RNA-seq data and mapping them on the underlying metabolic network. To examine the variation in metabolic pathways during fruit maturity, reference transcripts were mapped to KEGG metabolic pathways using KEGG Automatic Annotation Server (Moriya et al., 2007). Furthermore, differentially expressed genes identified in both cultivars were mapped to KEGG pathways to identify the upregulated and downregulated parts of metabolic networks during fruit maturation.

4.1 RNA-seq data for mango

RNA-seq data available in the lab for two South Asian mango cultivars, ‘Sindhri’ and ‘Kala Chaunsa’, at two developmental stages, immature and mature, was used to study the metabolic changes occurring during mango fruit development. Three replicates per stage per cultivar containing between thirty million and thirty-two million reads with a length of 100bp (**Table 4.1**) generated using Illumina paired-end sequencing technology (McCombie et al., 2019) were used. RNA-seq data was aligned to the

Table 4.1: Alignment mapping summary for ‘Sindhri’ and ‘Kala Chaunsa’ samples

Cultivar	Stage	Sample Id	Total number of reads	Mapped Data		Filtered Data	
				Number of reads	%	Number of reads	%
Kala Chaunsa	Immature	KIR1	32,428,882	31,202,189	96.22	27,476,093	84.73
		KIR3	32,329,452	30,939,513	95.70	26,496,707	81.96
		KIR4	34,121,408	32,908,495	96.45	29,198,535	85.57
	Mature	KMR1	30,905,024	29,696,887	96.09	26,122,889	84.53
		KMR4	33,241,602	32,166,568	96.77	29,606,117	89.06
		KMR5	33,155,404	32,123,988	96.89	29,695,679	89.57
Sindhri	Immature	SIR2	33,344,688	32,163,165	96.46	29,316,825	87.92
		SIR3	33,894,142	32,219,329	95.06	28,730,455	84.77
		SIR5	34,695,624	33,335,851	96.08	29,530,425	85.11
	Mature	SMR2	33,127,308	31,783,591	95.94	28,983,437	87.49
		SMR3	33,470,882	32,269,214	96.41	29,322,827	87.61
		SMR4	31,858,380	30,539,331	95.86	26,767,373	84.02

transcriptomic reference for ‘Kala Chaunsa’ (already available in the lab) using the Burrows Wheeler Aligner (BWA) (Li and Durbin, 2009). The reference was generated in the lab using Trinity package (Grabherr et al., 2011) and contained 67,643 contigs. To avoid the alignment of reads of the same gene at multiple locations on the reference transcriptome, the contigs were filtered for transcript isoforms and the ‘longest isoform’ was selected reducing the contig count to 34,123 contigs. Aligned reads were subsequently filtered to remove reads with low mapping quality (Phred score < 20). The alignment results are summarized in **Table 4.1**.

4.2 Metabolic annotation of mango genes

KEGG Automated Annotation Server (KAAS) (Moriya et al., 2007) was used to assign the metabolic annotation to the transcripts of reference genome. KAAS uses sequence similarity to assign KEGG Orthology (KO) numbers to each transcript. KO numbers represents a group of genes, and it is directly linked to an object in the KEGG pathway map or other biological process (Moriya et al., 2007). KEGG contains metabolic

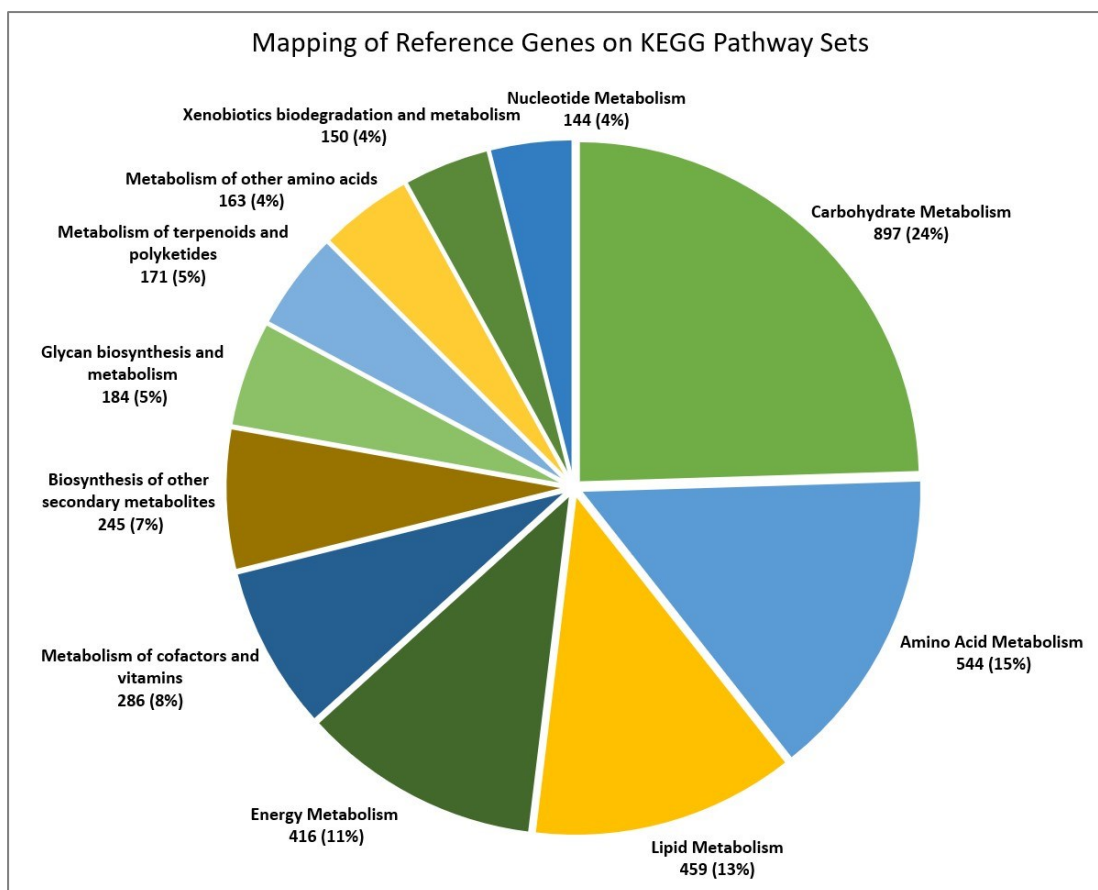


Figure 4.1: Distribution of reference transcripts assigned to different KEGG pathway sets. The transcripts were assigned to different KEGG pathways sets using KAAS.

annotations for several plant genomes. Thirty-five of these genomes (**Appendix K**) including *Citrus sinensis* (sweet orange; KEGG organism code: cit), *Malus domestica* (apple; KEGG organism code: mdm) and *Musa acuminata* (banana; KEGG organism code: mus) were used as representative set to assign KEGG annotations to mango genes. Metabolic enzymes of various KEGG pathways were assigned to transcripts through KO numbers. The annotation results are summarized in **Figure 4.1**. Over 60% of the reference transcripts mapped to carbohydrate, amino acid, lipid and energy metabolisms. Number of transcripts mapped to different KEGG pathway maps belonging to these pathway sets is listed in **Table 4.2**.

Table 4.2: Number of transcripts mapped to carbohydrate, amino acid, lipid and energy metabolisms

KEGG Pathway Set	KEGG Pathway Map	Number of Transcripts Mapped
Amino Acid Metabolism	Alanine, aspartate and glutamate metabolism	46
Amino Acid Metabolism	Arginine and proline metabolism	54
Amino Acid Metabolism	Arginine biosynthesis	34
Amino Acid Metabolism	Cysteine and methionine metabolism	85
Amino Acid Metabolism	Glycine, serine and threonine metabolism	54
Amino Acid Metabolism	Histidine metabolism	22
Amino Acid Metabolism	Lysine biosynthesis	11
Amino Acid Metabolism	Lysine degradation	23
Amino Acid Metabolism	Phenylalanine metabolism	34
Amino Acid Metabolism	Phenylalanine, tyrosine and tryptophan biosynthesis	46
Amino Acid Metabolism	Tryptophan metabolism	37
Amino Acid Metabolism	Tyrosine metabolism	38
Amino Acid Metabolism	Valine, leucine and isoleucine biosynthesis	14
Amino Acid Metabolism	Valine, leucine and isoleucine degradation	46
Carbohydrate Metabolism	5-Branched dibasic acid metabolism	6
Carbohydrate Metabolism	Amino sugar and nucleotide sugar metabolism	107
Carbohydrate Metabolism	Ascorbate and aldarate metabolism	42
Carbohydrate Metabolism	Butanoate metabolism	19
Carbohydrate Metabolism	Citrate cycle	43
Carbohydrate Metabolism	Fructose and mannose metabolism	61
Carbohydrate Metabolism	Galactose metabolism	47
Carbohydrate Metabolism	Glycolysis / Gluconeogenesis	109
Carbohydrate Metabolism	Glyoxylate and dicarboxylate metabolism	59
Carbohydrate Metabolism	Inositol phosphate metabolism	64
Carbohydrate Metabolism	Pentose and glucuronate interconversions	59
Carbohydrate Metabolism	pentose phosphate pathway	45
Carbohydrate Metabolism	Propanoate metabolism	43
Carbohydrate Metabolism	Pyruvate metabolism	83
Carbohydrate Metabolism	Starch and sucrose metabolism	110
Energy Metabolism	Carbon fixation in photosynthetic organisms	65
Energy Metabolism	Methane metabolism	46
Energy Metabolism	Nitrogen metabolism	25
Energy Metabolism	Oxidative phosphorylation	132
Energy Metabolism	Photosynthesis	71
Energy Metabolism	Photosynthesis - antenna proteins	16
Energy Metabolism	Sulfur metabolism	32
Lipid Metabolism	alpha-Linolenic acid metabolism	32
Lipid Metabolism	Arachidonic acid metabolism	17
Lipid Metabolism	Biosynthesis of unsaturated fatty acids	19
Lipid Metabolism	Cutin, suberine and wax biosynthesis	21

Continued from previous page		
Lipid Metabolism	Ether lipid metabolism	18
Lipid Metabolism	Fatty acid biosynthesis	47
Lipid Metabolism	Fatty acid degradation	36
Lipid Metabolism	Fatty acid elongation	27
Lipid Metabolism	Glycerolipid metabolism	75
Lipid Metabolism	Glycerophospholipid metabolism	81
Lipid Metabolism	Linoleic acid metabolism	9
Lipid Metabolism	Sphingolipid metabolism	31
Lipid Metabolism	Steroid biosynthesis	28
Lipid Metabolism	Steroid hormone biosynthesis	12
Lipid Metabolism	Synthesis and degradation of ketone bodies	6

In carbohydrate metabolism, the highest number of transcripts were assigned to amino sugar and nucleotide sugar metabolism, glycolysis, and starch and sucrose metabolism. Metabolic enzymes included glucose-6-phosphate isomerase (EC 5.3.1.9), 6-phosphofructokinase 1 (EC 2.7.1.11), pyruvate kinase (EC 2.7.1.40) and glyceraldehyde 3-phosphate dehydrogenase (EC 1.2.1.12) which play important role in glycolysis. Several other enzymes of various pathways of carbohydrate metabolism were also identified including ATP citrate (pro-S)-lyase (EC 2.3.3.8), isocitrate dehydrogenase (EC 1.1.1.42), malate dehydrogenase (EC 1.1.1.37) and citrate synthase (EC 2.3.3.1) which are part of citrate cycle. Some of these enzymes, which are differentially expressed in ‘Sindhri’ and ‘Kala Chaunsa’ during fruit maturation are discussed in **Section 4.4**.

In addition to their involvement in different metabolic pathways, reference transcripts were also mapped to BRITe protein families involved in metabolic process using KAAS. Results show that majority of the transcripts were mapped to protein kinases, peptidases and phosphatases (**Figure 4.2**). In addition to their role in metabolic pathways, protein kinases also play an important role in regulating other cellular pathways by phosphorylating other enzymes (Krebs, 1972). Similarly, peptidases have been reported to provide defense against pathogen invasion and hydrolysis of the

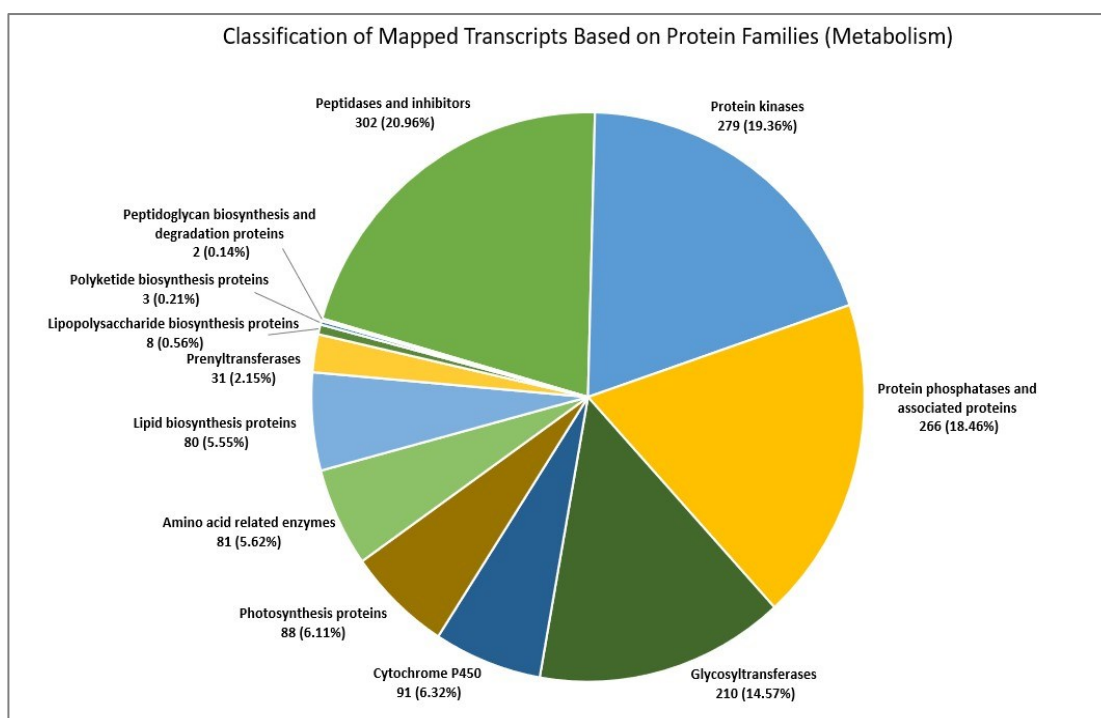


Figure 4.2: Mapping of reference transcripts classified according to BRITE hierarchy of protein families. The transcripts were assigned to protein families using KAAS.

storage proteins (Bonner, 2020), and phosphatases operate in a coordinated manner with the protein kinases in determining the cellular response to a physiological stimulus (Schweighofer and Meskiene, 2015).

4.3 Identification of genes differential expressed during fruit maturation

Differential expression analysis was performed for two mango cultivars including ‘Sindhri’ and ‘Kala Chaunsa’ aligned against the ‘Kala Chaunsa’ transcriptomic reference containing 34,123 genes (see **Section 4.1**) to identify genes that are differentially expressed between immature and mature stages. First, the number of reads aligned against each genes were counted for each sample using HTSeq (Anders et al., 2015). These read counts were subsequently analyzed using two R packages, DESeq2 (Love et al., 2014) and EdgeR (Robinson et al., 2009) to identify significant

changes in the gene expression levels between immature and mature stages of ‘Sindhri’ and ‘Kala Chaunsa’ fruit.

DESeq2 predicts differential expression by the use of negative binomial generalized linear models (Love et al., 2014). To identify differentially expressed genes between immature and mature stages, a data matrix containing the read counts was developed for the samples along with their replicates. Columns in this data matrix corresponded to the sample Ids and the rows corresponded to the transcripts. Sample Ids were further grouped as Mature and Immature. After the design of the matrix, normalization and variance estimation was performed followed by generation of the list of significantly differentially expressed genes between immature and mature samples. The \log_2 fold change (\log_2FC) values, which are proportional to the extent of differential expression of a gene, were used to identify the differentially expressed genes. Cut-off values of 1 and -1 for \log_2FC were used to identify up and down regulated genes. EdgeR also takes count matrices as an input and filters out the genes that are expressed at a very low level (Robinson et al., 2009). In addition to using negative binomial distribution like DESeq2, it also employs empirical Bayes estimation by borrowing information from different genes to shrink the dispersions towards a consensus value. Significant changes in the expression of 4,493 and 4,173 genes of ‘Sindhri’ were identified between immature and mature stages using DESeq2 and EdgeR, respectively (**Figure 4.3**). In the case of ‘Kala Chaunsa’, 491 and 438 genes were found to be differentially expressed by DESeq2 and EdgeR, respectively (**Figure 4.3**). Genes predicted to be differentially expressed by both the methods, 3,722 in ‘Sindhri’ and 294 in ‘Kala Chaunsa’ (**Figure 4.3**), were used for subsequent metabolic pathway analyses.

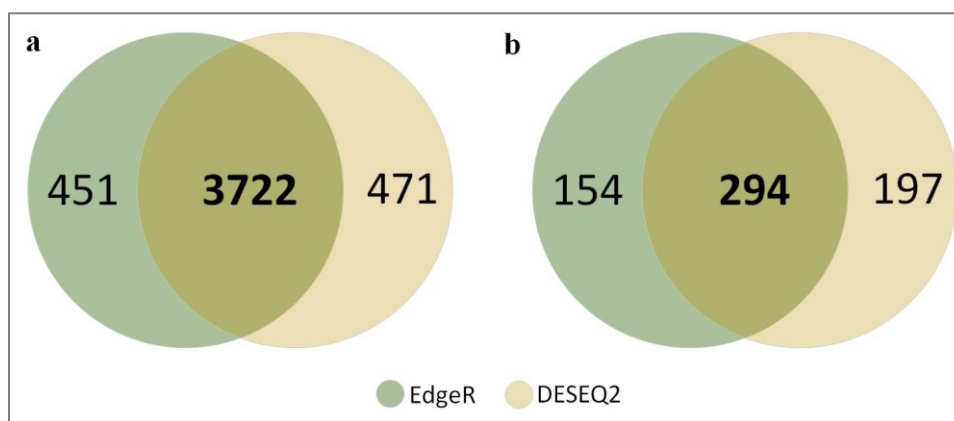


Figure 4.3: Number of differentially expressed genes identified during mango fruit development. (a) Number of differentially expressed genes identified in 'Sindhri' (b) Number of differentially expressed genes identified in 'Kala Chaunsa'.

4.4 Assignment of genes differentially expressed during ripening on KEGG pathways

Metabolic annotations of mango obtained using KAAS (Section 4.2) were further processed using custom scripts to assign metabolic enzymes and pathways to the common differentially expressed genes identified by DESEQ2 and EdgeR for both cultivars i.e., 'Sindhri' and 'Kala Chaunsa'. Differentially expressed genes mapped on to KEGG pathways for each of these cultivars are discussed in the subsequent sections.

4.4.1 Metabolic mapping of differentially expressed genes in 'Kala Chaunsa'

To identify metabolic pathways involved in the development of 'Kala Chaunsa' fruit, differentially expressed genes identified in this cultivar were mapped to KEGG pathway maps. The results are summarized in **Figure 4.4**. Most of the differentially expressed genes were found to be involved in carbohydrate, lipid, energy, and terpenoids and polyketides metabolisms. A previous study has also shown the upregulation of carbohydrate metabolism during development and ripening in mango

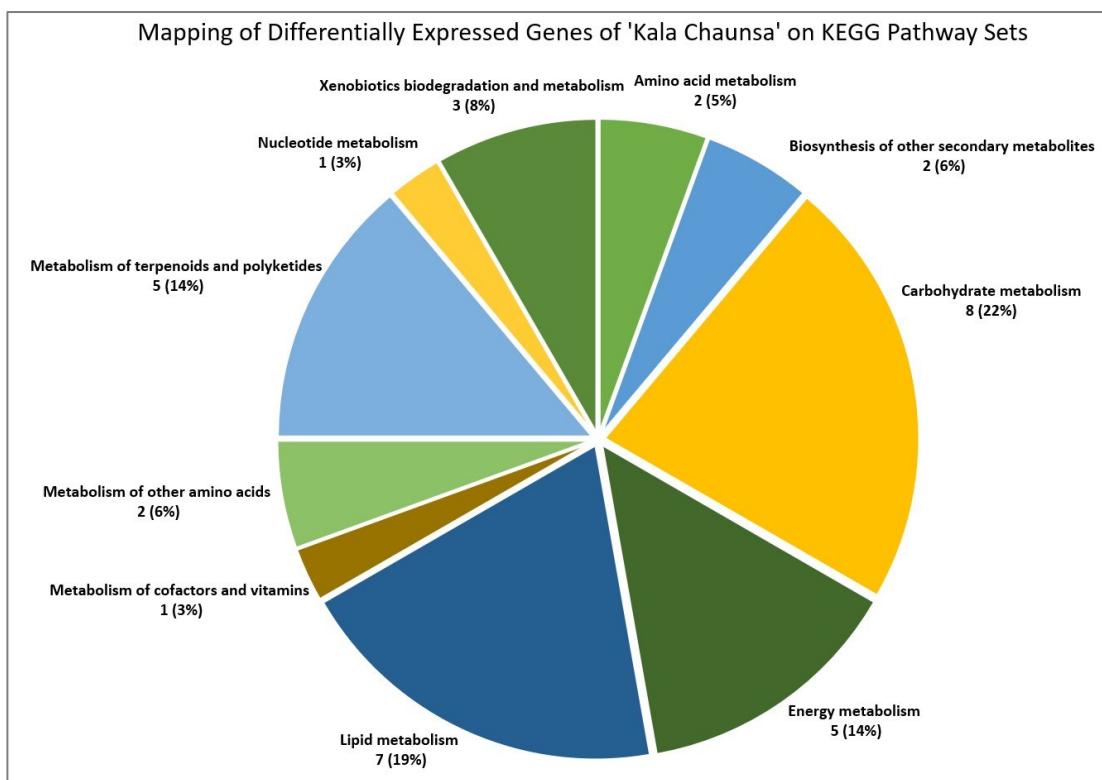


Figure 4.4: Distribution of differentially expressed genes of ‘Kala Chaunsa’ assigned to different KEGG pathway sets. Common genes predicted as differentially expressed by DEseq2 as well as EdgeR were used for annotation.

fruits and mutant sweet orange (Wu et al., 2014). The number of upregulated and downregulated genes assigned to different KEGG pathways are shown in **Figure 4.5**.

Out of 22 distinct differentially expressed genes assigned to KEGG pathways, 9 were upregulated (**Table 4.3**) and 13 were downregulated (**Table 4.4**) at the mature stage compared to immature stage. One of the upregulated enzymes, pyruvate kinase (EC 2.7.1.40), which produces phosphoenolpyruvate was found to be upregulated at mature stage in agreement with previous findings that it is activated during mango fruit maturation (Litz, 2009) and is part of KEGG maps of glycolysis (**Figure 4.6**) and pyruvate metabolism (**Figure 4.7**). Similarly, enzyme glutathione-S-transferase (EC 2.5.1.18), involved in glutathione metabolism, was also predicted to be upregulated at the mature stage (**Figure 4.8**). Studies have shown its positive role in plant growth and development (Gong et al., 2005; Moons, 2005). Overexpression of trehalose-6-

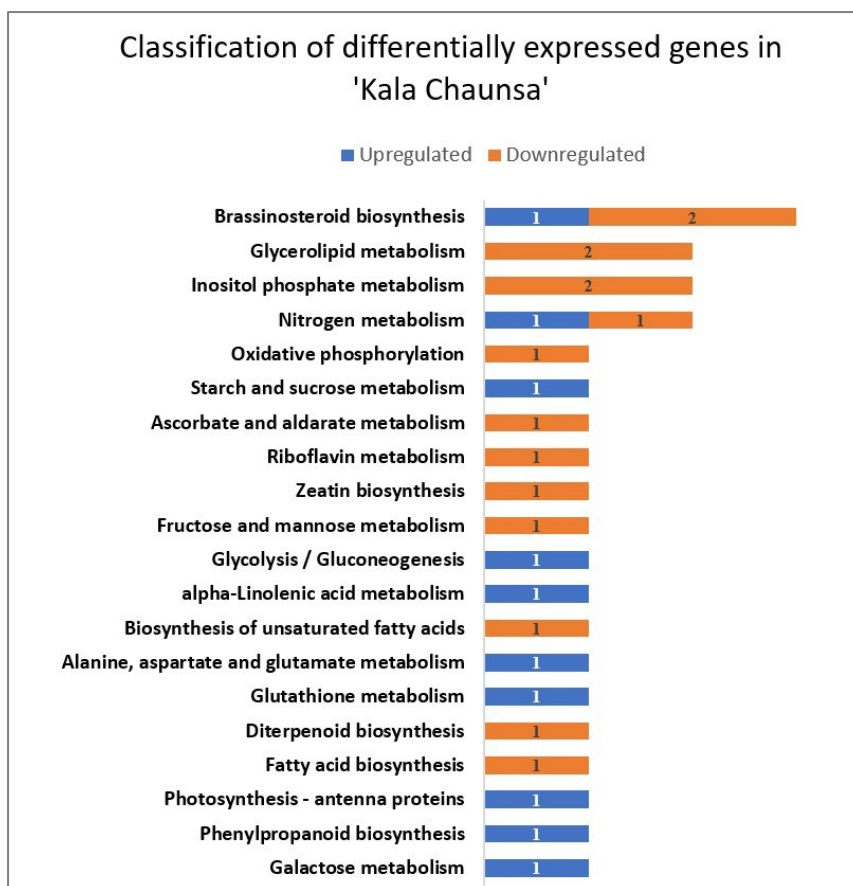


Figure 4.5: Number of genes differentially expressed during fruit maturation in various KEGG pathway maps in ‘Kala Chaunsa’.

phosphate phosphatase (EC 3.1.3.12), which is an indicator of stress tolerance induced by the maturation phase (Lin et al., 2019) was also noted (**Figure 4.9**). Besides this, the enzyme glutamate dehydrogenase (EC 1.4.1.3), which converts L-Glutamate to Oxoglutarate and Ammonia was also found to upregulated in ‘Kala Chaunsa’. The enzyme is involved in KEGG pathway maps of arginine biosynthesis (**Figure 4.10**) and nitrogen metabolism (**Figure 4.11**). On the other hand, carbonic anhydrase (EC 4.2.1.1), which is involved in photosynthesis and the supply of CO₂ to other enzymes in higher plants (Coleman, 2000; Bhat et al., 2017), was predicted to be downregulated in ‘Kala Chaunsa’ during fruit maturation (**Figure 4.10**). Brassinosteroid-6-oxidase (EC 1.14.-.-), an enzyme involved in the synthesis of growth promoting hormone brassinolide via catasterone biosynthesis pathway (**Figure 4.12**) which is involved in the ripening of mango fruit (Zaharah et al., 2012) was also found to be upregulated but

Table 4.3: Metabolic mapping of upregulated genes in ‘Kala Chaunsa’

KEGG Orthology Number	Gene(s)	Enzyme(s)	KEGG pathway map
K00873	pyk	pyruvate kinase (EC 2.7.1.40)	Glycolysis / Gluconeogenesis Pyruvate metabolism Alanine, aspartate and glutamate metabolism
K00261	GLUD1_2 gdhA	glutamate dehydrogenase (NAD(P)+) (EC 1.4.1.3)	Arginine biosynthesis D-Glutamine and D-glutamate metabolism Nitrogen metabolism
K06617	RFS	raffinose synthase (EC 2.4.1.82)	Galactose metabolism
K01087	otsB	trehalose 6-phosphate phosphatase (EC 3.1.3.12)	Starch and sucrose metabolism
K00799	GST	glutathione S-transferase (EC 2.5.1.18)	Glutathione metabolism Metabolism of xenobiotics by cytochrome P450
K09590	CYP85A1 BR6OX1	brassinosteroid-6-oxidase 1 (EC 1.14.-.-)	Brassinosteroid biosynthesis
K08910	LHCA4	light-harvesting complex I chlorophyll a/b binding protein 4	Photosynthesis - antenna proteins
K10529	DOX	alpha-dioxygenase (EC 1.14.99.-)	alpha-Linolenic acid metabolism
K09755	CYP84A F5H	ferulate-5-hydroxylase (EC 1.14.-.-)	Phenylpropanoid biosynthesis

typhasterol/6-deoxytyphasterol 2-alpha-hydroxylase (CYP92A6), another enzyme of the same pathway, was predicted to be downregulated in ‘Kala Chaunsa’ during maturation. This enzyme was, however, found to be upregulated in the ‘Sindhri’ cultivar (see **Section 4.4.2**) Similarly, ferulate-5-hydroxylase (EC 1.14.-.-) involved in phenylpropanoid biosynthesis was predicted to be downregulated in ‘Kala Chaunsa’ but upregulated in ‘Sindhri’ (**Figure 4.13**) at the mature stage compared to immature stage. These variations suggest cultivar specific differences in metabolic pathways relating to fruit ripening.

4.4.2 Metabolic mapping of differentially expressed genes in ‘Sindhri’

To identify metabolic pathways involved in the development of ‘Sindhri’ fruit, differentially expressed genes identified in this cultivar were mapped to KEGG

Table 4.4: Metabolic mapping of downregulated genes in ‘Kala Chaunsa’

KEGG Orthology Number	Gene(s)	Enzyme(s)	KEGG pathway map
K22849	DGAT3	diacylglycerol O-acyltransferase 3, plant (EC 2.3.1.20)	Glycerolipid metabolism
K10703	HACD PHS1 PAS2	very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase (EC 4.2.1.134)	Biosynthesis of unsaturated fatty acids Fatty acid elongation
K14379	ACP5	tartrate-resistant acid phosphatase type 5 (EC 3.1.3.2)	Riboflavin metabolism
K00469	MIOX	inositol oxygenase (EC 1.13.99.1)	Ascorbate and aldarate metabolism Inositol phosphate metabolism
K13496	UGT73C	UDP-glucosyltransferase 73C (EC 2.4.1.-)	Zeatin biosynthesis
K13508	GPAT	glycerol-3-phosphate acyltransferase (EC 2.3.1.15 2.3.1.198)	Glycerolipid metabolism Glycerophospholipid metabolism
K10781	FATB	fatty acyl-ACP thioesterase B (EC 3.1.2.14 3.1.2.21)	Fatty acid biosynthesis
K01858	INO1 ISYNA1	myo-inositol-1-phosphate synthase (EC 5.5.1.4)	Inositol phosphate metabolism
K20623	CYP92A6	typhasterol/6-deoxotyphasterol 2alpha-hydroxylase	Brassinosteroid biosynthesis
K04122	GA3 CYP701	ent-kaurene oxidase (EC 1.14.14.86)	Diterpenoid biosynthesis
K19355	MAN	mannan endo-1,4-beta-mannosidase (EC 3.2.1.78)	Fructose and mannose metabolism
K01674	cah	carbonic anhydrase (EC 4.2.1.1)	Nitrogen metabolism
K03935	NDUFS2	NADH dehydrogenase (ubiquinone) Fe-S protein 2 (EC 7.1.1.2 1.6.99.3)	Oxidative phosphorylation

pathway maps. The results are summarized in **Figure 4.14**. As observed in ‘Kala Chaunsa’, higher number of differentially expressed genes of ‘Sindhri’ were involved in carbohydrate, amino acid and lipid metabolism (Wu et al., 2014). Number of upregulated and downregulated genes assigned to various KEGG pathway maps are shown in **Figure 4.15**.

Out of 193 distinct differentially expressed genes assigned to KEGG pathways, 111 genes were found to be upregulated (**Appendix L**) while 84 genes were found to be downregulated (**Appendix M**) at mature stage compared to immature stage. Based on their log₂FC value, top twenty upregulated and downregulated gene mappings are listed in **Table 4.5** and **Table 4.6** respectively. Four enzymes involved in KEGG glycolysis map were found to be upregulated in ‘Sindhri’ during fruit maturation (**Figure 4.6**) including pyruvate dehydrogenase (EC 2.7.1.40) which was found to be upregulated in ‘Kala Chaunsa’ also (see **Section 4.4.1**). Enzyme aldehyde

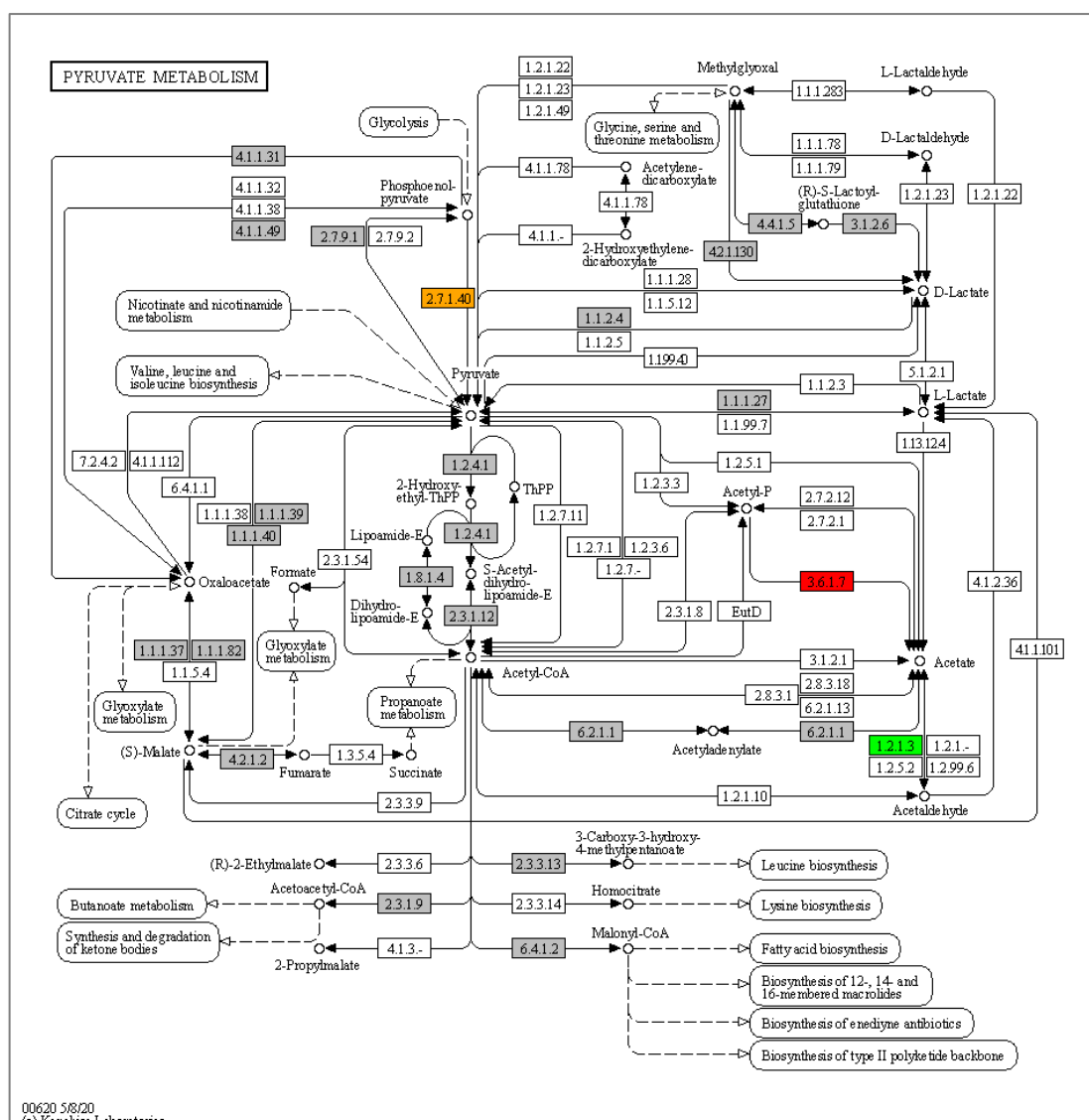


Figure 4.7: Metabolic mapping of genes differentially expressed during fruit maturation on pyruvate metabolism. Metabolic annotations of reference transcripts are highlighted in colored boxes. Enzyme upregulated in both ‘Sindhri’ and ‘Kala Chaunsa’ cultivars during fruit maturation is shown in orange color, enzyme predicted to be upregulated during maturation in ‘Sindhri’ is highlighted in green color and enzymes downregulated in ‘Sindhri’ during maturation are highlighted in red color. Enzymes not found to be differentially expressed during fruit maturation are shown in grey color.

compared to immature stage in ‘Sindhri’. Similarly, enzyme pyruvate decarboxylase (EC 4.1.1.1) listed in **Appendix L**, which is involved in the conversion of pyruvate to downstream metabolites was also predicted to be upregulated at the mature stage in ‘Sindhri’ (**Figure 4.6**). Glutamine synthetase (EC 6.3.1.2) involved in KEGG maps of arginine biosynthesis (**Figure 4.10**) and nitrogen metabolism (**Figure 4.11**), which incorporates ammonia to L-Glutamine (Scarpeci et al., 2007) was found to be upregulated in ‘Sindhri’ at the mature stage.

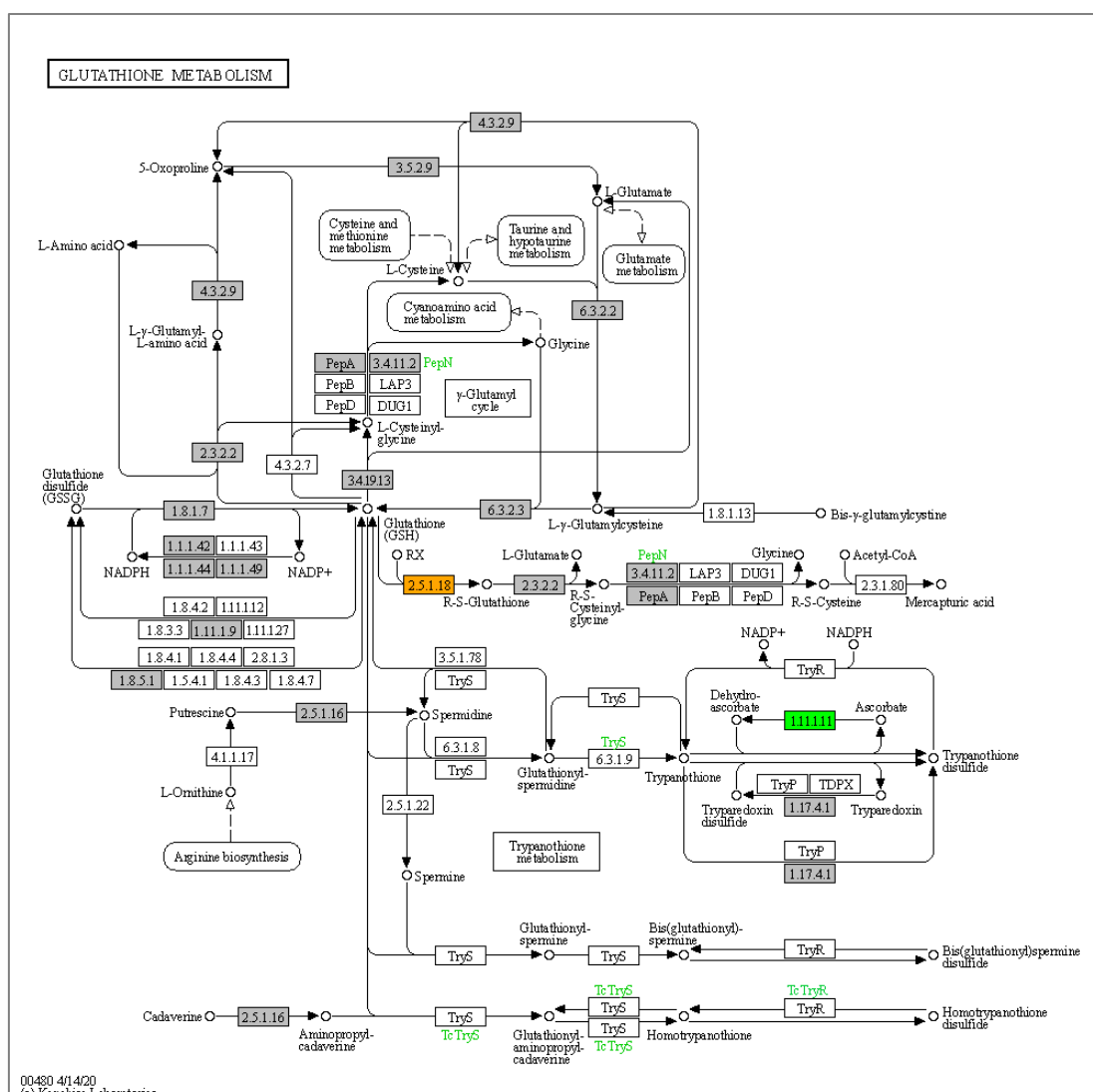


Figure 4.8: Metabolic mapping of genes differentially expressed during fruit ripening on glutathione metabolism. Metabolic annotations of reference transcripts are highlighted in colored boxes. Enzyme upregulated in both ‘Sindhri’ and ‘Kala Chaunsa’ cultivars during fruit ripening is shown in orange color and enzyme predicted to be upregulated during ripening in ‘Sindhri’ is highlighted in green color. Enzymes not found to be differentially expressed during fruit maturation are shown in grey color.

While no enzyme involved in glycolysis was predicted to be downregulated during maturation in ‘Kala Chaunsa’, aldose 1-epimerase (EC 5.1.3.3) was found to be downregulated in ‘Sindhri’ (**Figure 4.6**). Acetyl-CoA carboxylase (EC 6.4.1.2), which is part of pyruvate metabolism (**Figure 4.7**) and is responsible for the conversion of acetyl-CoA to malonyl-CoA which is further linked to fatty acid biosynthesis in plants (Sasaki and Nagano, 2004) was also found to be downregulated in ‘Sindhri’ during maturation.

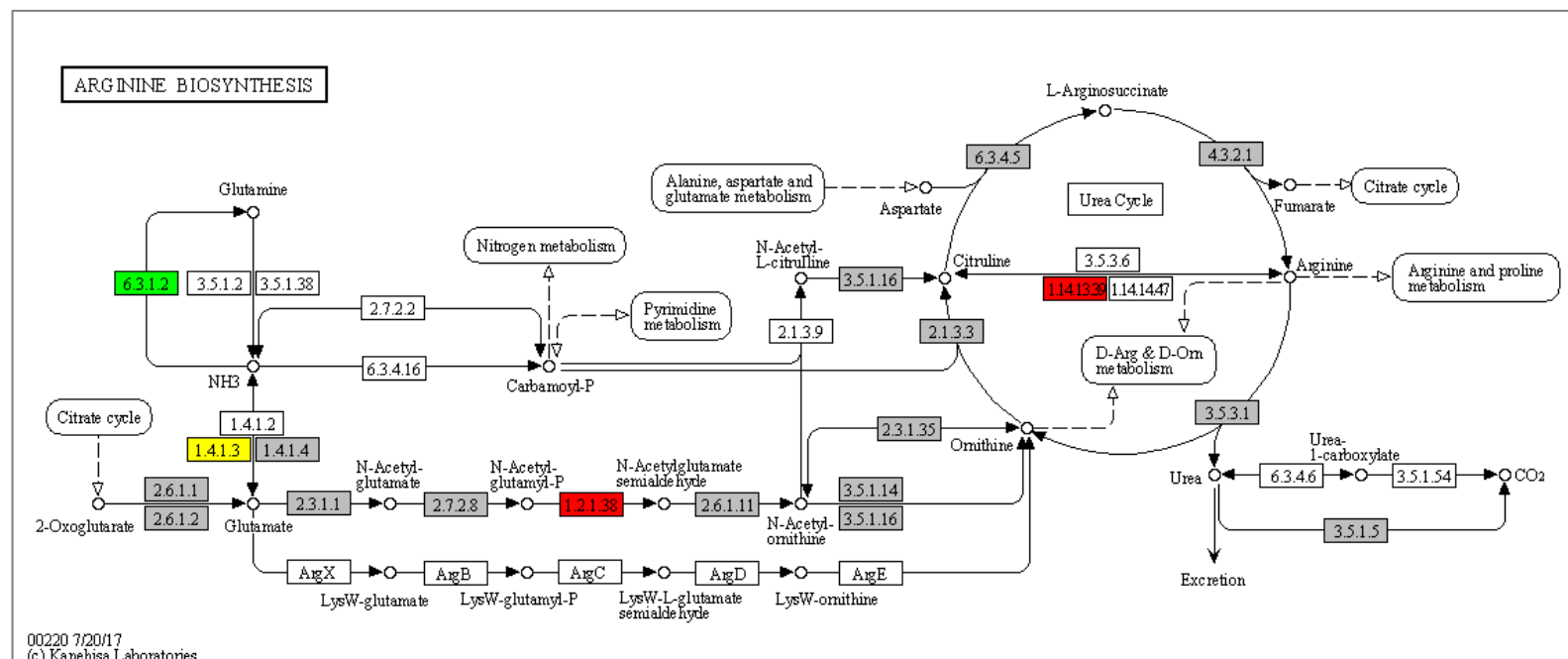


Figure 4.10: Metabolic mapping of genes differentially expressed during fruit maturation on arginine biosynthesis. Metabolic annotations of reference transcripts are highlighted in colored boxes. Enzyme upregulated in 'Sindhri' during fruit maturation is shown in green color, enzymes downregulated in 'Sindhri' during maturation are highlighted in red color and enzyme upregulated in 'Kala Chaunsa' is highlighted in yellow color. Enzymes not found to be differentially expressed during fruit maturation are shown in grey color.

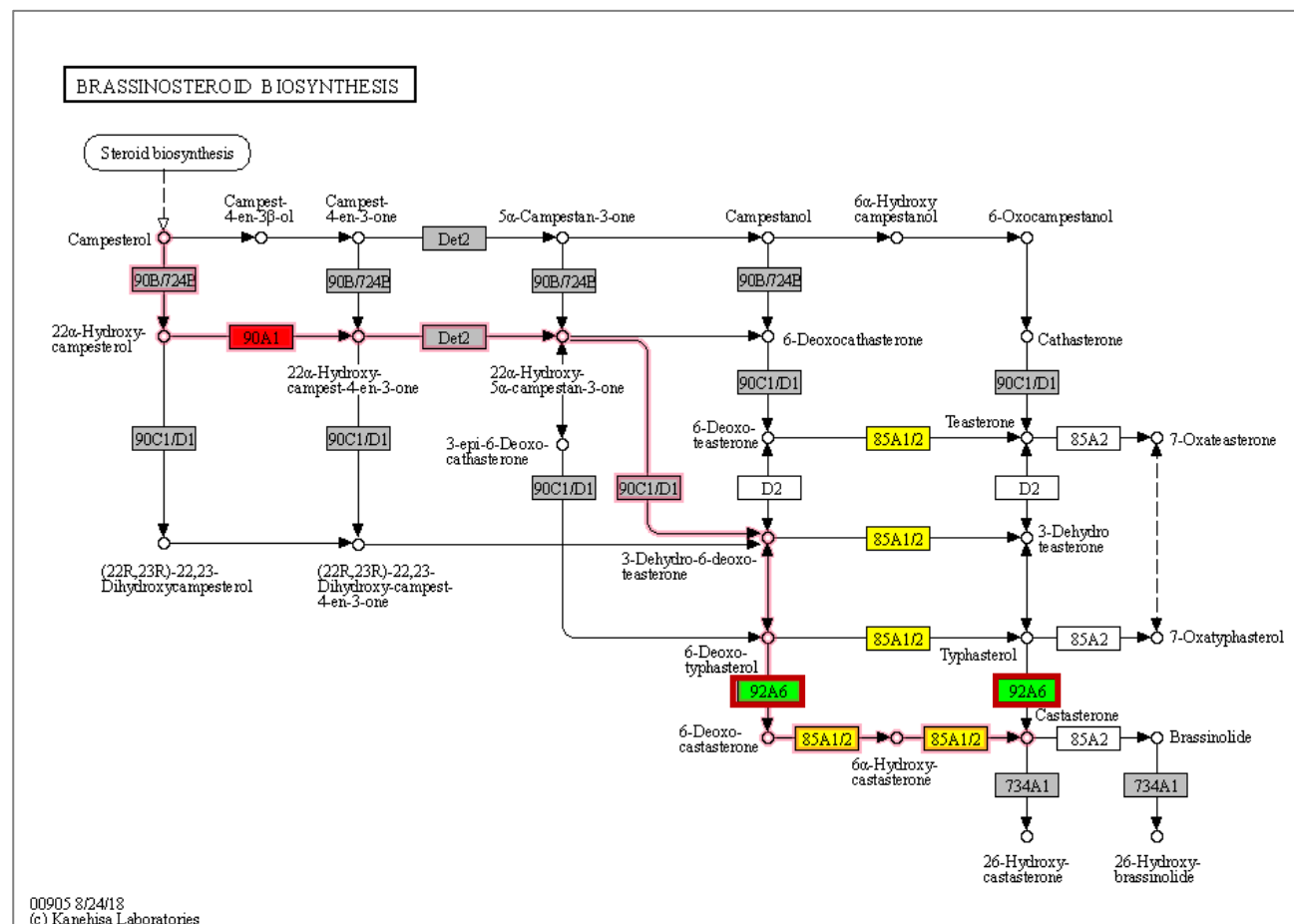


Figure 4.12: Metabolic mapping of genes differentially expressed during fruit maturation on brassinosteroid biosynthesis. Metabolic annotations of reference transcripts are highlighted in colored boxes. color. Enzyme upregulated in ‘Sindhri’ are highlighted in green color, enzymes downregulated in ‘Sindhri’ during maturation are highlighted in red color and enzyme predicted to be downregulated in ‘Kala Chaunsa’ during maturation are highlighted in yellow color. Solid border of red color represents the enzyme predicted to be upregulated in ‘Sindhri’ during maturation but downregulated in ‘Kala Chaunsa’. KEGG pathway module of plant terpenoid biosynthesis is highlighted with solid border of pink color. Enzymes not found to be differentially expressed during fruit maturation are shown in grey color.

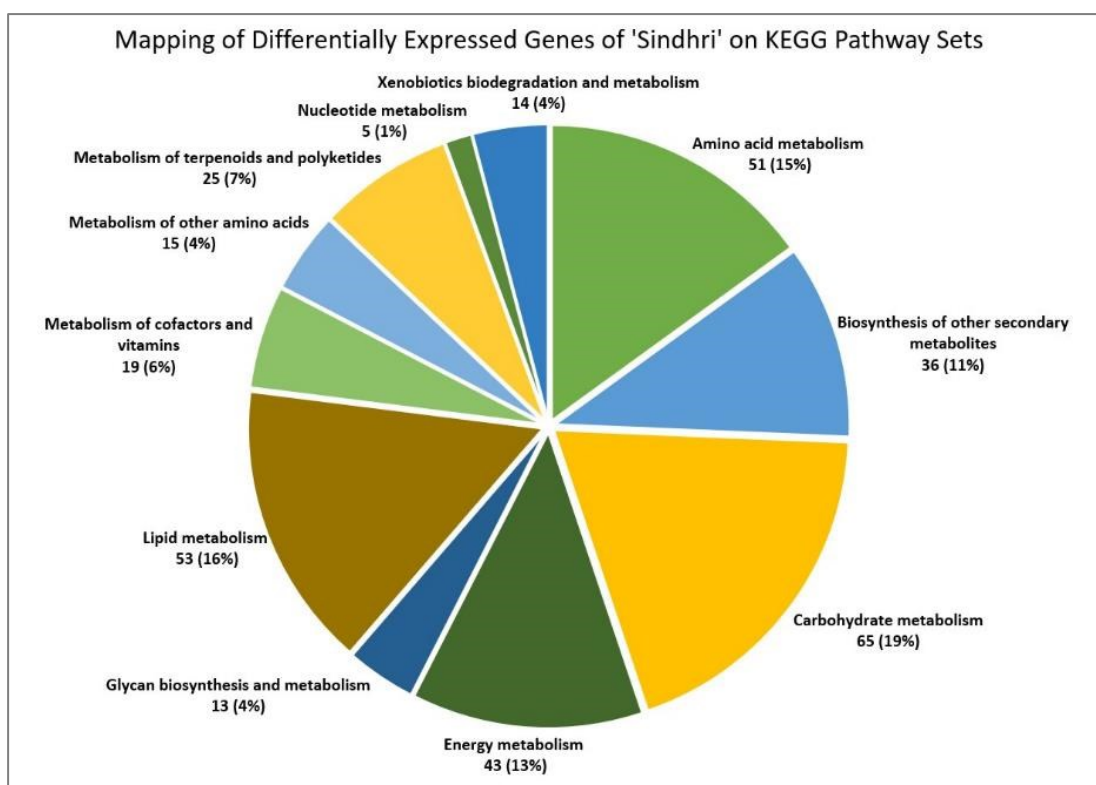


Figure 4.14: Distribution of differentially expressed genes of ‘Sindhri’ assigned to KEGG pathway sets. Common genes predicted by DEseq2 and EdgeR were used for assigning KEGG pathways.

Brassinosteroids actively participate in fruit development and ripening (Zaharah et al., 2012; Nolan et al., 2020). Different enzymes involved in brassinolide biosynthesis via catalsterone pathway were predicted to be differentially expressed during maturation in ‘Sindhri’ (**Figure 4.12**). For example, ferulate-5-hydroxylase (EC 1.14.-.-) involved in phenylpropanoid biosynthesis was predicted to be downregulated in ‘Sindhri’ at the mature stage (**Figure 4.13**). As noted above, this enzyme was found to be upregulated in ‘Kala Chaunsa’ at the mature stage compared to immature stage (see **Section 4.4.1**). Moreover, various enzymes involved in the conversion of phenylalanine to 4-coumaryl-CoA through flavonoid biosynthesis pathway (Nabavi et al., 2020), which is also part of KEGG map of phenylpropanoid biosynthesis (**Figure 4.13**) were found to be differentially expressed during fruit maturation in ‘Sindhri’ with phenylalanine ammonia-lyase (EC 4.3.1.24) being upregulated and 4-coumarate-CoA ligase (EC 6.2.1.12) being downregulated at the mature stage.

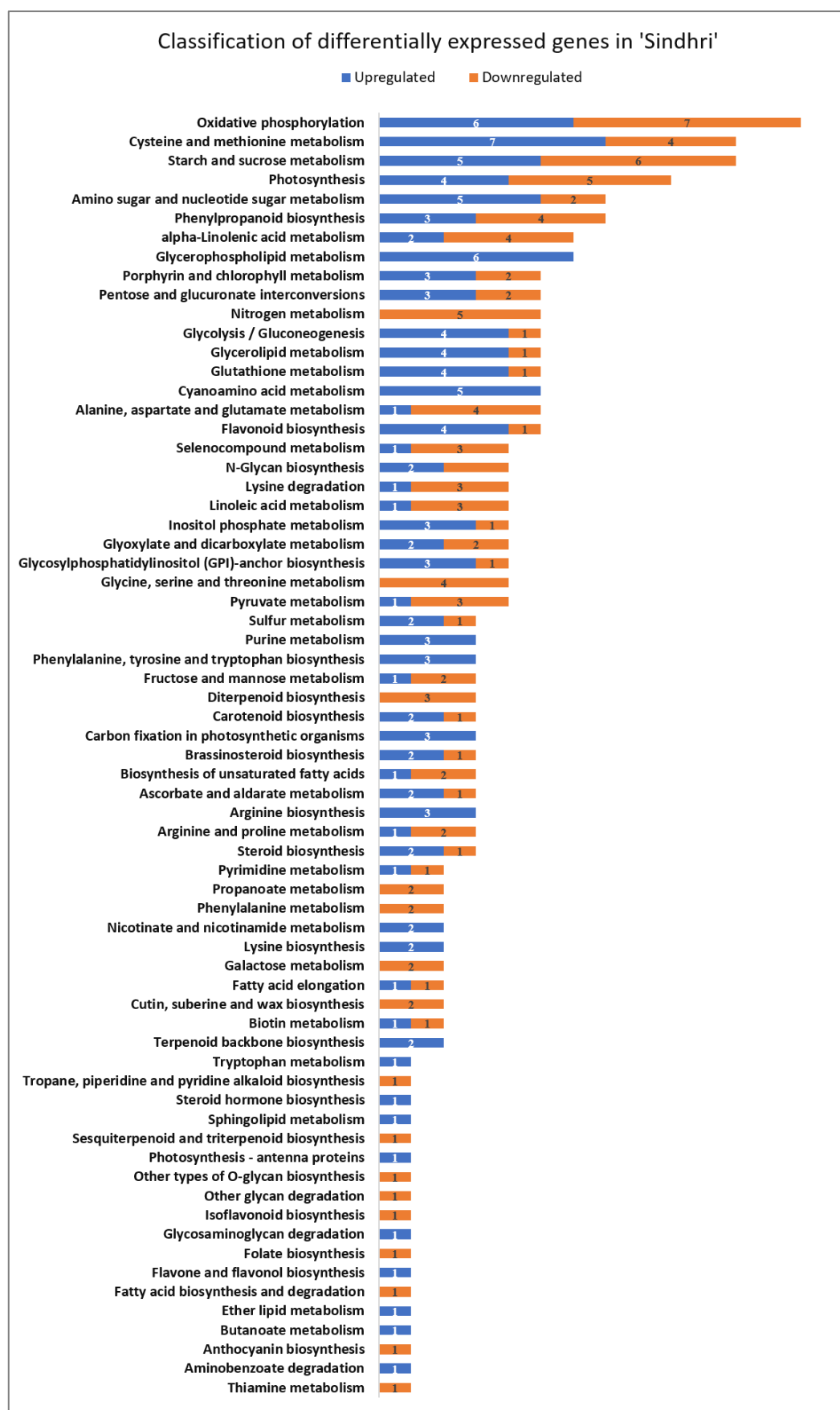


Figure 4.15: Number of genes differentially expressed during fruit maturation in various KEGG pathway maps in 'Sindhri'.

Table 4.5: Metabolic mapping of top 20 upregulated genes in ‘Sindhri’

KEGG Orthology	Gene(s)	Enzyme	KEGG Pathway Map
K00128	ALDH	aldehyde dehydrogenase (NAD ⁺) (EC 1.2.1.3)	Arginine and proline metabolism, beta-Alanine metabolism Tryptophan metabolism Pyruvate metabolism Lysine degradation Histidine metabolism Fatty acid degradation Glycerolipid metabolism Glycolysis / Gluconeogenesis
K00318	PRODH fadM putB	proline dehydrogenase (EC 1.5.5.2)	Arginine and proline metabolism
K00789	metK	S-adenosylmethionine synthetase (EC 2.5.1.6)	Cysteine and methionine metabolism
K00799	GST	glutathione S-transferase (EC 2.5.1.18)	Glutathione metabolism Metabolism of xenobiotics by cytochrome P450
K00873	pyk	pyruvate kinase (EC 2.7.1.40)	Glycolysis / Gluconeogenesis Pyruvate metabolism Purine metabolism
K00913	ITPK1	inositol-tetrakisphosphate 1-kinase (EC 2.7.1.159 2.7.1.134)	Inositol phosphate metabolism
K01087	otsB	trehalose 6-phosphate phosphatase (EC 3.1.3.12)	Starch and sucrose metabolism
K01535	PMA1 PMA2	H ⁺ -transporting ATPase (EC 7.1.2.1)	Oxidative phosphorylation
K01723	AOS	hydroperoxide dehydratase (EC 4.2.1.92)	alpha-Linolenic acid metabolism
K03921	FAB2 SSI2 desA1	acyl-(acyl-carrier-protein) desaturase (EC 1.14.19.2 1.14.19.11 1.14.19.26)	Biosynthesis of unsaturated fatty acids Fatty acid biosynthesis
K04125	GA2OX2	gibberellin 2beta-dioxygenase (EC 1.14.11.13)	Diterpenoid biosynthesis
K05907	APR	adenylyl-sulfate reductase (glutathione) (EC 1.8.4.9)	Sulfur metabolism
K05933	ACO2 ACO1	minocyclopropanecarboxylate oxidase (EC 1.14.17.4)	Cysteine and methionine metabolism
K10775	PAL	phenylalanine ammonia-lyase (EC 4.3.1.24)	Phenylalanine metabolism Phenylpropanoid biosynthesis
K13248	PHOSPHO2	pyridoxal phosphate phosphatase (EC 3.1.3.74)	Vitamin B6 metabolism
K14652	ribBA	3,4-dihydroxy 2-butanone 4-phosphate synthase / GTP cyclohydrolase II (EC 4.1.99.12 3.5.4.25)	Folate biosynthesis, Riboflavin metabolism
K18696	GDE1	glycerophosphodiester phosphodiesterase (EC 3.1.4.46)	Glycerophospholipid metabolism
K18819	GOLS	inositol 3-alpha-galactosyltransferase (EC 2.4.1.123)	Galactose metabolism
K20772	ACS1_2_6	1-aminocyclopropane-1-carboxylate synthase 1/2/6 (EC 4.4.1.14)	Cysteine and methionine metabolism
K22395	CAD	cinnamyl-alcohol dehydrogenase (EC 1.1.1.195)	Phenylpropanoid biosynthesis

Table 4.6: Metabolic mapping of top 20 downregulated genes in ‘Sindhri’

KEGG Orthology	Gene(s)	Enzyme	KEGG Pathway Map
K00600	glyA SHMT	glycine hydroxymethyltransferase (EC 2.1.2.1)	Cyanoamino acid metabolism Glyoxylate and dicarboxylate metabolism Glycine, serine and threonine metabolism Methane metabolism
K03858	PIGH GPI15	phosphatidylinositol N-acetylglucosaminyltransferase subunit H	Glycosylphosphatidylinositol (GPI)-anchor biosynthesis
K09755	CYP84AF5 H	ferulate-5-hydroxylase (EC 1.14.-.-) 5-	Phenylpropanoid biosynthesis
K00549	metE	methyltetrahydropteroyltriglutamate--homocysteine methyltransferase (EC 2.1.1.14)	Cysteine and methionine metabolism
K00587	ICMT STE14	protein-S-isoprenylcysteine O-methyltransferase (EC 2.1.1.100)	Terpenoid backbone biosynthesis
K01051	PME4	pectinesterase (EC 3.1.1.11)	Pentose and glucuronate interconversions
K01191	MAN2C1	alpha-mannosidase (EC 3.2.1.24)	Other glycan degradation
K01728	Pel	pectate lyase (EC 4.2.2.2)	Pentose and glucuronate interconversions
K01785	GALM	aldose 1-epimerase (EC 5.1.3.3)	Galactose metabolism Glycolysis / Gluconeogenesis
K01904	4CL	4-coumarate--CoA ligase (EC 6.2.1.12)	Phenylpropanoid biosynthesis Ubiquinone and other terpenoid-quinone biosynthesis
K02692	psaD	photosystem I subunit II	Photosynthesis
K02115	ATPF1G atpG	F-type H ⁺ -transporting ATPase subunit gamma	Oxidative phosphorylation Photosynthesis
K05280	CYP75B1	flavonoid 3'-monooxygenase (EC 1.14.14.82)	Flavone and flavonol biosynthesis Flavonoid biosynthesis
K07151	STT3	dolichyl-diphosphooligosaccharide---protein glycosyltransferase (EC 2.4.99.18)	N-Glycan biosynthesis
K07424	CYP3A	cytochrome P450 family 3 subfamily A (EC 1.14.14.1)	Linoleic acid metabolism, Steroid hormone biosynthesis, Retinol metabolism
K07964	HPSE	heparanase (EC 3.2.1.166)	Glycosaminoglycan degradation
K08902	psb27	photosystem II Psb27 protein	Photosynthesis
K13648	GAUT	alpha-1,4-galacturonosyltransferase (EC 2.4.1.43)	Amino sugar and nucleotide sugar metabolism
K16055	TPS	trehalose 6-phosphate synthase/phosphatase (EC 2.4.1.15 3.1.3.12)	Starch and sucrose metabolism
K20782	HPAT	hydroxyproline O-arabinosyltransferase (EC 2.4.2.58)	Other types of O-glycan biosynthesis

Furthermore, chalcone synthase (EC 2.3.1.74), an important enzyme of phenylpropanoid-flavonoid (PF) pathways which influence the level of flavonoid content in mango fruit peel and flesh (Hoang et al., 2015), was also found to be upregulated in ‘Sindhri’ (**Figure 4.16**). Carotenoids are responsible for the color of the fruit, upregulation of 15-cis-phytoene synthase (EC 2.4.1.32) which is part of β -carotene biosynthesis pathway, confirms the increase in the production of β -carotene during fruit maturation (Rocha Ribeiro et al., 2007; Ranganath et al., 2018) (**Figure 4.17**). Enzyme 9-cis-epoxycarotenoid dioxygenase (EC 1.13.11.51), which is involved in the production of abscisic acid in the carotenoid biosynthesis pathways also plays an important role in mango fruit maturation and ripening (Zaharah and Singh, 2012). This enzyme was also predicted to be upregulated in ‘Sindhri’ (**Figure 4.17**). Enzymes GTP cyclohydrolase II (EC 3.5.4.25) and 3,4-dihydroxy 2-butanone 4-phosphate synthase (EC 4.1.99.12) involved in the riboflavin biosynthesis pathway were also predicted to be upregulated at the mature stage in ‘Sindhri’ (**Figure 4.18**) suggesting the increase in the production of vitamin B2, also known as riboflavin, at the mature stage (Barbosa Gámez et al., 2017).

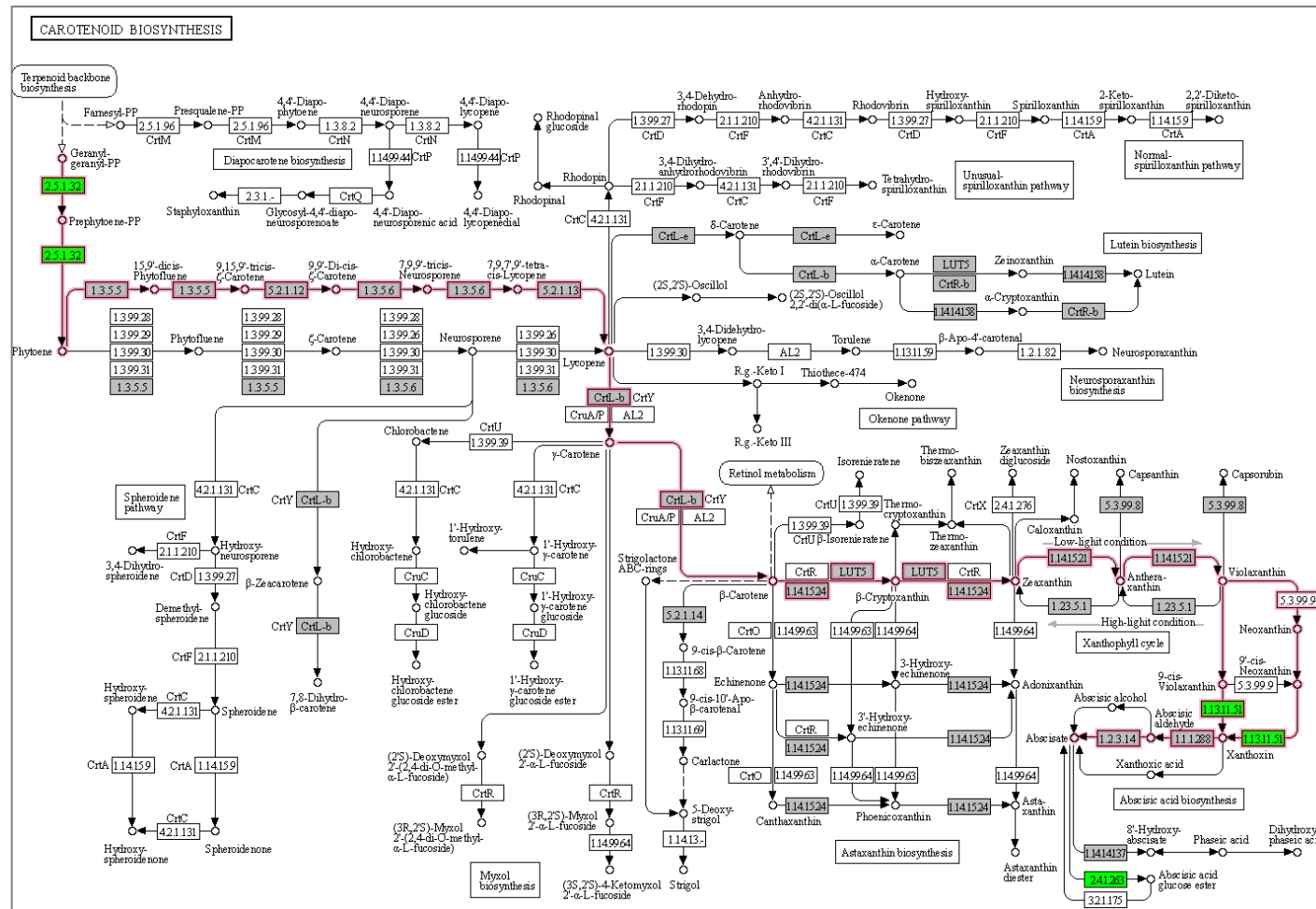


Figure 4.17: Metabolic mapping of genes differentially expressed during fruit maturation on carotenoid biosynthesis. Metabolic annotations of reference transcripts are highlighted in colored boxes. Enzymes predicted to be upregulated in ‘Sindhri’ during maturation are highlighted in green color. KEGG pathway modules of beta-carotene and abscisic acid biosynthesis are highlighted with solid border of pink color. Enzymes not found to be differentially expressed during fruit maturation are shown in grey color.

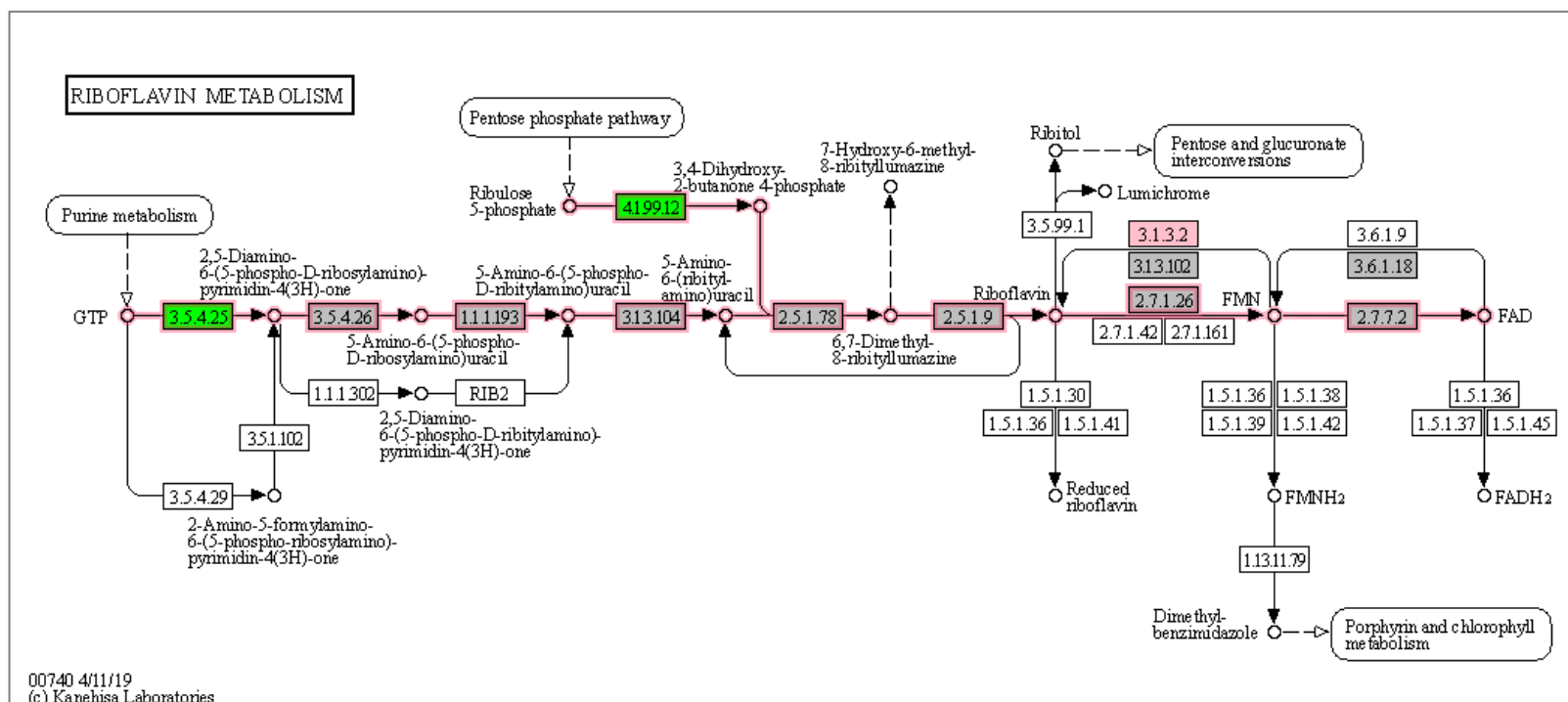


Figure 4.18: Metabolic mapping of genes differentially expressed during fruit maturation on riboflavin metabolism. Metabolic annotations of reference transcripts are highlighted in colored boxes. Enzymes predicted to be upregulated in ‘Sindhri’ during maturation are highlighted in green color and enzyme downregulated during maturation in ‘Kala Chaunsa’ is highlighted in pink color. KEGG pathway module of riboflavin biosynthesis is highlighted with solid border of pink color. Enzymes not found to be differentially expressed during fruit maturation are shown in grey color.

4.5 Discussion

Fruits provide useful food reserves and are an essential source of micronutrients, vitamins, and other phytochemicals. The quality of fruit is influenced by variety, nutritional status, and environmental conditions during the growth of the parent plant (Tharanathan et al., 2006). They also play a significant role in enhancing human nutrition by providing a source of energy, necessary growth factors, carbohydrates, dietary fiber, and antioxidants (Rocha Ribeiro et al., 2007), which are essential for maintaining normal health. The mango fruits are utilized for domestic and commercial purposes at every stage of growth (Litz, 2009). The maturity of mango fruit has been correlated with change in various physical characteristics such as color, shape and size, and chemical parameters such as total soluble solids, acidity, starch, phenolic compounds, and carotenoids (Tharanathan et al., 2006). Recent advancements in the generation and analysis of experimental data have revealed previously uncharacterized mechanisms during fruit development indicating the role different biochemical pathways play during fruit maturation.

In this chapter, metabolic annotations for mango were obtained by mapping the transcriptomic reference of an elite South Asian mango cultivar ‘Kala Chaunsa’ onto KEGG pathways (see **Section 4.2**). Metabolic annotation of transcriptomes using KAAS is commonly performed to understand the expression of key metabolic pathways (Rajesh et al., 2016; Naganeeswaran et al., 2020). KAAS performs similarity search against KEGG GENES database (Kanehisa et al., 2019) for assigning KEGG Orthology (KO) numbers to the query transcripts. For mango transcriptome, multiple transcripts were mapped to same KO numbers by KAAS (see **Table 4.3**, **Table 4.4**, **Appendix L** and **Appendix M**) where each KO number is linked to an enzyme involved in one or

more KEGG pathway maps. Mapping of multiple transcripts to the same KO number could be due to multiple reasons. One, it might be due to the presence of paralogs in the mango genome. Alternatively, multiple genes might be coding for the enzymes having similar functions resulting in overlapping annotations. Also, discrepancies in the KEGG annotations of representative gene sets can also result in redundant assignments. Despite these overlapping, the results presented in this chapter have provided a comprehensive metabolic annotation for mango fruit, which will enable further insights into the metabolic capabilities of mango fruit.

Metabolic annotation obtained using KAAS were subsequently mapped to various KEGG maps (see **Section 4.2**). While this identified key metabolic pathways active during fruit maturation, it also showed that some of the enzymes in the metabolic pathways were not mapped to any of the transcripts by KAAS leaving gaps in the metabolic pathways (see **Section 4.4**). Although a variety of different plants were used as representative set in this chapter, the mapping is limited by accuracy of annotations of the selected representative set available in the KEGG (Moriya et al., 2007). In addition, these gaps might be due to the absence of genes corresponding to these enzymes from the transcriptomic reference used for metabolic mapping. This can be improved by using transcriptomic references from multiple experiments providing a bigger pool of genes for enzyme mapping.

Next, genes differentially expressed between immature and mature fruit stages in two mango cultivars, ‘Sindhri’ and ‘Kala Chaunsa’ were identified. Analysis of differential expressed genes in ‘Sindhri’ and ‘Kala Chaunsa’ was performed using two R packages DESeq2 (Love et al., 2014) and EdgeR (Robinson et al., 2009) (see **Section 4.3**). Interestingly, both R packages predicted lower number of differentially expressed genes between immature and mature stages in ‘Kala Chaunsa’ compared to ‘Sindhri’.

While it is possible that fewer genes are involved in fruit maturation in ‘Kala Chaunsa’ compared to ‘Sindhri’, this might also be due to sampling error as visual and manual inspection was used for fruit stage determination in the study from which transcriptomic data was obtained (Hijazi et al, in preparation). Changes in the key metabolic pathways during fruit maturation have been reported in several studies (Rocha Ribeiro et al., 2007; Barbosa Gámez et al., 2017; Ranganath et al., 2018), hence several metabolic pathways are expected to be differentially expressed during maturation in ‘Kala Chaunsa’. Using objective classification of fruit stage (Skolik et al., 2019) and adding transcriptomic data for more biological replicates can further improve the differential expression analysis presented in this chapter.

The differentially expressed genes were subsequently mapped to the enzymes involved in various KEGG pathways to understand the changes that occur in the metabolism during fruit maturation. Associating expression data with metabolic pathways provided a functional context to observed gene expression patterns in both cultivars during fruit maturation. A higher portion of mapped enzymes was found to participate in the metabolism of carbohydrates, lipids and amino acids, and secondary metabolite pathways, including phenylpropanoid biosynthesis, flavonoid biosynthesis, and carotenoid biosynthesis. Differential regulation of metabolic pathways between ‘Sindhri’ and ‘Kala Chaunsa’ was observed in brassinosteroid biosynthesis and phenylpropanoid biosynthesis, where enzymes were found to be upregulated in one cultivar but downregulated in the other suggesting metabolic variation between these South Asian mango cultivars (see **Section 4.4**). While mapping of transcriptomic data on KEGG pathways provided an overview of metabolic changes during fruit maturation in both cultivars, the availability of a genome-scale metabolic model would enhance the application of this approach by the integration of expression data into the metabolic

model for defining constraints on reaction flux and prediction of metabolic phenotypes (Colijn et al., 2009; Rezola et al., 2015).

Inferring metabolic phenotypes based on changes in the expression levels are based on the fact that RNA is translated into proteins, including metabolic enzymes and that the cell enhances (or reduces) the mRNA transcription rate to enhance (or reduced) a function for which the coded protein is required (not required) (Hoppe, 2012). Comparison of mRNA expression levels and enzymatic activities have revealed low correlations between the transcriptome and metabolome inside a cell indicating that transcriptomic analysis is not sufficient to understand protein dynamics or biochemical regulation (Glanemann et al., 2003; Gibon et al., 2006; Wienkoop et al., 2008). Addition of proteomics and metabolomics data would provide more resolution to understand the changes in enzyme activity and metabolite profiles (Gligorijević and Pržulj, 2015; Yugi et al., 2016) thus enabling a better understanding of metabolic regulation in mango during fruit maturation.

5 Discussion

The understanding of biological networks is a fundamental problem in computational biology since cellular processes are regulated by the interaction between different layers of biological networks, including phenome, metabolome, proteome, transcriptome, epigenome, and genome (Rajasundaram and Selbig, 2016). Metabolic networks are one of the intricate networks comprising chemical interactions and transformations between metabolites and enzymes. Series of metabolic reactions convert the source metabolites into target metabolites and form metabolic pathways. Metabolic pathways provide an intermediate link to connect phenotypic variations to underlying regulations at the transcriptomic level, which is further used to enhance the quality or quantity of a specific trait or characteristic (Barbosa Gámez et al., 2017). Over the years, computational tools have been used to study metabolic networks, from the modeling of enzymes to studying the dynamics of metabolic pathways (Jing et al., 2014). With the advent of efficient and low-cost ‘omics technologies, pathway analyses have become more diverse and capable of integrating ‘omics data (Yeang and Vingron, 2006; Zhu et al., 2012; Fondi and Liò, 2015). Moreover, the availability of metabolic data and pathway information of genome-sequenced organisms provides an opportunity to compare and analyze functional specialization and diversity among metabolic networks of various organisms, including bacteria and plants. Several tools have been developed to analyze metabolic data available in public databases, but these tools lack the integration of ‘omics data for metabolic network analysis and pathway prediction. Comparing metabolic networks and pathways by integrating context-specific ‘omics data can provide further insight into the metabolic capabilities of an organism and evolutionary studies.

In this thesis, a web-based tool called MAPPS: Metabolic network Aalysis and Pathway Prediction Server was presented that offers a wide range of analyses relating to the metabolic pathway prediction and network comparisons, and provides an intuitive way to answer different biological questions focusing on metabolic differences between multiple organisms as well as relating to metabolic evolution (**Chapter 2**). MAPPS enables users to perform prediction and comparison of metabolic pathways between one or more metabolites, comparisons of metabolic networks between organisms, and computing reactions and pathways at different levels of the phylogeny in addition to allowing specialized analyses including identification of potential drug targets, *in silico* metabolic engineering, analysis of host-pathogen interactions, and ancestral network building. MAPPS provides several advantages over existing software, including the provision of custom metabolic networks, filtering of networks based on ‘omics data, network- and pathway-based comparison between multiple organism sets, and evolutionary analysis using metabolic data. MAPPS uses the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa et al., 2017) as the primary source of data. As a result, the accuracy and usefulness of MAPPS are limited by the availability of metabolic annotations in KEGG. This was highlighted in **Section 3.6**, which discussed the absence of one-stop nitrogen assimilation pathways in various pseudomonads due to missing connections in KEGG RCLASS. A large number of metabolic annotations currently available in public databases like KEGG are based on the similarity of predicted genes to genes of known function. These automatic annotations can be referred to as *predicted metabolisms* in contrast to those experimentally validated, the *known metabolisms*. These predicted metabolisms are subject to non-negligible errors or noise, which must be considered while analyzing metabolic networks. This can be done by using homologous information from known

metabolisms during pathway prediction and network analysis, in an approach similar to comparative genome annotation (Behr et al., 1999), to assign quality scores to the computed pathways for predicted metabolisms. This approach would provide an advantage of comparative prediction whereby metabolic annotations in one species will help or support the annotation in the other species.

Metabolic networks have been used for the optimization of phenotype as well as for the production of beneficial compounds using synthetic biology and metabolic engineering methods (Boyle and Silver, 2012; García-Granados et al., 2019). Scientists have engineered metabolic systems for the commercial production of a wide range of high-value biofuels (Majidian et al., 2018; Choi et al., 2020) and natural products (Pickens et al., 2011; Nielsen, 2019) into a bacterial host. Metabolic pathway prediction tools help in finding enzymes and non-redundant pathways of maximum yield and facilitate genome comparisons by aligning pathways to identify required links in the pathway (Fisher et al., 2014; Kampers et al., 2019). Potentially, the entire set of metabolic pathways can be (re)designed by using *in silico* approaches and implemented in specialized host organisms (Fisher et al., 2014; O'Connor, 2015). Many of the metabolic reactions are not active all the time; hence the integration of multiple layers of biological networks populated with 'omics data can enhance the efficiency and accuracy of metabolic predictions. Integrating proteomic and metabolomic data for host selection and optimization can help to improve the designing of heterologous pathways. Besides, spatial and temporal variations inside a eukaryotic cell direct its metabolic profile. Enzymes and metabolites are not homogeneously present in the cell nor present at all times, as assumed in most of the pathway computation tools (Agapakis et al., 2012; Zecchin et al., 2015). Compartmentalization and adding context to metabolic data can further enhance the capability of metabolic pathway computation tools for *in silico*

pathway designing (Hinzpeter et al., 2017). Incorporation of compartment information into metabolic networks can facilitate the identification of drug targets in a specific context instead of considering metabolism as a whole.

Next, a detailed analysis of metabolic diversity and functional specialization in 111 species of *Pseudomonas* was presented (**Chapter 3**). Comparison of metabolic network and pathways of carbohydrate, amino acid, and energy metabolism were performed using MAPPS and custom scripts to identify conserved and unique reactions and pathways in different *Pseudomonas* strains. Moreover, pseudomonads were grouped based on metabolic similarity, and these groupings were compared to the phylogenetic relationships inferred using the sequences for 16S rRNA and four housekeeping genes. While differences in metabolic networks and pathways were identified at the individual strain level, overall results suggest that species with similar lifestyle tend to have a high degree of metabolic similarity and that species have adapted their metabolic networks to suit their lifestyles. Mapping of transcriptomic and metabolomic data collected for different *Pseudomonas* strains surviving in diverse environments can provide further insight into the adaptability of metabolism in *Pseudomonas* (Chernov et al., 2019; Y. Deng et al., 2019). Similarly, the integration of proteomics data can provide a quantitative scale to filter predicted metabolic pathways for analyzing metabolic diversity and active pathways under different conditions (Montero-Blay et al., 2020).

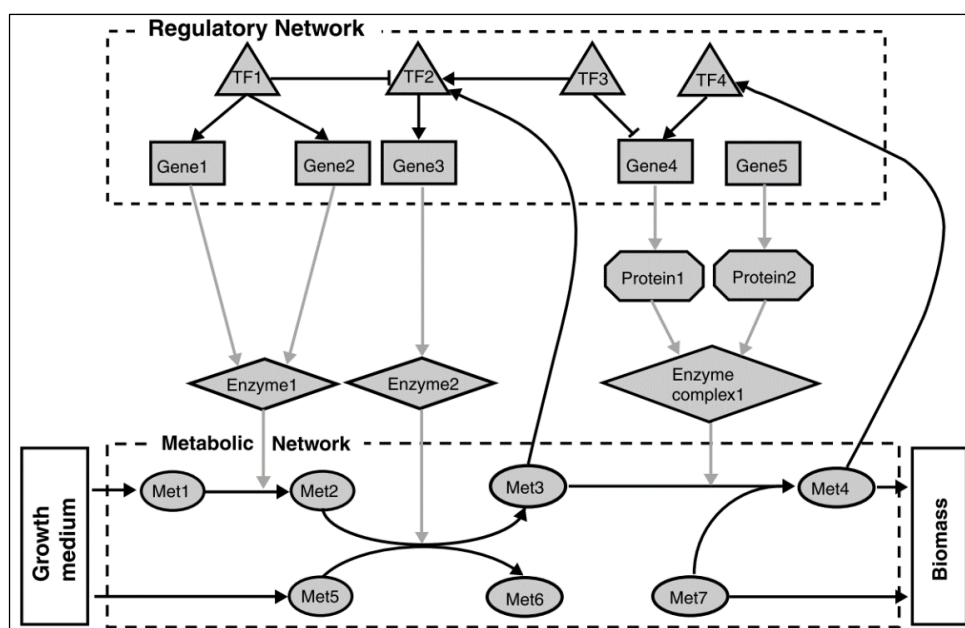
Over evolutionary time organisms, especially bacteria, adapt metabolically to different environments and interact with other organisms. Analysis of metabolic networks supplemented by ‘omics data and the environmental information can provide useful insights into the influence of external factors on metabolism enabling the study of the relationships between metabolic networks, phenotypes, and the environment.

This can help in answering questions pertinent to studying the evolution of various organisms, such as what are the selective pressures that influence the structure and function of biological networks, and what is it that an organism is trying to become good at? In some cases, it will be obvious as there might be a new nutrient source and a single new acquired reaction that allows this metabolite to be utilized. However, in many cases, there can be a complex situation without a simple nutrient source-reaction relationship. This is especially true when there is a multiplicity of metabolism-environment observations requiring a formal model to capture to understand this complex relationship.

Finally, to understand the effect of transcriptional changes during fruit development on underlying metabolic networks, differentially expressed genes of two different South Asian mango cultivars, ‘Sindhri’ and ‘Kala Chaunsa’, were identified and mapped to KEGG pathways in this thesis (**Chapter 4**). This metabolic mapping revealed variation in various metabolic pathways of ‘Sindhri’ and ‘Kala Chaunsa’. Most of the changes related to metabolism were predicted in carbohydrate, lipid, and amino acid metabolism in both the cultivars with key differences identified at the cultivar level. While mapping of expression data onto the metabolic networks provided useful insights into metabolic pathways upregulated or downregulated during fruit development, its reliability is limited by the accuracy of identification of differentially expressed genes. Adding other types of ‘omics data to the metabolic networks can increase the resolution of metabolic annotations. For example, it has been shown how metabolomics data can be used to understand the physiological changes in mango (Ganneru et al., 2019). Analyzing metabolite profiles obtained using metabolomics at different stages of fruit development will strengthen the results generated by

transcriptomic data and provide a reference to validate metabolic annotations in addition to the identification of missing links in the biochemical pathways.

Advancements in experimental protocols and reduction in the cost of ‘omics technologies have enabled researchers to use multi-omics data to analyze molecular changes at different levels of the biological systems simultaneously introducing the notion of integrative ‘omics (Choi and Pavelka, 2012). Analyzing different networks, including gene-regulatory networks and signaling networks along with metabolic networks, can help understand the effect of environmental changes and genetic perturbations on metabolic networks (M. Kim et al., 2016; Rajasundaram and Selbig, 2016) enabling a better understanding of the overall working of a cell. Gene regulatory networks, for example, have a direct impact on virtually all processes in the cell because they control the expression of the genes present in an organism, including the metabolic enzymes. A cell needs to regulate its enzyme production and behavior depending on its requirements. This is achieved through various regulatory mechanisms (Gonçalves et al., 2013), which allow the organism to adapt to environmental changes (Rodriguez-Martinez et al., 2016). Analyzing these networks together would help in a better understanding of the working of a cell as a whole. Using graph theoretical models as a framework for coupled regulatory-metabolic networks (Yeang and Vingron, 2006; Shlomi et al., 2007) similar to the one shown in **Figure 5.1** can be devised. By connecting the metabolic enzymes to their underlying genes, this will provide a platform to analyze diverse datasets including gene expression data and metabolic pathways within the same modeling framework to enable a better understanding of the functioning of metabolic networks and to relate the metabolic capabilities to phenotypes (Zur et al., 2010). Integrating this framework into MAPPS can allow users to predict metabolic pathways between metabolites based on gene expression data and to predict



metabolic networks active as well as preferred metabolic routes inside a cell can be predicted. Integrating these models into MAPPS can allow users to identify and compare metabolic pathways between metabolites that are preferred by different organisms and to study their effect on the evolution of these organisms.

Moreover, in the last decade, the focus has been shifted from the classical reductionist approach toward the holistic approach for understanding the interrelatedness of different layers of the cellular environment (Barabasi and Oltvai, 2004). Systems biology approaches developed for the understanding of cellular regulation put metabolic networks in focus, due to their role in shaping the molecular phenotype (Sweetlove and Ratcliffe, 2011). This, in turn, has led to the development of genome-scale models of metabolic networks. A genome-scale metabolic network of an organism provides a theoretical framework for querying the system and understanding its functional properties (Terzer et al., 2009). Hundreds of genome-scale metabolic models have been reconstructed in which reactions are assigned to sub-cellular compartments using sequence and proteomics derived localization information. Studies have created metabolic models of multiple tissues by connecting them through transport reactions (import and export) derived from the literature (Pfau et al., 2018; diCenzo et al., 2019). Similarly, metabolic networks comprising of gene-protein-reaction association data are now being built using information available in the public databases like BioCyc (Caspi et al., 2016). Besides this, *in silico* metabolic modeling has become increasingly established in modern biology research to explore, understand and analyze the changes in the metabolic fluxes in response to genetic and environmental perturbations (Harper et al., 2018). Various approaches have been proposed in the last decade to integrate transcriptomics data into metabolic flux analysis (Blazier and Papin, 2012), including methods that use the gene expression data to apply constraints on the

flux of a reaction (Colijn et al., 2009). These models and methods can be integrated into metabolic network analysis and pathway prediction tools to help them study and compare metabolic networks under different experimental conditions, thus allowing a better understanding of metabolic networks in today's post-genomic era.

6 References

- Aditi, Shariff, M., Beri, K., 2017. Exacerbation of bronchiectasis by *Pseudomonas monteilii*: a case report. *BMC Infect. Dis.* 17, 511.
- Agapakis, C.M., Boyle, P.M., Silver, P.A., 2012. Natural strategies for the spatial optimization of metabolism in synthetic biology. *Nat. Chem. Biol.*
- Ahyong, V., Berdan, C.A., Burke, T.P., Nomura, D.K., Welch, M.D., 2019. A Metabolic Dependency for Host Isoprenoids in the Obligate Intracellular Pathogen *Rickettsia parkeri* Underlies a Sensitivity to the Statin Class of Host-Targeted Therapeutics. *mSphere* 4.
- Akhtar, S., Mahmood, S., Naz, S., Nasir, M., Sultan, M.T., 2009. Sensory evaluation of mangoes (*Mangifera Indica* L.) grown in different regions of Pakistan. *Pakistan J. Bot.* 41, 2821–2829.
- Anders, S., Pyl, P.T., Huber, W., 2015. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169.
- Anzai, Y., Kim, H., Park, J.Y., Wakabayashi, H., Oyaizu, H., 2000. Phylogenetic affiliation of the pseudomonads based on 16S rRNA sequence. *Int. J. Syst. Evol. Microbiol.* 50, 1563–1589.
- Arnau, V.G., Sánchez, L.A., Delgado, O.D., 2015. *Pseudomonas yamanorum* sp. Nov., a psychrotolerant bacterium isolated from a subantarctic environment. *Int. J. Syst. Evol. Microbiol.* 65, 424–431.
- Azim, M.K., Khan, I.A., Zhang, Y., 2014. Characterization of mango (*Mangifera indica* L.) transcriptome and chloroplast genome. *Plant Mol. Biol.* 85, 193–208.
- Bairoch, A., 2000. The ENZYME database in 2000. *Nucleic Acids Res.* 28, 304–305.
- Bally, I.S.E., 1999. Changes in the cuticular surface during the development of mango (*Mangifera indica* L.) cv. Kensington Pride. *Sci. Hortic. (Amsterdam)*. 79, 13–22.
- Bally, I.S.E., Dillon, N.L., 2018. Mango (*Mangifera indica* l.) breeding, in: *Advances in Plant Breeding Strategies: Fruits*. Springer International Publishing, pp. 811–896.
- Barabasi, a, Oltvai, Z.N., 2004. Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5, 101–13.
- Barbosa Gámez, I., Caballero Montoya, K.P., Ledesma, N., Sáyago Ayerdi, S.G., García Magaña, M. de L., Bishop von Wettberg, E.J., Montalvo-González, E., 2017. Changes in the nutritional quality of five *Mangifera* species harvested at two maturity stages. *J. Sci. Food Agric.* 97, 4987–4994.
- Bartell, J.A., Blazier, A.S., Yen, P., Thøgersen, J.C., Jelsbak, L., Goldberg, J.B., Papin, J.A., 2017. Reconstruction of the metabolic network of *Pseudomonas aeruginosa* to interrogate virulence factor synthesis. *Nat. Commun.* 8, 14631.
- Barupal, D.K., Haldiya, P.K., Wohlgemuth, G., Kind, T., Kothari, S.L., Pinkerton, K.E., Fiehn, O., 2012. MetaMapp: mapping and visualizing metabolomic data by integrating information from biochemical pathways and chemical and mass spectral similarity. *BMC Bioinformatics* 13, 1.

- Bator, I., Wittgens, A., Rosenau, F., Tiso, T., Blank, L.M., 2020. Comparison of Three Xylose Pathways in *Pseudomonas putida* KT2440 for the Synthesis of Valuable Products. *Front. Bioeng. Biotechnol.* 7, 1–18.
- Behr, M.A., Wilson, M.A., Gill, W.P., Salamon, H., Schoolnik, G.K., Rane, S., Small, P.M., 1999. Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* (80-.). 284, 1520–1523.
- Bennasar, A., Rosselló-Mora, R., Lalucat, J., Moore, E.R.B., 1996. 16S rRNA gene sequence analysis relative to genomovars of *Pseudomonas stutzeri* and proposal of *Pseudomonas balearica* sp. nov. *Int. J. Syst. Bacteriol.* 46, 200–205.
- Berge, C., Minieka, E., 1973. *Graphs and Hypergraphs*, vol. 7.
- Berthelot, K., Estevez, Y., Deffieux, A., Peruch, F., 2012. Isopentenyl diphosphate isomerase: A checkpoint to isoprenoid biosynthesis. *Biochimie*.
- Bhat, F.A., Ganai, B.A., Uqab, B., 2017. Carbonic Anhydrase: Mechanism, Structure and Importance in Higher Plants. *Pelagia Res. Libr. Asian J. Plant Sci. Res.* 7, 17–23.
- Blazier, A.S., Papin, J.A., 2012. Integration of expression data in genome-scale metabolic network reconstructions. *Front. Physiol.*
- Blum, T., Kohlbacher, O., 2008. MetaRoute: Fast search for relevant metabolic routes for interactive network navigation and visualization. *Bioinformatics* 24, 2108–2109.
- Bonner, P.L., 2020. Peptidases in Plant Tissue, in: ELS. Wiley, pp. 1–12.
- Boscari, A., del Giudice, J., Ferrarini, A., Venturini, L., Zaffini, A.-L., Delledonne, M., Puppo, A., 2013. Expression Dynamics of the *Medicago truncatula* Transcriptome during the Symbiotic Interaction with *Sinorhizobium meliloti*: Which Role for Nitric Oxide? *Plant Physiol.* 161, 425–439.
- Botelho, J., Grosso, F., Peixe, L., 2019. Antibiotic resistance in *Pseudomonas aeruginosa* – Mechanisms, epidemiology and evolution. *Drug Resist. Updat.* 44, 100640.
- Bowden, S.D., Rowley, G., Hinton, J.C.D., Thompson, A., 2009. Glucose and glycolysis are required for the successful infection of macrophages and mice by *Salmonella enterica* serovar Typhimurium. *Infect. Immun.* 77, 3117–3126.
- Boyle, P.M., Silver, P.A., 2012. Parts plus pipes: Synthetic biology approaches to metabolic engineering. *Metab. Eng.* 14, 223–232.
- Buckel, W., 2001. Unusual enzymes involved in five pathways of glutamate fermentation. *Appl. Microbiol. Biotechnol.* 57, 263–273.
- Buescher, J.M., Antoniewicz, M.R., Boros, L.G., Burgess, S.C., Brunengraber, H., Clish, C.B., DeBerardinis, R.J., Feron, O., Frezza, C., Ghesquiere, B., Gottlieb, E., Hiller, K., Jones, R.G., Kamphorst, J.J., Kibbey, R.G., Kimmelman, A.C., Locasale, J.W., Lunt, S.Y., Maddocks, O.D., Malloy, C., Metallo, C.M., Meuillet, E.J., Munger, J., Nöh, K., Rabinowitz, J.D., Ralser, M., Sauer, U., Stephanopoulos, G., St-Pierre, J., Tennant, D.A., Wittmann, C., Vander Heiden, M.G., Vazquez, A., Voutsden, K., Young, J.D., Zamboni, N., Fendt, S.M., 2015. A roadmap for interpreting ¹³ C metabolite labeling patterns from cells. *Curr. Opin. Biotechnol.*

- Cabot, G., Zamorano, L., Moyà, B., Juan, C., Navas, A., Blázquez, J., Oliver, A., 2016. Evolution of *Pseudomonas aeruginosa* antimicrobial resistance and fitness under low and high mutation rates. *Antimicrob. Agents Chemother.* 60, 1767–1778.
- Calderón, C.E., Ramos, C., De Vicente, A., Cazorla, F.M., 2015. Comparative genomic analysis of *Pseudomonas chlororaphis* PCL1606 reveals new insight into antifungal compounds involved in biocontrol. *Mol. Plant-Microbe Interact.* 28, 249–260.
- Caspi, R., Billington, R., Ferrer, L., Foerster, H., Fulcher, C.A., Keseler, I.M., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L.A., Ong, Q., Paley, S., Subhraveti, P., Weaver, D.S., Karp, P.D., 2016. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* 44, D471–D480.
- Chae, L., Lee, I., Shin, J., Rhee, S.Y., 2012. Towards understanding how molecular networks evolve in plants. *Curr. Opin. Plant Biol.* 15, 177–184.
- Chae, T.U., Choi, S.Y., Kim, J.W., Ko, Y.S., Lee, S.Y., 2017. Recent advances in systems metabolic engineering tools and strategies. *Curr. Opin. Biotechnol.* 47, 67–82.
- Chernov, V.M., Chernova, O.A., Mouzykantov, A.A., Lopukhov, L.L., Aminov, R.I., 2019. Omics of antimicrobials and antimicrobial resistance. *Expert Opin. Drug Discov.*
- Choi, H., Pavelka, N., 2012. When one and one gives more than two: Challenges and opportunities of integrative omics. *Front. Genet.* 2.
- Choi, K.R., Jiao, S., Lee, S.Y., 2020. Metabolic engineering strategies toward production of biofuels. *Curr. Opin. Chem. Biol.*
- Chou, C.H., Chang, W.C., Chiu, C.M., Huang, C.C., Huang, H. Da, 2009. FMM: A web server for metabolic pathway reconstruction and comparative analysis. *Nucleic Acids Res.* 37, 129–134.
- Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Sayers, E.W., 2016. GenBank. *Nucleic Acids Res.* 44, D67–D72.
- Coleman, J.R., 2000. Carbonic Anhydrase and Its Role in Photosynthesis. Springer, Dordrecht, pp. 353–367.
- Colijn, C., Brandes, A., Zucker, J., Lun, D.S., Weiner, B., Farhat, M.R., Cheng, T.Y., Moody, D.B., Murray, M., Galagan, J.E., 2009. Interpreting expression data with metabolic flux models: Predicting *Mycobacterium tuberculosis* mycolic acid production. *PLoS Comput. Biol.* 5, 1000489.
- Costa-Silva, J., Domingues, D., Lopes, F.M., 2017. RNA-Seq differential expression analysis: An extended review and a software tool. *PLoS One.*
- Couillerot, O., Prigent-Combaret, C., Caballero-Mellado, J., Moëne-Loccoz, Y., 2009. *Pseudomonas fluorescens* and closely-related fluorescent pseudomonads as biocontrol agents of soil-borne phytopathogens. *Lett. Appl. Microbiol.* 48, 505–512.
- Croes, D., Couche, F., Wodak, S.J., Van Helden, J., 2006. Inferring meaningful pathways in weighted metabolic networks. *J. Mol. Biol.* 356, 222–236.
- Dandekar, T., Schuster, S., Snel, B., Huynen, M., Bork, P., 1999. Pathway alignment:

- application to the comparative analysis of glycolytic enzymes. *Biochem. J.* 343 Pt 1, 115–24.
- Deng, J.L., Xu, Y.H., Wang, G., 2019. Identification of potential crucial genes and key pathways in breast cancer using bioinformatic analysis. *Front. Genet.* 10, 695.
- Deng, Y., Ruan, Y., Ma, B., Timmons, M.B., Lu, H., Xu, X., Zhao, H., Yin, X., 2019. Multi-omics analysis reveals niche and fitness differences in typical denitrification microbial aggregations. *Environ. Int.* 132, 105085.
- Di Gioia, D., Luziatelli, F., Negroni, A., Ficca, A.G., Fava, F., Ruzzi, M., 2011. Metabolic engineering of *Pseudomonas fluorescens* for the production of vanillin from ferulic acid. *J. Biotechnol.* 156, 309–316.
- diCenzo, G.C., Michelangelo Tesi, Thomas Pfau, 2019. A Virtual Nodule Environment (ViNE) for modelling the inter-kingdom metabolic integration during symbiotic nitrogen fixation. *BioRxiv*.
- Didelot, X., Falush, D., 2007. Inference of bacterial microevolution using multilocus sequence data. *Genetics* 175, 1251–1266.
- Dieppois, G., Opota, O., Jorge, L., Lemaitre, B., 2015. *Pseudomonas entomophila*: A Versatile Bacterium with Entomopathogenic Properties, in: Ramos, J.L., Goldberg, J.B., Filloux, A. (Eds.), *Pseudomonas: Volume 7: New Aspects of Pseudomonas Biology*. Springer, pp. 1–316.
- Duan, G., Christian, N., Schwachtje, J., Walther, D., Ebenhöf, O., 2013. The metabolic interplay between plants and phytopathogens. *Metabolites* 3, 1–23.
- Fabregat, A., Sidiropoulos, K., Garapati, P., Gillespie, M., Hausmann, K., Haw, R., Jassal, B., Jupe, S., K€orninger, F., McKay, S., Matthews, L., May, B., Milacic, M., Rothfels, K., Shamovsky, V., Webber, M., Weiser, J., Williams, M., Wu, G., Stein, L., Hermjakob, H., D'Eustachio, P., 2016. The reactome pathway knowledgebase. *Nucleic Acids Res.* 44, D481–D487.
- Faust, K., Croes, D., van Helden, J., 2011. Prediction of metabolic pathways from genome-scale metabolic networks. *BioSystems* 105, 109–121.
- Fell, D.A., 1992. Metabolic control analysis: a survey of its theoretical and experimental development. *Biochem. J.* 330, 313–330.
- Felsenstein, J., 2004. *Inferring Phylogenies*. Sunderland (MA): Sinauer Associates.
- Felsenstein, J., 1996. Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *Methods Enzymol.* 266, 418–427.
- Ferguson, B.J., Indrasumunar, A., Hayashi, S., Lin, M.H., Lin, Y.H., Reid, D.E., Gresshoff, P.M., 2010. Molecular analysis of legume nodule development and autoregulation. *J. Integr. Plant Biol.* 52, 61–76.
- Feuillet, C., Leach, J.E., Rogers, J., Schnable, P.S., Eversole, K., 2011. Crop genome sequencing: Lessons and rationales. *Trends Plant Sci.*
- Fisher, A.K., Freedman, B.G., Bevan, D.R., Senger, R.S., 2014. A review of metabolic and enzymatic engineering strategies for designing and optimizing performance of microbial cell factories. *Comput. Struct. Biotechnol. J.*
- Fitch, W.M., Margoliash, E., 1967. Construction of Phylogenetic Trees. *Science* (80-.). 155, 279–284.

- Flamholz, A., Noor, E., Bar-Even, A., Milo, R., 2012. EQuilibrator - The biochemical thermodynamics calculator. *Nucleic Acids Res.* 40, 770–775.
- Flores, O., Prince, C., Nuñez, M., Vallejos, A., Mardones, C., Yañez, C., Besoain, X., Bastías, R., 2018. Genetic and phenotypic characterization of indole-producing isolates of *Pseudomonas syringae* pv. *actinidiae* obtained from Chilean Kiwifruit Orchards. *Front. Microbiol.* 9, 1907.
- Fondi, M., Liò, P., 2015. Multi -omics and metabolic modelling pipelines: Challenges and tools for systems microbiology. *Microbiol. Res.*
- Forst, C. V, Flamm, C., Hofacker, I.L., Stadler, P.F., 2006. Algebraic comparison of metabolic networks, phylogenetic inference, and metabolic innovation. *BMC Bioinformatics* 7, 67.
- Freilich, S., Zarecki, R., Eilam, O., Segal, E.S., Henry, C.S., Kupiec, M., Gophna, U., Sharan, Roded, Ruppín, Eytan, Sharan, R., Ruppín, E., Hekstra, D., Leibler, S., Khoruts, A., Dicksved, J., Jansson, J., Sadowsky, M., Klitgord, N., Segrè, D., MacArthur, R., O'Dwyer, J., Kembel, S., Green, J., Shapiro, J., Tikhonov, M., Tilman, D., Vetsigian, K., Jajoo, R., Kishony, R., 2011. Competitive and cooperative metabolic interactions in bacterial communities. *Nat. Commun.* 2, 589.
- Fridovich-Keil, J.L., 2006. Galactosemia: The Good the Bad and the Unknown. *J. Cell. Physiol.*
- Frimmersdorf, E., Horatzek, S., Pelnikevich, A., Wiehlmann, L., Schomburg, D., 2010. How *Pseudomonas aeruginosa* adapts to various environments: A metabolomic approach. *Environ. Microbiol.* 12, 1734–1747.
- Fuhrer, T., Fischer, E., Sauer, U., 2005. Experimental identification and quantification of glucose metabolism in seven bacterial species. *J. Bacteriol.* 187, 1581–1590.
- Fukushima, M., Kakinuma, K., Kawaguchi, R., 2002. Strains on the Basis of the *gyrB* Gene Sequence 40, 2779–2785.
- Funk, C.D., 2001. Prostaglandins and leukotrienes: Advances in eicosanoid biology. *Science* (80-).
- Gani, M., Rao, S., Miller, M., Scoular, S., 2019. *Pseudomonas* Mendocina bacteremia: A case study and review of literature. *Am. J. Case Rep.* 20, 453–458.
- Ganneru, S., Shaik, H., Peddi, K., Mudiam, M.K.R., 2019. Evaluating the metabolic perturbations in *Mangifera indica* (mango) ripened with various ripening agents/practices through gas chromatography-mass spectrometry based metabolomics . *J. Sep. Sci.* 1–9.
- García-Granados, R., Lerma-Escalera, J.A., Morones-Ramírez, J.R., 2019. Metabolic engineering and synthetic biology: Synergies, future, and challenges. *Front. Bioeng. Biotechnol.* 7, 36.
- Garrido-Sanz, D., Meier-Kolthoff, J.P., Göker, M., Martín, M., Rivilla, R., Redondo-Nieto, M., 2016. Genomic and genetic diversity within the *Pseudomonas fluoresces* complex. *PLoS One* 11, e0150183.
- Gibon, Y., Usadel, B., Blaesing, O.E., Kamlage, B., Hoehne, M., Trethewey, R., Stitt, M., 2006. Integration of metabolite with transcript and enzyme activity profiling during diurnal cycles in *Arabidopsis* rosettes. *Genome Biol.* 7, R76.

- Gillaspy, G., Ben-David, H., Gruissem, W., 1993. Fruits: A developmental perspective. *Plant Cell*.
- Giovannoni, J., 2001. Molecular Biology of Fruit Maturation and Ripening. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 52, 725–749.
- Glaeser, S.P., Kämpfer, P., 2015. Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. *Syst. Appl. Microbiol.*
- Glagoleva, A.Y., Shmakov, N.A., Shoeva, O.Y., Vasiliev, G. V., Shatskaya, N. V., Börner, A., Afonnikov, D.A., Khlestkina, E.K., 2017. Metabolic pathways and genes identified by RNA-seq analysis of barley near-isogenic lines differing by allelic state of the Black lemma and pericarp (Blp) gene. *BMC Plant Biol.* 17, 182.
- Glanemann, C., Loos, A., Gorret, N., Willis, L.B., O'Brien, X.M., Lessard, P.A., Sinskey, A.J., 2003. Disparity between changes in mRNA abundance and enzyme activity in *Corynebacterium glutamicum*: Implications for DNA microarray analysis. *Appl. Microbiol. Biotechnol.* 61, 61–68.
- Gligorijević, V., Pržulj, N., 2015. Methods for biological data integration: Perspectives and challenges. *J. R. Soc. Interface* 12, 20150571-.
- Gomila, M., Peña, A., Mulet, M., Lalucat, J., García-Valdés, E., 2015. Phylogenomics and systematics in *Pseudomonas*. *Front. Microbiol.* 6, 214.
- Gonçalves, E., Bucher, J., Ryll, A., Niklas, J., Mauch, K., Klamt, S., Rocha, M., Saez-Rodriguez, J., 2013. Bridging the layers: towards integration of signal transduction, regulation and metabolism into mathematical models. *Mol. Biosyst.* 9, 1576.
- Gong, H., Jiao, Y., Hu, W.W., Pua, E.C., 2005. Expression of glutathione-S-transferase and its role in plant growth and development in vivo and shoot morphogenesis in vitro. *Plant Mol. Biol.* 57, 53–66.
- Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., Di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652.
- Graudenzi, A., Maspero, D., Di Filippo, M., Gnugnoli, M., Isella, C., Mauri, G., Medico, E., Antoniotti, M., Damiani, C., 2018. Integration of transcriptomic data and metabolic networks in cancer samples reveals highly significant prognostic power. *J. Biomed. Inform.* 87, 37–49.
- Gross, H., Loper, J.E., 2009. Genomics of secondary metabolite production by *Pseudomonas* spp. *Nat. Prod. Rep.*
- Hadadi, N., Hafner, J., Shajkofci, A., Zisaki, A., Hatzimanikatis, V., 2016. ATLAS of Biochemistry: A Repository of All Possible Biochemical Reactions for Synthetic Biology and Metabolic Engineering Studies. *ACS Synth. Biol.* 5, 1155–1166.
- Handa, A.K., Tiznado-Hernández, M.E., Mattoo, A.K., 2012. Fruit development and ripening: A molecular perspective, in: *Plant Biotechnology and Agriculture*. Elsevier Inc., pp. 405–424.
- Handorf, T., Ebenhöf, O., 2007. MetaPath Online: A web server implementation of the

- network expansion algorithm. *Nucleic Acids Res.* 35, 613–618.
- Harper, L., Balasubramanian, D., Ohneck, E.A., Sause, W.E., Chapman, J., Mejia-Sosa, B., Lhakhang, T., Heguy, A., Tsirigos, A., Ueberheide, B., Boyd, J.M., Lun, D.S., Torres, V.J., 2018. *Staphylococcus aureus* responds to the central metabolite pyruvate to regulate virulence. *MBio* 9.
- Hastings, J., De Matos, P., Dekker, A., Ennis, M., Harsha, B., Kale, N., Muthukrishnan, V., Owen, G., Turner, S., Williams, M., Steinbeck, C., 2013. The ChEBI reference database and ontology for biologically relevant chemistry: Enhancements for 2013. *Nucleic Acids Res.* 41, D456–D463.
- Hattori, M., Tanaka, N., Kanehisa, M., Goto, S., 2010. SIMCOMP/SUBCOMP: Chemical structure search servers for network analyses. *Nucleic Acids Res.* 38, 652–656.
- Haynes, W.A., Higdon, R., Stanberry, L., Collins, D., Kolker, E., 2013. Differential Expression Analysis for Pathways. *PLoS Comput. Biol.* 9, e1002967.
- Heath, A.P., Bennett, G.N., Kaviraki, L.E., 2011. Identifying branched metabolic pathways by merging linear metabolic pathways. *Lect. Notes Comput. Sci.* (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 6577 LNBI, 70–84.
- Heather, J.M., Chain, B., 2016. The sequence of sequencers: The history of sequencing DNA. *Genomics*.
- Heinken, A., Sahoo, S., Fleming, R.M.T., Thiele, I., 2013. Systems-level characterization of a host-microbe metabolic symbiosis in the mammalian gut. *Gut Microbes* 4, 28–40.
- Herrgård, M.J., Lee, B.S., Portnoy, V., Palsson, B., 2006. Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in *Saccharomyces cerevisiae*. *Genome Res.* 16, 627–635.
- Heuston, S., Begley, M., Gahan, C.G.M., Hill, C., 2012. Isoprenoid biosynthesis in bacterial pathogens. *Microbiol. (United Kingdom)*.
- Hinzpeter, F., Gerland, U., Tostevin, F., 2017. Optimal Compartmentalization Strategies for Metabolic Microcompartments. *Biophys. J.* 112, 767–779.
- Hoang, V.L.T., Innes, D.J., Shaw, P.N., Monteith, G.R., Gidley, M.J., Dietzgen, R.G., 2015. Sequence diversity and differential expression of major phenylpropanoid-flavonoid biosynthetic genes among three mango varieties. *BMC Genomics* 16, 561.
- Hoppe, A., 2012. What mRNA abundances can tell us about metabolism. *Metabolites*.
- Huang, R., Wallqvist, A., Covell, D.G., 2006. Comprehensive analysis of pathway or functionally related gene expression in the National Cancer Institute's anticancer screen. *Genomics* 87, 315–328.
- Huang, S., Chaudhary, K., Garmire, L.X., 2017. More is better: Recent progress in multi-omics data integration methods. *Front. Genet.* 8, 84.
- Huang, Y., Zhong, C., Lin, H.X., Wang, J., 2017. A Method for Finding Metabolic Pathways Using Atomic Group Tracking. *PLoS One* 12, e0168725.
- Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., Sidiropoulos,

- K., Cook, J., Gillespie, M., Haw, R., Loney, F., May, B., Milacic, M., Rothfels, K., Sevilla, C., Shamovsky, V., Shorser, S., Varusai, T., Weiser, J., Wu, G., Stein, L., Hermjakob, H., D'Eustachio, P., 2020. The reactome pathway knowledgebase. *Nucleic Acids Res.* 48, D498–D503.
- Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., Barabasi, A.-L., 2000. The large-scale organization of metabolic networks. *Nature* 407, 651–654.
- Jiao, Z., Wu, N., Hale, L., Wu, W., Wu, D., Guo, Y., 2013. Characterisation of *Pseudomonas chlororaphis* subsp. *aurantiaca* strain Pa40 with the ability to control wheat sharp eyespot disease. *Ann. Appl. Biol.* 163, n/a-n/a.
- Jing, L.S., Shah, F.F.M., Mohamad, M.S., Hamran, N.L., Salleh, A.H.M., Deris, S., Alashwal, H., 2014. Database and tools for metabolic network analysis. *Biotechnol. Bioprocess Eng.* 19, 568–585.
- Jung, H., Winefield, C., Bombarely, A., Prentis, P., Waterhouse, P., 2019. Tools and Strategies for Long-Read Sequencing and De Novo Assembly of Plant Genomes. *Trends Plant Sci.*
- Kampers, L.F.C., Van Heck, R.G.A., Donati, S., Saccenti, E., Volkers, R.J.M., Schaap, P.J., Suarez-Diez, M., Nikel, P.I., Martins Dos Santos, V.A.P., 2019. In silico-guided engineering of *Pseudomonas putida* towards growth under micro-oxic conditions. *Microb. Cell Fact.* 18, 179.
- Kanehisa, M., 2006. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.* 34, D354–D357.
- Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., Morishima, K., 2017. KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45, D353–D361.
- Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K., Tanabe, M., 2019. New approach for understanding genome variations in KEGG. *Nucleic Acids Res.* 47, D590–D595.
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., Tanabe, M., 2016. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44, D457–D462.
- Karp, P.D., Billington, R., Caspi, R., Fulcher, C.A., Latendresse, M., Kothari, A., Keseler, I.M., Krummenacker, M., Midford, P.E., Ong, Q., Ong, W.K., Paley, S.M., Subhraveti, P., 2018. The BioCyc collection of microbial genomes and metabolic pathways. *Brief. Bioinform.* 20, 1085–1093.
- Karp, P.D., Paley, S.M., Krummenacker, M., Latendresse, M., Dale, J.M., Lee, T.J., Kaipa, P., Gilham, F., Spaulding, A., Popescu, L., Altman, T., Paulsen, I., Keseler, I.M., Caspi, R., 2009. Pathway tools version 13.0 : integrated software for pathway / genome informatics and systems biology. *Brief. Bioinform.* 2, 40–79.
- Khosraviani, M., Zamani, M.S., Bidkhori, G., 2015. FogLight: An efficient matrix-based approach to construct metabolic pathways by search space reduction. *Bioinformatics* 32, 398–408.
- Kim, M., Rai, N., Zorraquino, V., Tagkopoulos, I., 2016. Multi-omics integration accurately predicts cellular state in unexplored conditions for *Escherichia coli*. *Nat. Commun.* 7, 13090.

- Kim, S., Thiessen, P.A., Bolton, E.E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B.A., Wang, J., Yu, B., Zhang, J., Bryant, S.H., 2016. PubChem substance and compound databases. *Nucleic Acids Res.* 44, D1202–D1213.
- King, Z.A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J.A., Ebrahim, A., Palsson, B.O., Lewis, N.E., 2016. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res.* 44, D515–D522.
- Klepikova, A. V., Kasianov, A.S., Gerasimov, E.S., Logacheva, M.D., Penin, A.A., 2016. A high resolution map of the *Arabidopsis thaliana* developmental transcriptome based on RNA-seq profiling. *Plant J.* 88, 1058–1070.
- Koehorst, J.J., Van Dam, J.C.J., Van Heck, R.G.A., Saccenti, E., Dos Santos, V.A.P.M., Suarez-Diez, M., Schaap, P.J., 2016. Comparison of 432 *Pseudomonas* strains through integration of genomic, functional, metabolic and expression data. *Sci. Rep.* 6, 1–13.
- Koprivova, A., Kopriva, S., 2014. Molecular mechanisms of regulation of sulfate assimilation: first steps on a long road. *Front. Plant Sci.* 5, 589.
- Krebs, E.G., 1972. Protein Kinases, in: *Current Topics in Cellular Regulation*. Academic Press, pp. 99–133.
- Kreimer, A., Borenstein, E., Gophna, U., Ruppín, E., 2008. The evolution of modularity in bacterial metabolic networks. *Proc. Natl. Acad. Sci.* 105, 6976–6981.
- Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549.
- Kuwahara, H., Alazmi, M., Cui, X., Gao, X., 2016. MRE: a web tool to suggest foreign enzymes for the biosynthesis pathway design with competing endogenous reactions in mind. *Nucleic Acids Res.* 44, gkw342.
- La Rosa, R., Johansen, H.K., Molina, S., 2018. Convergent metabolic specialization through distinct evolutionary paths in *Pseudomonas aeruginosa*. *MBio* 9.
- Lakshminarayana, S., 1973. Respiration and ripening patterns in the life cycle of the mango fruit. *J. Hortic. Sci.* 48, 227–233.
- Lalucat, J., Bennasar, A., Bosch, R., García-Valdés, E., Palleroni, N.J., 2006. Biology of *Pseudomonas stutzeri*. *Microbiol. Mol. Biol. Rev.* 70, 510–547.
- Lee, J., Cho, Y.J., Yang, J.Y., Jung, Y.J., Hong, S.G., Kim, O.S., 2017. Complete genome sequence of *Pseudomonas antarctica* PAMC 27494, a bacteriocin-producing psychrophile isolated from Antarctica. *J. Biotechnol.* 259, 15–18.
- Lee, S.Y., 2012. Biochemical pathways: an atlas of biochemistry and molecular biology, *Biotechnology Journal*.
- Lehninger, 2004. *Lehninger Principles of Biochemistry* W.H.Freeman2004.pdf. New York: WH Freeman.
- Letunic, I., Bork, P., 2019. Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* 47, 256–259.
- Levy, R., Carr, R., Kreimer, A., Freilich, S., Borenstein, E., 2015. NetCooperate: a

- network-based tool for inferring host-microbe and microbe-microbe cooperation. *BMC Bioinformatics* 16, 164.
- Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760.
- Lin, Q., Yang, J., Wang, Q., Zhu, H., Chen, Z., Dao, Y., Wang, K., 2019. Overexpression of the trehalose-6-phosphate phosphatase family gene AtTPPF improves the drought tolerance of *Arabidopsis thaliana*. *BMC Plant Biol.* 19, 1–15.
- Litz, R.E., 2009. The Mango, 2nd Edition: Botany, Production and Uses. Cabi 1–18.
- Loewen, P.C., Villeneuve, J., Fernando, W.G.D., de Kievit, T., 2014. Genome sequence of *Pseudomonas chlororaphis* strain PA23. *Genome Announc.* 2.
- Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 1–21.
- Lu, X., Li, L., Wu, R., Feng, X., Li, Z., Yang, H., Wang, C., Guo, H., Galkin, A., Herzberg, O., Mariano, P.S., Martin, B.M., Dunaway-Mariano, D., 2006. Kinetic analysis of *Pseudomonas aeruginosa* arginine deiminase mutants and alternate substrates provides insight into structural determinants of function. *Biochemistry* 45, 1162–1172.
- Luria, N., Sela, N., Yaari, M., Feygenberg, O., Kobiler, I., Lers, A., Prusky, D., 2014. De-novo assembly of mango fruit peel transcriptome reveals mechanisms of mango response to hot water treatment. *BMC Genomics* 15, 957.
- Ma, H., Zeng, A.-P.A., 2003. Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. *Bioinformatics* 19, 270–277.
- Madeira, F., Park, Y. mi, Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R.N., Potter, S.C., Finn, R.D., Lopez, R., 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* 47, W636–W641.
- Mailloux, R.J., Lemire, J., Appanna, V.D., 2011. Metabolic networks to combat oxidative stress in *Pseudomonas fluorescens*. *Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol.* 99, 433–442.
- Majidian, P., Tabatabaei, M., Zeinolabedini, M., Naghshbandi, M.P., Chisti, Y., 2018. Metabolic engineering of microorganisms for biofuel production. *Renew. Sustain. Energy Rev.*
- Manaia, C.M., Moore, E.R.B., 2002. *Pseudomonas thermotolerans* sp. nov., a thermotolerant species of the genus *Pseudomonas* sensu stricto. *Int. J. Syst. Evol. Microbiol.* 52, 2203–2209.
- Marco-Puche, G., Lois, S., Benítez, J., Trivino, J.C., 2019. RNA-Seq Perspectives to Improve Clinical Diagnosis. *Front. Genet.*
- Martz, E.O., Lakes, R.S., Park, J.B., 1996. Hysteresis behaviour and specific damping capacity of negative Poisson's ratio foams. *Cell. Polym.* 15, 349–364.
- McClymont, K., Soyer, O.S., 2013. Metabolic tinker: An online tool for guiding the design of synthetic metabolic pathways. *Nucleic Acids Res.* 41.
- McCombie, W.R., McPherson, J.D., Mardis, E.R., 2019. Next-generation sequencing

- technologies. Cold Spring Harb. Perspect. Med. 9.
- McKee, T., McKee, J., 2016. The Molecular Basis of Life, Oxford University Press. Oxford: Oxford University Press.
- Mentzen, W.I., Peng, J., Ransom, N., Nikolau, B.J., Wurtele, E., 2008. Articulation of three core metabolic processes in Arabidopsis: Fatty acid biosynthesis, leucine catabolism and starch metabolism. BMC Plant Biol. 8, 76.
- Mishra, D., Chaturvedi, V.K., Snijesh, V.P., Shaik, N.A., Singh, M.P., 2019. Other Biological Databases, Essentials of Bioinformatics, Volume I.
- Mithani, A., Hein, J., Preston, G.M., 2011. Comparative analysis of metabolic networks provides insight into the evolution of plant pathogenic and nonpathogenic lifestyles in Pseudomonas. Mol. Biol. Evol. 28, 483–499.
- Mithani, A., Preston, G.M., Hein, J., 2010. A Bayesian approach to the evolution of metabolic networks on a phylogeny. PLoS Comput. Biol. 6.
- Mithani, A., Preston, G.M., Hein, J., 2009a. Rahnuma: Hypergraph-based tool for metabolic pathway prediction and network comparison. Bioinformatics 25, 1831–1832.
- Mithani, A., Preston, G.M., Hein, J., 2009b. A stochastic model for the evolution of metabolic networks with neighbor dependence. Bioinformatics 25, 1528–1535.
- Monias, B.L., 1928. Classification of bacterium alcaligenes, pyocyaneum and fluorescens. J. Infect. Dis. 43, 330–334.
- Monk, J., Nogales, J., Palsson, B.O., 2014. Optimizing genome-scale network reconstructions. Nat. Biotechnol.
- Montero-Blay, A., Piñero-Lambea, C., Miravet-Verde, S., Lluch-Senar, M., Serrano, L., 2020. Inferring Active Metabolic Pathways from Proteomics and Essentiality Data. Cell Rep. 31, 107722.
- Montes, C., Altimira, F., Canchignia, H., CastroÁlvaro, Sánchez, E., Miccono, M., Tapia, E., SequeidaÁlvaro, Valdés, J., Tapia, P., González, C., Prieto, H., 2016. A draft genome sequence of Pseudomonas veronii R4: A grapevine (Vitis vinifera L.) root-associated strain with high biocontrol potential. Stand. Genomic Sci. 11, 76.
- Moons, A., 2005. Regulatory and Functional Interactions of Plant Growth Regulators and Plant Glutathione S-Transferases (GSTs). Vitam. Horm.
- Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C., Kanehisa, M., 2007. KAAS: An automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35, W182–W185.
- Moriya, Y., Shigemizu, D., Hattori, M., Tokimatsu, T., Kotera, M., Goto, S., Kanehisa, M., 2010. PathPred: An enzyme-catalyzed metabolic pathway prediction server. Nucleic Acids Res. 38, 138–143.
- Mukherjee, S.K., 1953. The mango-its botany, cultivation, uses and future improvement, especially as observed in india. Econ. Bot. 7, 130–162.
- Mulet, M., Lalucat, J., García-Valdés, E., 2010. DNA sequence-based analysis of the Pseudomonas species. Environ. Microbiol. 12, 1513–1530.
- Nabavi, S.M., Šamec, D., Tomczyk, M., Milella, L., Russo, D., Habtemariam, S.,

- Suntar, I., Rastrelli, L., Daglia, M., Xiao, J., Giampieri, F., Battino, M., Sobarzo-Sanchez, E., Nabavi, S.F., Yousefi, B., Jeandet, P., Xu, S., Shirooie, S., 2020. Flavonoid biosynthetic pathways in plants: Versatile targets for metabolic engineering. *Biotechnol. Adv.*
- Naganeeswaran, S., Fayas, T.P., Rajesh, M.K., 2020. Dataset of transcriptome assembly of date palm embryogenic calli and functional annotation. *Data Br.* 31, 105760.
- Nazish, T., Shabbir, G., Ali, A., Sami-ul-Allah, S., Naeem, M., Javed, M., Batool, S., Arshad, H., Hussain, S.B., Aslam, K., Seher, R., Tahir, M., Baber, M., 2017. Molecular diversity of pakistani mango (*Mangifera indica* L.) varieties based on microsatellite markers. *Genet. Mol. Res.* 16, 1–8.
- Nelson, K.E., Weinel, C., Paulsen, I.T., Dodson, R.J., Hilbert, H., Martins dos Santos, V.A.P., Fouts, D.E., Gill, S.R., Pop, M., Holmes, M., Brinkac, L., Beanan, M., DeBoy, R.T., Daugherty, S., Kolonay, J., Madupu, R., Nelson, W., White, O., Peterson, J., Khouri, H., Hance, I., Chris Lee, P., Holtzapple, E., Scanlan, D., Tran, K., Moazzez, A., Utterback, T., Rizzo, M., Lee, K., Kosack, D., Moestl, D., Wedler, H., Lauber, J., Stjepandic, D., Hoheisel, J., Straetz, M., Heim, S., Kiewitz, C., Eisen, J., Timmis, K.N., Dusterhöft, A., Tümmeler, B., Fraser, C.M., 2002. Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environ. Microbiol.* 4, 799–808.
- Nielsen, J., 2019. Cell factory engineering for improved production of natural products. *Nat. Prod. Rep.*
- Nikel, P.I., de Lorenzo, V., 2018. *Pseudomonas putida* as a functional chassis for industrial biocatalysis: From native biochemistry to trans-metabolism. *Metab. Eng.*
- Nikiforova, V.J., Gakière, B., Kempa, S., Adamik, M., Willmitzer, L., Hesse, H., Hoefgen, R., 2004. Towards dissecting nutrient metabolism in plants: A systems biology case study on sulphur metabolism. *J. Exp. Bot.* 55, 1861–1870.
- Nolan, T.M., Vukasinović, N., Liu, D., Russinova, E., Yin, Y., 2020. Brassinosteroids: Multidimensional regulators of plant growth, development, and stress responses, in: *Plant Cell*. American Society of Plant Biologists, pp. 298–318.
- O'Connor, S.E., 2015. Engineering of Secondary Metabolism. *Annu. Rev. Genet.* 49, 71–94.
- Oberhardt, M.A., Goldberg, J.B., Hogardt, M., Papin, J.A., 2010. Metabolic network analysis of *Pseudomonas aeruginosa* during chronic cystic fibrosis lung infection. *J. Bacteriol.* 192, 5534–5548.
- Oberhardt, M.A., Palsson, B., Papin, J.A., 2009. Applications of genome-scale metabolic reconstructions. *Mol. Syst. Biol.*
- Oberhardt, M.A., Puchalka, J., Fryer, K.E., Martins Dos Santos, V.A.P., Papin, J.A., 2008. Genome-scale metabolic network analysis of the opportunistic pathogen *Pseudomonas aeruginosa* PAO1. *J. Bacteriol.* 190, 2790–2803.
- Ogata, H., Goto, S., Fujibuchi, W., Kanehisa, M., 1998. Computation with the KEGG pathway database. *BioSystems* 47, 119–128.
- Pandit, S.S., Chidley, H.G., Kulkarni, R.S., Pujari, K.H., Giri, A.P., Gupta, V.S., 2009.

- Cultivar relationships in mango based on fruit volatile profiles. *Food Chem.* 114, 363–372.
- Pandit, S.S., Kulkarni, R.S., Giri, A.P., Köllner, T.G., Degenhardt, J., Gershenzon, J., Gupta, V.S., 2010. Expression profiling of various genes during the fruit development and ripening of mango. *Plant Physiol. Biochem.* 48, 426–433.
- Paulsen, I.T., Press, C.M., Ravel, J., Kobayashi, D.Y., Myers, G.S.A., Mavrodi, D. V., DeBoy, R.T., Seshadri, R., Ren, Q., Madupu, R., Dodson, R.J., Durkin, A.S., Brinkac, L.M., Daugherty, S.C., Sullivan, S.A., Rosovitz, M.J., Gwinn, M.L., Zhou, L., Schneider, D.J., Cartinhour, S.W., Nelson, W.C., Weidman, J., Watkins, K., Tran, K., Khouri, H., Pierson, E.A., Pierson, L.S., Thomashow, L.S., Loper, J.E., 2005. Complete genome sequence of the plant commensal *Pseudomonas fluorescens* Pf-5. *Nat. Biotechnol.* 23, 873–878.
- Perez-Garcia, O., Lear, G., Singhal, N., 2016. Metabolic network modeling of microbial interactions in natural and engineered environmental systems. *Front. Microbiol.* 7.
- Perumal, D., Lim, C.S., Sakharkar, M.K., 2009. A Comparative Study of Metabolic Network Topology between a Pathogenic and a Non-Pathogenic Bacterium for Potential Drug Target Identification. *Summit on Translat. Bioinforma.* 2009, 100–4.
- Perumal, D., Samal, A., Sakharkar, K.R., Sakharkar, M.K., 2011. Targeting multiple targets in *Pseudomonas aeruginosa* PAO1 using flux balance analysis of a reconstructed genome-scale metabolic network. *J. Drug Target.* 19, 1–13.
- Pfau, T., Christian, N., Masakapalli, S.K., Sweetlove, L.J., Poolman, M.G., Ebenhö, O., 2018. The intertwined metabolism during symbiotic nitrogen fixation elucidated by metabolic modelling. *Sci. Rep.* 8, 12504.
- Pickens, L.B., Tang, Y., Chooi, Y.-H., 2011. Metabolic Engineering for the Production of Natural Products. *Annu. Rev. Chem. Biomol. Eng.* 2, 211–236.
- Pireddu, L., Szafron, D., Lu, P., Greiner, R., 2006. The Path-A metabolic pathway prediction web server. *Nucleic Acids Res.* 34, W714-9.
- Placzek, S., Schomburg, I., Chang, A., Jeske, L., Ulbrich, M., Tillack, J., Schomburg, D., 2017. BRENDA in 2017: New perspectives and new tools in BRENDA. *Nucleic Acids Res.* 45, D380–D388.
- Planes, F.J., Beasley, J.E., 2009. Path finding approaches and metabolic pathways. *Discret. Appl. Math.* 157, 2244–2256.
- Ponomarova, O., Patil, K.R., 2015. Metabolic interactions in microbial communities: Untangling the Gordian knot. *Curr. Opin. Microbiol.* 27, 37–44.
- Prather, K.L.J., Martin, C.H., 2008. De novo biosynthetic pathways: rational design of microbial chemical factories. *Curr. Opin. Biotechnol.* 19, 468–474.
- Preston, G.M., Haubold, B., Rainey, P.B., 1998. Bacterial genomics and adaptation to life on plants: Implications for the evolution of pathogenicity and symbiosis. *Curr. Opin. Microbiol.* 1, 589–597.
- Puchałka, J., Oberhardt, M.A., Godinho, M., Bielecka, A., Regenhardt, D., Timmis, K.N., Papin, J.A., Martins Dos Santos, V.A.P., 2008. Genome-scale reconstruction and analysis of the *Pseudomonas putida* KT2440 metabolic network facilitates applications in biotechnology. *PLoS Comput. Biol.* 4, e1000210.

- Rahman, S.A., Advani, P., Schunk, R., Schrader, R., Schomburg, D., 2005. Metabolic pathway analysis web service (Pathway Hunter Tool at CUBIC). *Bioinformatics* 21, 1189–1193.
- Rajasundaram, D., Selbig, J., 2016. More effort - more results: Recent advances in integrative “omics” data analysis. *Curr. Opin. Plant Biol.* 30, 57–61.
- Rajesh, M.K., Fayas, T.P., Naganeeswaran, S., Rachana, K.E., Bhavyashree, U., Sajini, K.K., Karun, A., 2016. De novo assembly and characterization of global transcriptome of coconut palm (*Cocos nucifera* L.) embryogenic calli using Illumina paired-end sequencing. *Protoplasma* 253, 913–928.
- Ramette, A., Frapolli, M., Saux, M.F. Le, Gruffaz, C., Meyer, J.M., Défago, G., Sutra, L., Moënne-Loccoz, Y., 2011. *Pseudomonas protegens* sp. nov., widespread plant-protecting bacteria producing the biocontrol compounds 2,4-diacetylphloroglucinol and pyoluteorin. *Syst. Appl. Microbiol.* 34, 180–188.
- Ramon, C., Gollub, M.G., Stelling, J., 2018. Integrating –omics data into genome-scale metabolic network models: principles and challenges. *Essays Biochem.* 62, 563–574.
- Ranganath, K.G., Shivashankara, K.S., Roy, T.K., Dinesh, M.R., Geetha, G.A., Pavithra, K.C.G., Ravishankar, K. V., 2018. Profiling of anthocyanins and carotenoids in fruit peel of different colored mango cultivars. *J. Food Sci. Technol.* 55, 4566–4577.
- Ravikrishnan, A., Nasre, M., Raman, K., 2018. Enumerating all possible biosynthetic pathways from metabolic networks. *Sci. Rep.* 8, 9932.
- Ravishankar, K.V., Mani, B.H.R., Anand, L., Dinesh, M.R., 2011. Development of new microsatellite markers from Mango (*Mangifera indica*) and cross-species amplification. *Am. J. Bot.* 98.
- Reed, J.L., Vo, T.D., Schilling, C.H., Palsson, B.O., 2003. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol.* 4.
- Remold, S.K., Brown, C.K., Farris, J.E., Hundley, T.C., Perpich, J.A., Purdy, M.E., 2011. Differential Habitat Use and Niche Partitioning by *Pseudomonas* Species in Human Homes. *Microb. Ecol.* 62, 505–517.
- Ren, Z., Lee, J., Moosa, M.M., Nian, Y., Hu, L., Xu, Z., McCoy, J.G., Ferreón, A.C.M., Im, W., Zhou, M., 2018. Structure of an EIIC sugar transporter trapped in an inward-facing conformation. *Proc. Natl. Acad. Sci. U. S. A.* 115, 5962–5967.
- Rezola, A., Pey, J., Tobalina, L., Rubio, A., Beasley, J.E., Planes, F.J., 2015. Advances in network-based metabolic pathway analysis and gene expression data integration. *Brief. Bioinform.* 16, 265–279.
- Rico, A., Preston, G.M., 2008. *Pseudomonas syringae* pv. tomato DC3000 uses constitutive and apoplast-induced nutrient assimilation pathways to catabolize nutrients that are abundant in the tomato apoplast. *Mol. Plant-Microbe Interact.* 21, 269–282.
- Rienksma, R.A., Schaap, P.J., Martins dos Santos, V.A.P., Suarez-Diez, M., 2019. Modeling Host-Pathogen Interaction to Elucidate the Metabolic Drug Response of Intracellular *Mycobacterium tuberculosis*. *Front. Cell. Infect. Microbiol.* 9, 1–14.
- Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2009. edgeR: A Bioconductor package

- for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140.
- Rocha Ribeiro, S.M., De Queiroz, J.H., Lopes Ribeiro De Queiroz, M.E., Campos, F.M., Pinheiro Sant’Ana, H.M., 2007. Antioxidant in mango (*Mangifera indica* L.) pulp. *Plant Foods Hum. Nutr.* 62, 13–17.
- Rodriguez-Martinez, A., Ayala, R., Posma, J.M., Neves, A.L., Gauguier, D., Nicholson, J.K., Dumas, M.-E., 2016. MetaboSignal: a network-based approach for topological analysis of metabotype regulation via metabolic and signaling pathways. *Bioinformatics* btw697.
- Saha, R., Chowdhury, A., Maranas, C.D., 2014. Recent advances in the reconstruction of metabolic models and integration of omics data. *Curr. Opin. Biotechnol.* 29, 39–45.
- Sahraeian, S.M.E., Mohiyuddin, M., Sebra, R., Tilgner, H., Afshar, P.T., Au, K.F., Bani Asadi, N., Gerstein, M.B., Wong, W.H., Snyder, M.P., Schadt, E., Lam, H.Y.K., 2017. Gaining comprehensive biological insight into the transcriptome by performing a broad-spectrum RNA-seq analysis. *Nat. Commun.* 8, 1–15.
- Sakamoto, M., Ohkuma, M., 2011. Identification and classification of the genus bacteroides by multilocus sequence analysis. *Microbiology* 157, 3388–3397.
- Santos, C.N.S., Koffas, M., Stephanopoulos, G., 2011. Optimization of a heterologous pathway for the production of flavonoids from glucose. *Metab. Eng.* 13, 392–400.
- Sarkar, S.F., Guttman, D.S., 2004. Evolution of the Core Genome of *Pseudomonas syringae*, a Highly Clonal, Endemic Plant Pathogen. *Appl. Environ. Microbiol.* 70, 1999–2012.
- Sarle, W.S., Jain, A.K., Dubes, R.C., 1990. Algorithms for Clustering Data. *Technometrics* 32, 227.
- Sasaki, Y., Nagano, Y., 2004. Plant acetyl-CoA carboxylase: Structure, biosynthesis, regulation, and gene manipulation for plant breeding. *Biosci. Biotechnol. Biochem.*
- Sauer, U., 2006. Metabolic networks in motion: ¹³C-based flux analysis. *Mol. Syst. Biol.* 2.
- Scarpeci, T.E., Marro, M.L., Bortolotti, S., Boggio, S.B., Valle, E.M., 2007. Plant nutritional status modulates glutamine synthetase levels in ripe tomatoes (*Solanum lycopersicum* cv. Micro-Tom). *J. Plant Physiol.* 164, 137–145.
- Schiano, C., Costa, V., Aprile, M., Grimaldi, V., Maiello, C., Esposito, R., Soricelli, A., Colantuoni, V., Donatelli, F., Ciccodicola, A., Napoli, C., 2017. Heart failure: Pilot transcriptomic analysis of cardiac tissue by RNA-sequencing. *Cardiol. J.* 24, 539–553.
- Schweighofer, A., Meskiene, I., 2015. Phosphatases in plants. *Methods Mol. Biol.* 1306, 25–46.
- Seitz, A.P., Leadbetter, E.R., Godchaux, W., 1993. Utilization of sulfonates as sole sulfur source by soil bacteria including *Comamonas acidovorans*. *Arch. Microbiol.* 159, 440–444.
- Selvakumar, G., Joshi, P., Suyal, P., Mishra, P.K., Joshi, G.K., Bisht, J.K., Bhatt, J.C., Gupta, H.S., 2011. *Pseudomonas lurida* M2RH3 (MTCC 9245), a psychrotolerant

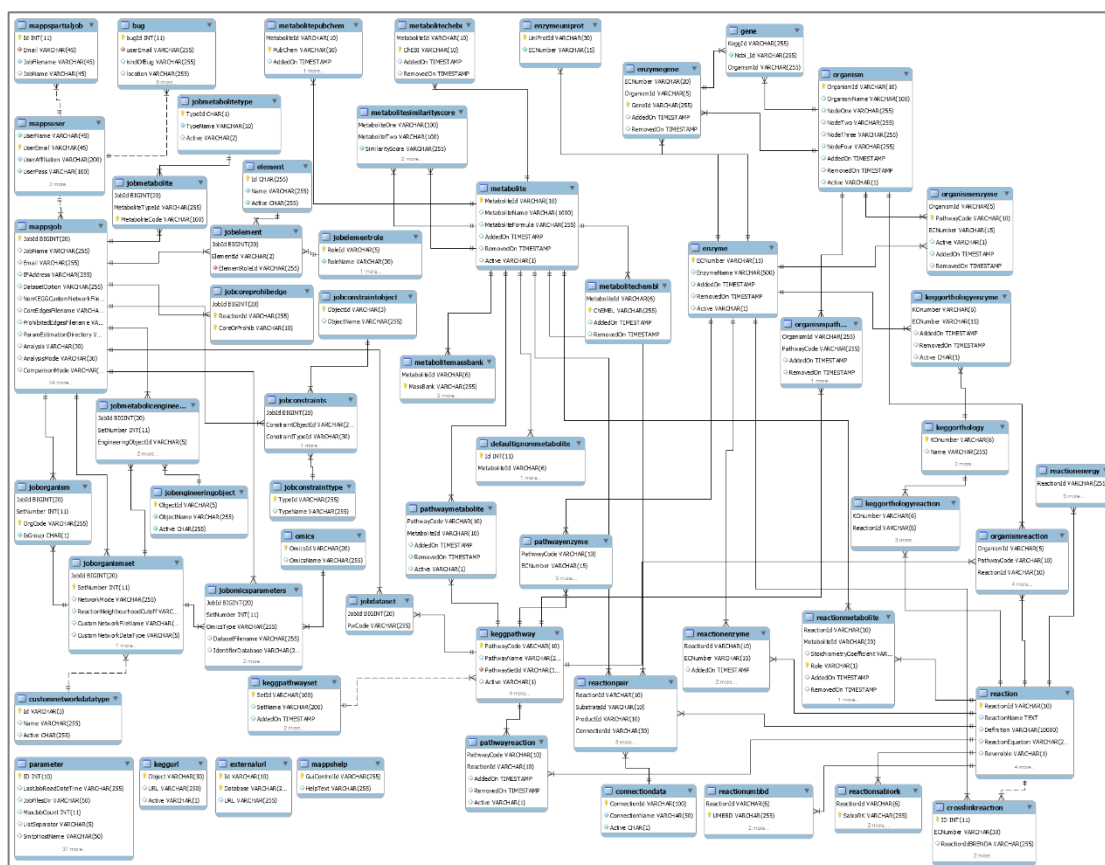
- bacterium from the Uttarakhand Himalayas, solubilizes phosphate and promotes wheat seedling growth. *World J. Microbiol. Biotechnol.* 27, 1129–1135.
- Setya, A., Murillo, M., Leustek, T., 1996. Sulfate reduction in higher plants: Molecular evidence for a novel 5'-adenylylsulfate reductase. *Proc. Natl. Acad. Sci. U. S. A.* 93, 13383–13388.
- Seymour, G.B., Ostergaard, L., Chapman, N.H., Knapp, S., Martin, C., 2013. Fruit development and ripening. *Annu. Rev. Plant Biol.* 64, 219–241.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504.
- Shi, C.Y., Yang, H., Wei, C.L., Yu, O., Zhang, Z.Z., Jiang, C.J., Sun, J., Li, Y.Y., Chen, Q., Xia, T., Wan, X.C., 2011. Deep sequencing of the *Camellia sinensis* transcriptome revealed candidate genes for major metabolic pathways of tea-specific compounds. *BMC Genomics* 12, 1–19.
- Shlomi, T., Eisenberg, Y., Sharan, R., Ruppin, E., 2007. A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Mol. Syst. Biol.* 3, 101.
- Silby, M.W., Winstanley, C., Godfrey, S.A.C., Levy, S.B., Jackson, R.W., 2011. *Pseudomonas* genomes: Diverse and adaptable. *FEMS Microbiol. Rev.* 35, 652–680.
- Skolik, P., Morais, C.L.M., Martin, F.L., McAinsh, M.R., 2019. Determination of developmental and ripening stages of whole tomato fruit using portable infrared spectroscopy and Chemometrics. *BMC Plant Biol.* 19, 236.
- Smith, E., Morowitz, H.J., 2004. Universality in intermediary metabolism. *Proc. Natl. Acad. Sci. U. S. A.* 101, 13168–13173.
- Spiers, A.J., Buckling, A., Rainey, P.B., 2000. The causes of *Pseudomonas* diversity. *Microbiology* 2345–2350.
- Spies, D., Ciaudo, C., 2015. Dynamics in Transcriptomics: Advancements in RNA-seq Time Course and Downstream Analysis. *Comput. Struct. Biotechnol. J.*
- Srivastava, A., Handa, A.K., 2005. Hormonal regulation of tomato fruit development: A molecular perspective. *J. Plant Growth Regul.*
- Stark, R., Grzelak, M., Hadfield, J., 2019. RNA sequencing: the teenage years. *Nat. Rev. Genet.* 20, 631–656.
- Subedi, D., Vijay, A.K., Kohli, G.S., Rice, S.A., Willcox, M., 2018. Comparative genomics of clinical strains of *Pseudomonas aeruginosa* strains isolated from different geographic sites. *Sci. Rep.* 8, 1–14.
- Sweetlove, L.J., Ratcliffe, R.G., 2011. Flux-balance modeling of plant metabolism. *Front. Plant Sci.* 2, 38.
- Tandon, D.K., Kalra, S.K., 1984. Pectin changes during the development of mango fruit cv Dashehari. *J. Hortic. Sci.* 59, 283–286.
- Tandon, D.K., Kalra, S.K., 1983. Changes in sugars, starch and amylase activity during development of mango fruit cv Dashehari. *J. Hortic. Sci.* 58, 449–453.

- Taylor, C.M., Wang, Q., Rosa, B.A., Huang, S.C.C., Powell, K., Schedl, T., Pearce, E.J., Abubucker, S., Mitreva, M., 2013. Discovery of Anthelmintic Drug Targets and Drugs Using Chokepoints in Nematode Metabolic Pathways. *PLoS Pathog.* 9, e1003505.
- Terzer, M., Maynard, N.D., Covert, M.W., Stelling, J., 2009. Genome-scale metabolic networks. *Wiley Interdiscip. Rev. Biol. Med.* 1, 285–297.
- Tharanathan, R.N., Yashoda, H.M., Prabha, T.N., 2006. Mango (*Mangifera indica* L.) , “The King of Fruits ”— An Overview. *Food Rev. Int.* 22, 95–123.
- The UniProt Consortium, 2015. UniProt: a hub for protein information. *Nucleic Acids Res.* 43, D204–12.
- Tokic, M., Hatzimanikatis, V., Miskovic, L., 2020. Large-scale kinetic metabolic models of *Pseudomonas putida* KT2440 for consistent design of metabolic engineering strategies. *Biotechnol. Biofuels* 13, 1–19.
- Tomar, N., De, R.K., 2013. Comparing methods for metabolic network analysis and an application to metabolic engineering. *Gene*.
- Udaondo, Z., Ramos, J.L., Segura, A., Krell, T., Daddaoua, A., 2018. Regulation of carbohydrate degradation pathways in *Pseudomonas* involves a versatile set of transcriptional regulators. *Microb. Biotechnol.*
- Wang, P., Luo, Y., Huang, J., Gao, S., Zhu, G., Dang, Z., Gai, J., Yang, M., Zhu, M., Zhang, H., Ye, X., Gao, A., Tan, X., Wang, S., Wu, S., Cahoon, E.B., Bai, B., Zhao, Z., Li, Q., Wei, J., Chen, H., Luo, R., Gong, D., Tang, K., Zhang, B., Ni, Z., Huang, G., Hu, S., Chen, Y., 2020. The genome evolution and domestication of tropical fruit mango. *Genome Biol.* 21.
- Weglarz, K.M., Havill, N.P., Burke, G.R., Von Dohlen, C.D., 2018. Partnering with a pest: Genomes of hemlock woolly adelgid symbionts reveal atypical nutritional provisioning patterns in dual-obligate bacteria. *Genome Biol. Evol.* 10, 1607–1621.
- Wienkoop, S., Morgenthal, K., Wolschin, F., Scholz, M., Selbig, J., Weckwerth, W., 2008. Integration of metabolomic and proteomic phenotypes. *Mol. Cell. Proteomics* 7, 1725–1736.
- Winsor, G.L., Griffiths, E.J., Lo, R., Dhillon, B.K., Shay, J.A., Brinkman, F.S.L., 2016. Enhanced annotations and features for comparing thousands of *Pseudomonas* genomes in the *Pseudomonas* genome database. *Nucleic Acids Res.* 44, D646–D653.
- Wu, H. xia, Jia, H. min, Ma, X. wei, Wang, S. biao, Yao, Q. sheng, Xu, W. tian, Zhou, Y. gang, Gao, Z. shan, Zhan, R. lin, 2014. Transcriptome and proteomic analysis of mango (*Mangifera indica* Linn) fruits. *J. Proteomics* 105, 19–30.
- Xia, D., Zheng, H., Liu, Z., Li, G., Li, J., Hong, J., Zhao, K., 2011. MRSD: A web server for Metabolic Route Search and Design. *Bioinformatics* 27, 1581–1582.
- Xin, X.F., Kvitko, B., He, S.Y., 2018. *Pseudomonas syringae*: What it takes to be a pathogen. *Nat. Rev. Microbiol.*
- Yamanishi, Y., Mihara, H., Osaki, M., Muramatsu, H., Esaki, N., Sato, T., Hizukuri, Y., Goto, S., Kanehisa, M., 2007. Prediction of missing enzyme genes in a bacterial metabolic network: Reconstruction of the lysine-degradation pathway of

- Pseudomonas aeruginosa*. FEBS J. 274, 2262–2273.
- Yeang, C.-H., Vingron, M., 2006. A joint model of regulatory and metabolic networks. BMC Bioinformatics 7, 332.
- Yeh, I., Hanekamp, T., Tsoka, S., Karp, P.D., Altman, R.B., 2004. Computational analysis of *Plasmodium falciparum* metabolism: Organizing genomic information to facilitate drug discovery. Genome Res. 14, 917–924.
- Yeung, M., Thiele, I., Palsson, B.O., 2007. Estimation of the number of extreme pathways for metabolic networks. BMC Bioinformatics 8, 363.
- Yu, S., Plan, M.R., Winter, G., Krömer, J.O., 2016. Metabolic engineering of *pseudomonas putida* KT2440 for the production of para-hydroxy benzoic acid. Front. Bioeng. Biotechnol. 4, 90.
- Yugi, K., Kubota, H., Hatano, A., Kuroda, S., 2016. Trans-Omics: How To Reconstruct Biochemical Networks Across Multiple “Omic” Layers. Trends Biotechnol. 34, 276–290.
- Zaharah, S.S., Singh, Z., 2012. Absciscic acid modulates mango fruit ripening, in: Acta Horticulturae. International Society for Horticultural Science, pp. 913–920.
- Zaharah, S.S., Singh, Z., Symons, G.M., Reid, J.B., 2012. Role of Brassinosteroids, Ethylene, Absciscic Acid, and Indole-3-Acetic Acid in Mango Fruit Ripening. J. Plant Growth Regul. 31, 363–372.
- Zecchin, A., Stapor, P.C., Goveia, J., Carmeliet, P., 2015. Metabolic pathway compartmentalization: An underappreciated opportunity? Curr. Opin. Biotechnol. 34, 73–81.
- Zengerer, V., Schmid, M., Bieri, M., Müller, D.C., Remus-Emsermann, M.N.P., Ahrens, C.H., Pelludat, C., 2018. *Pseudomonas orientalis* F9: A potent antagonist against phytopathogens with phytotoxic effect in the apple flower. Front. Microbiol. 9.
- Zhang, A., Sun, H., Yan, G., Wang, P., Wang, X., 2016. Mass spectrometry-based metabolomics: Applications to biomarker and metabolic pathway research. Biomed. Chromatogr. 30, 7–12.
- Zhang, Z.H., Jhaveri, D.J., Marshall, V.M., Bauer, D.C., Edson, J., Narayanan, R.K., Robinson, G.J., Lundberg, A.E., Bartlett, P.F., Wray, N.R., Zhao, Q.Y., 2014. A comparative study of techniques for differential expression analysis on RNA-seq data. PLoS One 9, e103207.
- Zhou, S., Lama, S., Sankaranarayanan, M., Park, S., 2019. Metabolic engineering of *Pseudomonas denitrificans* for the 1,3-propanediol production from glycerol. Bioresour. Technol. 292.
- Zhu, J., Sova, P., Xu, Q., Dombek, K.M., Xu, E.Y., Vu, H., Tu, Z., Brem, R.B., Bumgarner, R.E., Schadt, E.E., 2012. Stitching together multiple data dimensions reveals interacting metabolomic and transcriptomic networks that modulate cell regulation. PLoS Biol. 10, 1001301.
- Zur, H., Rupp, E., Shlomi, T., 2010. iMAT: An integrative metabolic analysis tool. Bioinformatics 26, 3140–3142.

Appendices

Appendix A: Database schema of MAPPs



Appendix B: List of ubiquitous metabolites ignored by default during pathway prediction in MAPPS

Metabolite ID	Metabolite Name
C00001	H ₂ O
C00002	ATP
C00003	NAD ⁺
C00004	NADH
C00005	NADPH
C00006	NADP ⁺
C00007	Oxygen
C00008	ADP
C00009	Orthophosphate
C00010	CoA
C00011	CO ₂
C00012	Peptide
C00013	Diphosphate
C00015	UDP
C00016	FAD
C00017	Protein
C00018	Pyridoxal phosphate
C00019	S-Adenosyl-L-methionine
C00020	AMP
C00027	Hydrogen peroxide
C00034	Manganese
C00035	GDP
C00044	GTP
C00053	3'-Phosphoadenylyl sulfate
C00054	Adenosine 3',5'-bisphosphate
C00055	CMP
C00059	Sulfate
C00063	CTP
C00066	tRNA
C00075	UTP
C00080	H ⁺
C00081	ITP
C00104	IDP
C00130	IMP
C00131	dATP
C00206	dADP
C00237	CO
C00286	dGTP
C00320	Thiosulfate
C00857	Deamino-NAD ⁺
C01322	RX
C01342	NH ₄ (⁺)
C01352	FADH ₂
C01528	Hydrogen selenide

Continued from previous page	
C02987	L-Glutamyl-tRNA(Glu)
C04432	tRNA containing 6-isopentenyladenosine
C05336	Selenomethionyl-tRNA(Met)
C05684	Selenite
C05697	Selenate
C06481	L-Seryl-tRNA(Sec)
C06482	L-Selenocysteinyl-tRNA(Sec)
C09306	Sulfur dioxide
C14818	Fe ²⁺
C14819	Fe ³⁺

Appendix C: List of conserved reactions related to carbohydrate and amino acid metabolism

Reaction Id	Reaction Definition	Type
R00014	Pyruvate + Thiamin diphosphate \rightleftharpoons 2-(alpha-Hydroxyethyl)thiamine diphosphate + CO ₂	conserved
R00177	Orthophosphate + Diphosphate + S-Adenosyl-L-methionine \rightleftharpoons ATP + L-Methionine + H ₂ O	conserved
R00200	ATP + Pyruvate \rightleftharpoons ADP + Phosphoenolpyruvate	conserved
R00226	(S)-2-Acetolactate + CO ₂ \rightleftharpoons 2 Pyruvate	conserved
R00238	2 Acetyl-CoA \rightleftharpoons CoA + Acetoacetyl-CoA	conserved
R00245	L-Glutamate 5-semialdehyde + NAD ⁺ + H ₂ O \rightleftharpoons L-Glutamate + NADH + H ⁺	conserved
R00253	ATP + L-Glutamate + Ammonia \rightleftharpoons ADP + Orthophosphate + L-Glutamine	conserved
R00256	L-Glutamine + H ₂ O \rightleftharpoons L-Glutamate + Ammonia	conserved
R00268	Oxalosuccinate \rightleftharpoons 2-Oxoglutarate + CO ₂	conserved
R00344	ATP + Pyruvate + HCO ₃ ⁻ \rightleftharpoons ADP + Orthophosphate + Oxaloacetate	conserved
R00351	Citrate + CoA \rightleftharpoons Acetyl-CoA + H ₂ O + Oxaloacetate	conserved
R00405	ATP + Succinate + CoA \rightleftharpoons ADP + Orthophosphate + Succinyl-CoA	conserved
R00416	UTP + N-Acetyl-alpha-D-glucosamine 1-phosphate \rightleftharpoons Diphosphate + UDP-N-acetyl-alpha-D-glucosamine	conserved
R00451	meso-2,6-Diaminoheptanedioate \rightleftharpoons L-Lysine + CO ₂	conserved
R00480	ATP + L-Aspartate \rightleftharpoons ADP + 4-Phospho-L-aspartate	conserved
R00575	2 ATP + L-Glutamine + HCO ₃ ⁻ + H ₂ O \rightleftharpoons 2 ADP + Orthophosphate + L-Glutamate + Carbamoyl phosphate	conserved
R00586	L-Serine + Acetyl-CoA \rightleftharpoons O-Acetyl-L-serine + CoA	conserved
R00621	2-Oxoglutarate + Thiamin diphosphate \rightleftharpoons 3-Carboxy-1-hydroxypropyl-ThPP + CO ₂	conserved
R00658	2-Phospho-D-glycerate \rightleftharpoons Phosphoenolpyruvate + H ₂ O	conserved
R00660	Phosphoenolpyruvate + UDP-N-acetyl-alpha-D-glucosamine \rightleftharpoons UDP-N-acetyl-3-(1-carboxyvinyl)-D-glucosamine + Orthophosphate	conserved
R00674	L-Serine + Indole \rightleftharpoons L-Tryptophan + H ₂ O	conserved
R00691	L-Arogenate \rightleftharpoons L-Phenylalanine + H ₂ O + CO ₂	conserved
R00694	L-Phenylalanine + 2-Oxoglutarate \rightleftharpoons Phenylpyruvate + L-Glutamate	conserved
R00707	(S)-1-Pyrroline-5-carboxylate + NAD ⁺ + 2 H ₂ O \rightleftharpoons L-Glutamate + NADH + H ⁺	conserved
R00708	(S)-1-Pyrroline-5-carboxylate + NADP ⁺ + 2 H ₂ O \rightleftharpoons L-Glutamate + NADPH + H ⁺	conserved
R00713	Succinate semialdehyde + NAD ⁺ + H ₂ O \rightleftharpoons Succinate + NADH + H ⁺	conserved
R00714	Succinate semialdehyde + NADP ⁺ + H ₂ O \rightleftharpoons Succinate + NADPH + H ⁺	conserved
R00734	L-Tyrosine + 2-Oxoglutarate \rightleftharpoons 3-(4-Hydroxyphenyl)pyruvate + L-Glutamate	conserved
R00742	ATP + Acetyl-CoA + HCO ₃ ⁻ \rightleftharpoons ADP + Orthophosphate + Malonyl-CoA	conserved
R00768	L-Glutamine + D-Fructose 6-phosphate \rightleftharpoons L-Glutamate + D-Glucosamine 6-phosphate	conserved
R00945	5,10-Methylenetetrahydrofolate + Glycine + H ₂ O \rightleftharpoons Tetrahydrofolate + L-Serine	conserved
R00959	D-Glucose 1-phosphate \rightleftharpoons alpha-D-Glucose 6-phosphate	conserved
R01015	D-Glyceraldehyde 3-phosphate \rightleftharpoons Glycerone phosphate	conserved
R01056	D-Ribose 5-phosphate \rightleftharpoons D-Ribulose 5-phosphate	conserved

Continued from previous page

R01057	alpha-D-Ribose 1-phosphate <=> D-Ribose 5-phosphate	conserved
R01061	D-Glyceraldehyde 3-phosphate + Orthophosphate + NAD+ <=> 3-Phospho-D-glyceroyl phosphate + NADH + H+	conserved
R01071	1-(5-Phospho-D-ribosyl)-ATP + Diphosphate <=> ATP + 5-Phospho-alpha-D-ribose 1-diphosphate	conserved
R01073	N-(5-Phospho-D-ribosyl)anthranilate + Diphosphate <=> Anthranilate + 5-Phospho-alpha-D-ribose 1-diphosphate	conserved
R01082	(S)-Malate <=> Fumarate + H2O	conserved
R01135	GTP + IMP + L-Aspartate <=> GDP + Orthophosphate + N6-(1,2-Dicarboxyethyl)-AMP	conserved
R01163	L-Histidinal + H2O + NAD+ <=> L-Histidine + NADH + H+	conserved
R01185	Inositol 1-phosphate + H2O <=> myo-Inositol + Orthophosphate	conserved
R01186	myo-Inositol 4-phosphate + H2O <=> myo-Inositol + Orthophosphate	conserved
R01187	1D-myo-Inositol 3-phosphate + H2O <=> myo-Inositol + Orthophosphate	conserved
R01221	Glycine + Tetrahydrofolate + NAD+ <=> 5,10-Methylenetetrahydrofolate + Ammonia + CO2 + NADH + H+	conserved
R01248	L-Proline + NAD+ <=> (S)-1-Pyrroline-5-carboxylate + NADH + H+	conserved
R01251	L-Proline + NADP+ <=> (S)-1-Pyrroline-5-carboxylate + NADPH + H+	conserved
R01253	L-Proline + Quinone <=> (S)-1-Pyrroline-5-carboxylate + Hydroquinone	conserved
R01325	Citrate <=> cis-Aconitate + H2O	conserved
R01373	Prephenate <=> Phenylpyruvate + H2O + CO2	conserved
R01397	Carbamoyl phosphate + L-Aspartate <=> Orthophosphate + N-Carbamoyl-L-aspartate	conserved
R01466	O-Phospho-L-homoserine + H2O <=> L-Threonine + Orthophosphate	conserved
R01518	2-Phospho-D-glycerate <=> 3-Phospho-D-glycerate	conserved
R01529	D-Ribulose 5-phosphate <=> D-Xylulose 5-phosphate	conserved
R01715	Chorismate <=> Prephenate	conserved
R01731	Oxaloacetate + L-Arogenate <=> L-Aspartate + Prephenate	conserved
R01736	(R)-S-Lactoylglutathione + H2O <=> Glutathione + (R)-Lactate	conserved
R01771	ATP + L-Homoserine <=> ADP + O-Phospho-L-homoserine	conserved
R01773	L-Homoserine + NAD+ <=> L-Aspartate 4-semialdehyde + NADH + H+	conserved
R01775	L-Homoserine + NADP+ <=> L-Aspartate 4-semialdehyde + NADPH + H+	conserved
R01776	Acetyl-CoA + L-Homoserine <=> CoA + O-Acetyl-L-homoserine	conserved
R01826	Phosphoenolpyruvate + D-Erythrose 4-phosphate + H2O <=> 2-Dehydro-3-deoxy-D-arabino-heptonate 7-phosphate + Orthophosphate	conserved
R01899	Isocitrate + NADP+ <=> Oxalosuccinate + NADPH + H+	conserved
R01900	Isocitrate <=> cis-Aconitate + H2O	conserved
R01933	2-Oxoadipate + CoA + NAD+ <=> Glutaryl-CoA + CO2 + NADH + H+	conserved
R01975	(S)-3-Hydroxybutanoyl-CoA + NAD+ <=> Acetoacetyl-CoA + NADH + H+	conserved
R02164	Quinone + Succinate <=> Hydroquinone + Fumarate	conserved
R02291	L-Aspartate 4-semialdehyde + Orthophosphate + NADP+ <=> 4-Phospho-L-aspartate + NADPH + H+	conserved
R02320	Nucleoside triphosphate + Pyruvate <=> NDP + Phosphoenolpyruvate	conserved
R02340	Indoleglycerol phosphate <=> Indole + D-Glyceraldehyde 3-phosphate	conserved
R02401	5-Oxopentanoate + NAD+ + H2O <=> Glutarate + NADH + H+	conserved
R02404	ATP + Itaconate + CoA <=> ADP + Orthophosphate + Itaconyl-CoA	conserved
R02412	ATP + Shikimate <=> ADP + Shikimate 3-phosphate	conserved

Continued from previous page

R02413	Shikimate + NADP+ \rightleftharpoons 3-Dehydroshikimate + NADPH + H+	conserved
R02570	Succinyl-CoA + Enzyme N6-(dihydrolipoyl)lysine \rightleftharpoons CoA + [Dihydrolipoyllysine-residue succinyltransferase] S-succinyl-dihydrolipoyllysine	conserved
R02722	L-Serine + Indoleglycerol phosphate \rightleftharpoons L-Tryptophan + D-Glyceraldehyde 3-phosphate + H ₂ O	conserved
R02734	N-Succinyl-LL-2,6-diaminoheptanedioate + H ₂ O \rightleftharpoons Succinate + LL-2,6-Diaminoheptanedioate	conserved
R02735	LL-2,6-Diaminoheptanedioate \rightleftharpoons meso-2,6-Diaminoheptanedioate	conserved
R02739	alpha-D-Glucose 6-phosphate \rightleftharpoons beta-D-Glucose 6-phosphate	conserved
R02740	alpha-D-Glucose 6-phosphate \rightleftharpoons beta-D-Fructose 6-phosphate	conserved
R02788	ATP + UDP-N-acetylmuramoyl-L-alanyl-D-glutamate + meso-2,6-Diaminoheptanedioate \rightleftharpoons ADP + Orthophosphate + UDP-N-acetylmuramoyl-L-alanyl-gamma-D-glutamyl-meso-2,6-diaminopimelate	conserved
R03012	L-Histidinol + NAD+ \rightleftharpoons L-Histidinal + NADH + H+	conserved
R03026	(S)-3-Hydroxybutanoyl-CoA \rightleftharpoons Crotonoyl-CoA + H ₂ O	conserved
R03045	3-Hydroxypropionyl-CoA \rightleftharpoons Propenoyl-CoA + H ₂ O	conserved
R03083	2-Dehydro-3-deoxy-D-arabino-heptonate 7-phosphate \rightleftharpoons 3-Dehydroquininate + Orthophosphate	conserved
R03084	3-Dehydroquininate \rightleftharpoons 3-Dehydroshikimate + H ₂ O	conserved
R03105	Mercaptopyruvate + Thioredoxin \rightleftharpoons Hydrogen sulfide + Pyruvate + Thioredoxin disulfide	conserved
R03132	O-Acetyl-L-serine + Thiosulfate \rightleftharpoons S-Sulfo-L-cysteine + Acetate	conserved
R03191	UDP-N-acetylmuramate + NAD+ \rightleftharpoons UDP-N-acetyl-3-(1-carboxyvinyl)-D-glucosamine + NADH + H+	conserved
R03192	UDP-N-acetylmuramate + NADP+ \rightleftharpoons UDP-N-acetyl-3-(1-carboxyvinyl)-D-glucosamine + NADPH + H+	conserved
R03243	L-Histidinol phosphate + 2-Oxoglutarate \rightleftharpoons 3-(Imidazol-4-yl)-2-oxopropyl phosphate + L-Glutamate	conserved
R03270	2-(alpha-Hydroxyethyl)thiamine diphosphate + Enzyme N6-(lipoyl)lysine \rightleftharpoons [Dihydrolipoyllysine-residue acetyltransferase] S-acetyl-dihydrolipoyllysine + Thiamin diphosphate	conserved
R03276	(S)-3-Hydroxybutanoyl-CoA \rightleftharpoons (R)-3-Hydroxybutanoyl-CoA	conserved
R03291	Hydroxyproline + NAD+ \rightleftharpoons L-1-Pyrroline-3-hydroxy-5-carboxylate + NADH + H+	conserved
R03293	Hydroxyproline + NADP+ \rightleftharpoons L-1-Pyrroline-3-hydroxy-5-carboxylate + NADPH + H+	conserved
R03316	3-Carboxy-1-hydroxypropyl-ThPP + Enzyme N6-(lipoyl)lysine \rightleftharpoons [Dihydrolipoyllysine-residue succinyltransferase] S-succinyl-dihydrolipoyllysine + Thiamin diphosphate	conserved
R03321	beta-D-Glucose 6-phosphate \rightleftharpoons beta-D-Fructose 6-phosphate	conserved
R03457	D-erythro-1-(Imidazol-4-yl)glycerol 3-phosphate \rightleftharpoons 3-(Imidazol-4-yl)-2-oxopropyl phosphate + H ₂ O	conserved
R03508	1-(2-Carboxyphenylamino)-1-deoxy-D-ribulose 5-phosphate \rightleftharpoons Indoleglycerol phosphate + CO ₂ + H ₂ O	conserved
R03509	N-(5-Phospho-D-ribosyl)anthranilate \rightleftharpoons 1-(2-Carboxyphenylamino)-1-deoxy-D-ribulose 5-phosphate	conserved
R03815	Dihydrolipoylprotein + NAD+ \rightleftharpoons Lipoylprotein + NADH + H+	conserved
R04035	1-(5-Phospho-D-ribosyl)-ATP + H ₂ O \rightleftharpoons Phosphoribosyl-AMP + Diphosphate	conserved
R04037	Phosphoribosyl-AMP + H ₂ O \rightleftharpoons 5-(5-Phospho-D-ribosylaminoformimino)-1-(5-phosphoribosyl)-imidazole-4-carboxamide	conserved
R04137	3-Hydroxyisovaleryl-CoA \rightleftharpoons 3-Methylcrotonyl-CoA + H ₂ O	conserved

Continued from previous page		
R04173	O-Phospho-L-serine + 2-Oxoglutarate \rightleftharpoons 3-Phosphonooxypyruvate + L-Glutamate	conserved
R04203	(2S,3S)-3-Hydroxy-2-methylbutanoyl-CoA + NAD ⁺ \rightleftharpoons 2-Methylacetoacetyl-CoA + NADH + H ⁺	conserved
R04204	(2S,3S)-3-Hydroxy-2-methylbutanoyl-CoA \rightleftharpoons 2-Methylbut-2-enoyl-CoA + H ₂ O	conserved
R04224	2-Methylprop-2-enoyl-CoA + H ₂ O \rightleftharpoons (S)-3-Hydroxyisobutyryl-CoA	conserved
R04365	Succinyl-CoA + 2,3,4,5-Tetrahydrodipicolinate + H ₂ O \rightleftharpoons CoA + N-Succinyl-2-L-amino-6-oxoheptanedioate	conserved
R04405	5-Methyltetrahydropteroyltri-L-glutamate + L-Homocysteine \rightleftharpoons Tetrahydropteroyltri-L-glutamate + L-Methionine	conserved
R04444	L-1-Pyrroline-3-hydroxy-5-carboxylate + NAD ⁺ + 2 H ₂ O \rightleftharpoons L-erythro-4-Hydroxyglutamate + NADH + H ⁺	conserved
R04445	L-1-Pyrroline-3-hydroxy-5-carboxylate + NADP ⁺ + 2 H ₂ O \rightleftharpoons L-erythro-4-Hydroxyglutamate + NADPH + H ⁺	conserved
R04640	5-(5-Phospho-D-ribosylaminoformimino)-1-(5-phosphoribosyl)-imidazole-4-carboxamide \rightleftharpoons N-(5'-Phospho-D-1'-ribulosylformimino)-5-amino-1-(5"-phospho-D-ribosyl)-4-imidazolecarboxamide	conserved
R04780	beta-D-Fructose 1,6-bisphosphate + H ₂ O \rightleftharpoons beta-D-Fructose 6-phosphate + Orthophosphate	conserved
R05051	L-erythro-4-Hydroxyglutamate + NADH + H ⁺ \rightleftharpoons L-4-Hydroxyglutamate semialdehyde + NAD ⁺ + H ₂ O	conserved
R05066	(S)-3-Hydroxyisobutyrate + NAD ⁺ \rightleftharpoons (S)-Methylmalonate semialdehyde + NADH + H ⁺	conserved
R05332	Acetyl-CoA + alpha-D-Glucosamine 1-phosphate \rightleftharpoons CoA + N-Acetyl-alpha-D-glucosamine 1-phosphate	conserved
R06941	(3S)-3-Hydroxyadipyl-CoA + NAD ⁺ \rightleftharpoons 3-Oxoadipyl-CoA + NADH + H ⁺	conserved
R06942	5-Carboxy-2-pentenoyl-CoA + H ₂ O \rightleftharpoons (3S)-3-Hydroxyadipyl-CoA	conserved
R07396	4-Methylthio-2-oxobutanoic acid + L-Glutamate \rightleftharpoons L-Methionine + 2-Oxoglutarate	conserved
R07618	Enzyme N6-(dihydrolipoyl)lysine + NAD ⁺ \rightleftharpoons Enzyme N6-(lipoyl)lysine + NADH + H ⁺	conserved
R08648	Pyruvate + 2-Oxobutanoate \rightleftharpoons (S)-2-Aceto-2-hydroxybutanoate + CO ₂	conserved
R10147	L-Aspartate 4-semialdehyde + Pyruvate \rightleftharpoons (2S,4S)-4-Hydroxy-2,3,4,5-tetrahydrodipicolinate + H ₂ O	conserved
R10993	ATP + L-Glutamate + (S)-2-Aminobutanoate \rightleftharpoons ADP + Orthophosphate + gamma-L-Glutamyl-L-2-aminobutyrate	conserved
R10994	ATP + gamma-L-Glutamyl-L-2-aminobutyrate + Glycine \rightleftharpoons ADP + Orthophosphate + Ophthalmate	conserved

Appendix D: List of variable reactions of carbohydrate and amino acid metabolism

Reaction Id	Reaction Equation	Type	No. of Organisms
R00114	2 L-Glutamate + NADP+ \rightleftharpoons L-Glutamine + 2-Oxoglutarate + NADPH + H+	variable	110
R00135	Peptide + H ₂ O \rightleftharpoons L-Proline + Peptide	variable	110
R00192	S-Adenosyl-L-homocysteine + H ₂ O \rightleftharpoons Adenosine + L-Homocysteine	variable	110
R00220	L-Serine \rightleftharpoons Pyruvate + Ammonia	variable	110
R00235	ATP + Acetate + CoA \rightleftharpoons AMP + Diphosphate + Acetyl-CoA	variable	110
R00236	Acetyl adenylate + CoA \rightleftharpoons AMP + Acetyl-CoA	variable	110
R00248	L-Glutamate + NADP+ + H ₂ O \rightleftharpoons 2-Oxoglutarate + Ammonia + NADPH + H+	variable	110
R00258	L-Alanine + 2-Oxoglutarate \rightleftharpoons Pyruvate + L-Glutamate	variable	110
R00259	Acetyl-CoA + L-Glutamate \rightleftharpoons CoA + N-Acetyl-L-glutamate	variable	110
R00287	UDP-glucose + H ₂ O \rightleftharpoons UMP + D-Glucose 1-phosphate	variable	110
R00289	UTP + D-Glucose 1-phosphate \rightleftharpoons Diphosphate + UDP-glucose	variable	110
R00316	ATP + Acetate \rightleftharpoons Diphosphate + Acetyl adenylate	variable	110
R00345	Orthophosphate + Oxaloacetate \rightleftharpoons H ₂ O + Phosphoenolpyruvate + CO ₂	variable	110
R00357	L-Aspartate + H ₂ O + Oxygen \rightleftharpoons Oxaloacetate + Ammonia + Hydrogen peroxide	variable	110
R00361	(S)-Malate + Quinone \rightleftharpoons Oxaloacetate + Hydroquinone	variable	110
R00472	(S)-Malate + CoA \rightleftharpoons Acetyl-CoA + H ₂ O + Glyoxylate	variable	110
R00479	Isocitrate \rightleftharpoons Succinate + Glyoxylate	variable	110
R00485	L-Asparagine + H ₂ O \rightleftharpoons L-Aspartate + Ammonia	variable	110
R00490	L-Aspartate \rightleftharpoons Fumarate + Ammonia	variable	110
R00566	L-Arginine \rightleftharpoons Agmatine + CO ₂	variable	110
R00582	O-Phospho-L-serine + H ₂ O \rightleftharpoons L-Serine + Orthophosphate	variable	110
R00590	L-Serine \rightleftharpoons Dehydroalanine + H ₂ O	variable	110
R00669	N-Acetylornithine + H ₂ O \rightleftharpoons Acetate + L-Ornithine	variable	110
R00670	L-Ornithine \rightleftharpoons Putrescine + CO ₂	variable	110
R00705	3-Oxopropanoate + CoA + NAD+ \rightleftharpoons Acetyl-CoA + CO ₂ + NADH + H+	variable	110
R00706	3-Oxopropanoate + CoA + NADP+ \rightleftharpoons Acetyl-CoA + CO ₂ + NADPH + H+	variable	110
R00711	Acetaldehyde + NADP+ + H ₂ O \rightleftharpoons Acetate + NADPH + H+	variable	110
R00751	L-Threonine \rightleftharpoons Glycine + Acetaldehyde	variable	110
R00754	Ethanol + NAD+ \rightleftharpoons Acetaldehyde + NADH + H+	variable	110
R00829	Succinyl-CoA + Acetyl-CoA \rightleftharpoons CoA + 3-Oxoadipyl-CoA	variable	110
R00885	GTP + D-Mannose 1-phosphate \rightleftharpoons Diphosphate + GDP-mannose	variable	110
R00897	O-Acetyl-L-serine + Hydrogen sulfide \rightleftharpoons L-Cysteine + Acetate	variable	110
R00907	L-Alanine + 3-Oxopropanoate \rightleftharpoons Pyruvate + beta-Alanine	variable	110

Continued from previous page			
R00926	Propionyladenylate + CoA \rightleftharpoons AMP + Propanoyl-CoA	variable	110
R00931	2-Methylcitrate + CoA \rightleftharpoons Propanoyl-CoA + Oxaloacetate + H ₂ O	variable	110
R00935	(S)-Methylmalonate semialdehyde + CoA + NAD ⁺ \rightleftharpoons Propanoyl-CoA + CO ₂ + NADH + H ⁺	variable	110
R00944	10-Formyltetrahydrofolate + H ₂ O \rightleftharpoons Formate + Tetrahydrofolate	variable	110
R00985	Chorismate + Ammonia \rightleftharpoons Anthranilate + Pyruvate + H ₂ O	variable	110
R00986	Chorismate + L-Glutamine \rightleftharpoons Anthranilate + Pyruvate + L-Glutamate	variable	110
R00994	2-Oxobutanoate + CO ₂ + NADH + H ⁺ \rightleftharpoons D-erythro-3-Methylmalate + NAD ⁺	variable	110
R00996	L-Threonine \rightleftharpoons 2-Oxobutanoate + Ammonia	variable	110
R01049	ATP + D-Ribose 5-phosphate \rightleftharpoons AMP + 5-Phospho-alpha-D-ribose 1-diphosphate	variable	110
R01070	beta-D-Fructose 1,6-bisphosphate \rightleftharpoons Glycerone phosphate + D-Glyceraldehyde 3-phosphate	variable	110
R01083	N6-(1,2-Dicarboxyethyl)-AMP \rightleftharpoons Fumarate + AMP	variable	110
R01086	N-(L-Arginino)succinate \rightleftharpoons Fumarate + L-Arginine	variable	110
R01090	L-Leucine + 2-Oxoglutarate \rightleftharpoons 4-Methyl-2-oxopentanoate + L-Glutamate	variable	110
R01213	alpha-Isopropylmalate + CoA \rightleftharpoons Acetyl-CoA + 3-Methyl-2-oxobutanoic acid + H ₂ O	variable	110
R01214	L-Valine + 2-Oxoglutarate \rightleftharpoons 3-Methyl-2-oxobutanoic acid + L-Glutamate	variable	110
R01215	L-Valine + Pyruvate \rightleftharpoons 3-Methyl-2-oxobutanoic acid + L-Alanine	variable	110
R01257	(S)-Malate + FAD \rightleftharpoons Oxaloacetate + FADH ₂	variable	110
R01334	2-Phosphoglycolate + H ₂ O \rightleftharpoons Glycolate + Orthophosphate	variable	110
R01354	ATP + Propanoate \rightleftharpoons Diphosphate + Propionyladenylate	variable	110
R01372	Phenylpyruvate + Oxygen \rightleftharpoons 2-Hydroxyphenylacetate + CO ₂	variable	110
R01374	D-Phenylalanine + H ₂ O + Quinone \rightleftharpoons Phenylpyruvate + Ammonia + Hydroquinone	variable	110
R01388	D-Glycerate + NAD ⁺ \rightleftharpoons Hydroxypyruvate + NADH + H ⁺	variable	110
R01398	Carbamoyl phosphate + L-Ornithine \rightleftharpoons Orthophosphate + L-Citrulline	variable	110
R01512	ATP + 3-Phospho-D-glycerate \rightleftharpoons ADP + 3-Phospho-D-glyceroyl phosphate	variable	110
R01513	3-Phospho-D-glycerate + NAD ⁺ \rightleftharpoons 3-Phosphonooxypyruvate + NADH + H ⁺	variable	110
R01641	Sedoheptulose 7-phosphate + D-Glyceraldehyde 3-phosphate \rightleftharpoons D-Ribose 5-phosphate + D-Xylulose 5-phosphate	variable	110
R01800	CDP-diacylglycerol + L-Serine \rightleftharpoons CMP + Phosphatidylserine	variable	110
R01818	D-Mannose 6-phosphate \rightleftharpoons D-Mannose 1-phosphate	variable	110
R01827	Sedoheptulose 7-phosphate + D-Glyceraldehyde 3-phosphate \rightleftharpoons D-Erythrose 4-phosphate + beta-D-Fructose 6-phosphate	variable	110

Continued from previous page

R01830	beta-D-Fructose 6-phosphate + D-Glyceraldehyde 3-phosphate \rightleftharpoons D-Erythrose 4-phosphate + D-Xylulose 5-phosphate	variable	110
R01954	ATP + L-Citrulline + L-Aspartate \rightleftharpoons AMP + Diphosphate + N-(L-Arginino)succinate	variable	110
R02035	D-Glucono-1,5-lactone 6-phosphate + H ₂ O \rightleftharpoons 6-Phospho-D-gluconate	variable	110
R02060	alpha-D-Glucosamine 1-phosphate \rightleftharpoons D-Glucosamine 6-phosphate	variable	110
R02199	L-Isoleucine + 2-Oxoglutarate \rightleftharpoons (S)-3-Methyl-2-oxopentanoic acid + L-Glutamate	variable	110
R02282	N-Acetylornithine + L-Glutamate \rightleftharpoons L-Ornithine + N-Acetyl-L-glutamate	variable	110
R02521	3-(4-Hydroxyphenyl)pyruvate + Oxygen \rightleftharpoons Homogentisate + CO ₂	variable	110
R02530	(R)-S-Lactoylglutathione \rightleftharpoons Glutathione + Methylglyoxal	variable	110
R02549	4-Aminobutyraldehyde + NAD ⁺ + H ₂ O \rightleftharpoons 4-Aminobutanoate + NADH + H ⁺	variable	110
R02568	D-Fructose 1-phosphate \rightleftharpoons Glycerone phosphate + D-Glyceraldehyde	variable	110
R02569	Acetyl-CoA + Enzyme N6-(dihydrolipoyl)lysine \rightleftharpoons CoA + [Dihydrolipoyllysine-residue acetyltransferase] S-acetyldihydrolipoyllysine	variable	110
R02649	ATP + N-Acetyl-L-glutamate \rightleftharpoons ADP + N-Acetyl-L-glutamate 5-phosphate	variable	110
R02670	2 3-Hydroxyanthranilate + 4 Oxygen \rightleftharpoons Cinnavalinate + 2 O ₂ + 2 Hydrogen peroxide + 2 H ⁺	variable	110
R02736	beta-D-Glucose 6-phosphate + NADP ⁺ \rightleftharpoons D-Glucono-1,5-lactone 6-phosphate + NADPH + H ⁺	variable	110
R03177	4-Guanidinobutanal + NAD ⁺ + H ₂ O \rightleftharpoons 4-Guanidinobutanoate + NADH + H ⁺	variable	110
R03313	L-Glutamate 5-semialdehyde + Orthophosphate + NADP ⁺ \rightleftharpoons L-Glutamyl 5-phosphate + NADPH + H ⁺	variable	110
R03391	CDP-4-dehydro-3,6-dideoxy-D-glucose + NAD ⁺ + H ₂ O \rightleftharpoons CDP-4-dehydro-6-deoxy-D-glucose + NADH + H ⁺	variable	110
R03392	CDP-4-dehydro-3,6-dideoxy-D-glucose + NADP ⁺ + H ₂ O \rightleftharpoons CDP-4-dehydro-6-deoxy-D-glucose + NADPH + H ⁺	variable	110
R03443	N-Acetyl-L-glutamate 5-semialdehyde + Orthophosphate + NADP ⁺ \rightleftharpoons N-Acetyl-L-glutamate 5-phosphate + NADPH + H ⁺	variable	110
R03896	(R)-2-Methylmalate \rightleftharpoons 2-Methylmaleate + H ₂ O	variable	110
R03898	2-Methylmaleate + H ₂ O \rightleftharpoons D-erythro-3-Methylmalate	variable	110
R03968	alpha-Isopropylmalate \rightleftharpoons 2-Isopropylmaleate + H ₂ O	variable	110
R04001	(2R,3S)-3-Isopropylmalate \rightleftharpoons 2-Isopropylmaleate + H ₂ O	variable	110
R04095	3-Methylbutanoyl-CoA + FAD \rightleftharpoons 3-Methylcrotonyl-CoA + FADH ₂	variable	110
R04125	S-Aminomethyldihydrolipoylprotein + Tetrahydrofolate \rightleftharpoons Dihydrolipoylprotein + 5,10-Methylenetetrahydrofolate + Ammonia	variable	110
R04187	L-Alanine + (S)-Methylmalonate semialdehyde \rightleftharpoons Pyruvate + L-3-Aminoisobutanoate	variable	110
R04198	2,3,4,5-Tetrahydrodipicolinate + NAD ⁺ + H ₂ O \rightleftharpoons (2S,4S)-4-Hydroxy-2,3,4,5-tetrahydrodipicolinate + NADH + H ⁺	variable	110

Continued from previous page			
R04199	2,3,4,5-Tetrahydrodipicolinate + NADP+ + H ₂ O \rightleftharpoons (2S,4S)-4-Hydroxy-2,3,4,5-tetrahydrodipicolinate + NADPH + H+	variable	110
R04425	(2S,3R)-3-Hydroxybutane-1,2,3-tricarboxylate \rightleftharpoons (Z)-But-2-ene-1,2,3-tricarboxylate + H ₂ O	variable	110
R04426	(2R,3S)-3-Isopropylmalate + NAD+ \rightleftharpoons (2S)-2-Isopropyl-3-oxosuccinate + NADH + H+	variable	110
R04441	(R)-2,3-Dihydroxy-3-methylbutanoate \rightleftharpoons 3-Methyl-2-oxobutanoic acid + H ₂ O	variable	110
R04475	N-Succinyl-LL-2,6-diaminoheptanedioate + 2-Oxoglutarate \rightleftharpoons N-Succinyl-2-L-amino-6-oxoheptanedioate + L-Glutamate	variable	110
R04617	ATP + UDP-N-acetylmuramoyl-L-alanyl-gamma-D-glutamyl-meso-2,6-diaminopimelate + D-Alanyl-D-alanine \rightleftharpoons ADP + Orthophosphate + UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-6-carboxy-L-lysyl-D-alanyl-D-alanine	variable	110
R04880	3,4-Dihydroxyphenylethyleneglycol + NAD+ \rightleftharpoons 3,4-Dihydroxymandelaldehyde + NADH + H+	variable	110
R05070	(R)-2,3-Dihydroxy-3-methylpentanoate \rightleftharpoons (S)-3-Methyl-2-oxopentanoic acid + H ₂ O	variable	110
R06171	L-Allothreonine \rightleftharpoons Glycine + Acetaldehyde	variable	110
R07363	1,2-Dihydroxy-5-(methylthio)pent-1-en-3-one + Oxygen \rightleftharpoons 3-(Methylthio)propanoate + Formate + CO	variable	110
R07364	1,2-Dihydroxy-5-(methylthio)pent-1-en-3-one + Oxygen \rightleftharpoons 4-Methylthio-2-oxobutanoic acid + Formate	variable	110
R08572	D-Glycerate + ATP \rightleftharpoons 2-Phospho-D-glycerate + ADP	variable	110
R09107	N-Acetyl-L-citrulline + H ₂ O \rightleftharpoons Acetate + L-Citrulline	variable	110
R10907	beta-D-Glucose 6-phosphate + NAD+ \rightleftharpoons D-Glucono-1,5-lactone 6-phosphate + NADH + H+	variable	110
R10991	(S)-2-Aminobutanoate + 2-Oxoglutarate \rightleftharpoons 2-Oxobutanoate + L-Glutamate	variable	110
R00022	Chitobiose + H ₂ O \rightleftharpoons 2 N-Acetyl-D-glucosamine	variable	109
R00199	ATP + Pyruvate + H ₂ O \rightleftharpoons AMP + Phosphoenolpyruvate + Orthophosphate	variable	109
R00239	ATP + L-Glutamate \rightleftharpoons ADP + L-Glutamyl 5-phosphate	variable	109
R00243	L-Glutamate + NAD+ + H ₂ O \rightleftharpoons 2-Oxoglutarate + Ammonia + NADH + H+	variable	109
R00341	ATP + Oxaloacetate \rightleftharpoons ADP + Phosphoenolpyruvate + CO ₂	variable	109
R00409	(2S,3R)-3-Hydroxybutane-1,2,3-tricarboxylate \rightleftharpoons Pyruvate + Succinate	variable	109
R00946	5-Methyltetrahydrofolate + L-Homocysteine \rightleftharpoons Tetrahydrofolate + L-Methionine	variable	109
R01072	5-Phosphoribosylamine + Diphosphate + L-Glutamate \rightleftharpoons L-Glutamine + 5-Phospho-alpha-D-ribose 1-diphosphate + H ₂ O	variable	109
R01288	O-Succinyl-L-homoserine + Hydrogen sulfide \rightleftharpoons L-Homocysteine + Succinate	variable	109
R01360	(S)-3-Hydroxy-3-methylglutaryl-CoA \rightleftharpoons Acetyl-CoA + Acetoacetate	variable	109
R01416	Agmatine + H ₂ O \rightleftharpoons N-Carbamoylputrescine + Ammonia	variable	109
R01648	4-Aminobutanoate + 2-Oxoglutarate \rightleftharpoons Succinate semialdehyde + L-Glutamate	variable	109

Continued from previous page

R01714	5-O-(1-Carboxyvinyl)-3-phosphoshikimate \rightleftharpoons Chorismate + Orthophosphate	variable	109
R01819	D-Mannose 6-phosphate \rightleftharpoons beta-D-Fructose 6-phosphate	variable	109
R02036	6-Phospho-D-gluconate \rightleftharpoons 2-Dehydro-3-deoxy-6-phospho-D-gluconate + H ₂ O	variable	109
R02085	(S)-3-Hydroxy-3-methylglutaryl-CoA \rightleftharpoons 3-Methylglutaconyl-CoA + H ₂ O	variable	109
R02274	5-Aminopentanoate + 2-Oxoglutarate \rightleftharpoons 5-Oxopentanoate + L-Glutamate	variable	109
R03425	Glycine + Lipoylprotein \rightleftharpoons S-Aminomethyldihydrolipoylprotein + CO ₂	variable	109
R04138	ATP + 3-Methylcrotonyl-CoA + HCO ₃ ⁻ \rightleftharpoons ADP + Orthophosphate + 3-Methylglutaconyl-CoA	variable	109
R04420	S-Methyl-5-thio-D-ribose 1-phosphate \rightleftharpoons S-Methyl-5-thio-D-ribulose 1-phosphate	variable	109
R00009	2 Hydrogen peroxide \rightleftharpoons Oxygen + 2 H ₂ O	variable	108
R00217	Oxaloacetate \rightleftharpoons Pyruvate + CO ₂	variable	108
R00291	UDP-glucose \rightleftharpoons UDP-alpha-D-galactose	variable	108
R00470	4-Hydroxy-2-oxoglutarate \rightleftharpoons Pyruvate + Glyoxylate	variable	108
R00578	ATP + L-Aspartate + L-Glutamine + H ₂ O \rightleftharpoons AMP + Diphosphate + L-Asparagine + L-Glutamate	variable	108
R01025	Choline + Acceptor \rightleftharpoons Betaine aldehyde + Reduced acceptor	variable	108
R01394	Hydroxypyruvate \rightleftharpoons 2-Hydroxy-3-oxopropanoate	variable	108
R01737	ATP + D-Gluconic acid \rightleftharpoons ADP + 6-Phospho-D-gluconate	variable	108
R02565	Betaine aldehyde + NAD ⁺ + H ₂ O \rightleftharpoons Betaine + NADH + 2 H ⁺	variable	108
R02566	Betaine aldehyde + NADP ⁺ + H ₂ O \rightleftharpoons Betaine + NADPH + 2 H ⁺	variable	108
R02922	Creatinine + H ₂ O \rightleftharpoons N-Methylhydantoin + Ammonia	variable	108
R04440	(R)-2,3-Dihydroxy-3-methylbutanoate + NADP ⁺ \rightleftharpoons 3-Hydroxy-3-methyl-2-oxobutanoic acid + NADPH + H ⁺	variable	108
R05068	(R)-2,3-Dihydroxy-3-methylpentanoate + NADP ⁺ \rightleftharpoons (R)-3-Hydroxy-3-methyl-2-oxopentanoate + NADPH + H ⁺	variable	108
R05069	(S)-2-Aceto-2-hydroxybutanoate \rightleftharpoons (R)-3-Hydroxy-3-methyl-2-oxopentanoate	variable	108
R05071	(S)-2-Acetolactate \rightleftharpoons 3-Hydroxy-3-methyl-2-oxobutanoic acid	variable	108
R05605	2-Dehydro-3-deoxy-6-phospho-D-gluconate \rightleftharpoons D-Glyceraldehyde 3-phosphate + Pyruvate	variable	108
R08968	N-Acetyl-D-glucosamine + ATP \rightleftharpoons N-Acetyl-alpha-D-glucosamine 1-phosphate + ADP	variable	108
R11263	2-Methylcitrate \rightleftharpoons 2-Methyl-trans-aconitate + H ₂ O	variable	108
R00131	Urea + H ₂ O \rightleftharpoons CO ₂ + 2 Ammonia	variable	107
R00216	(S)-Malate + NADP ⁺ \rightleftharpoons Pyruvate + CO ₂ + NADPH + H ⁺	variable	107
R00230	Acetyl-CoA + Orthophosphate \rightleftharpoons CoA + Acetyl phosphate	variable	107
R00410	Succinyl-CoA + Acetoacetate \rightleftharpoons Succinate + Acetoacetyl-CoA	variable	107
R00411	N-Succinyl-L-glutamate + H ₂ O \rightleftharpoons L-Glutamate + Succinate	variable	107

Continued from previous page

R00921	Propanoyl-CoA + Orthophosphate \rightleftharpoons Propanoyl phosphate + CoA	variable	107
R01175	Butanoyl-CoA + FAD \rightleftharpoons FADH ₂ + Crotonoyl-CoA	variable	107
R01364	Acetoacetate + Fumarate \rightleftharpoons 4-Fumarylacetoacetate + H ₂ O	variable	107
R01600	ATP + beta-D-Glucose \rightleftharpoons ADP + beta-D-Glucose 6-phosphate	variable	107
R01786	ATP + alpha-D-Glucose \rightleftharpoons ADP + alpha-D-Glucose 6-phosphate	variable	107
R02071	ATP + D-Fructose 1-phosphate \rightleftharpoons ADP + beta-D-Fructose 1,6-bisphosphate	variable	107
R02488	Glutaryl-CoA + Electron-transferring flavoprotein \rightleftharpoons Crotonoyl-CoA + Reduced electron-transferring flavoprotein + CO ₂	variable	107
R02519	Homogentisate + Oxygen \rightleftharpoons 4-Maleylacetoacetate	variable	107
R03232	Protein N(pi)-phospho-L-histidine + D-Fructose \rightleftharpoons Protein histidine + D-Fructose 1-phosphate	variable	107
R02283	N-Acetylornithine + 2-Oxoglutarate \rightleftharpoons N-Acetyl-L-glutamate 5-semialdehyde + L-Glutamate	variable	106
R01795	Tetrahydrobiopterin + L-Phenylalanine + Oxygen \rightleftharpoons Dihydrobiopterin + L-Tyrosine + H ₂ O	variable	105
R10343	Succinyl-CoA + Acetate \rightleftharpoons Acetyl-CoA + Succinate	variable	105
R00717	Glycolate + NAD ⁺ \rightleftharpoons Glyoxylate + NADH + H ⁺	variable	104
R01168	L-Histidine \rightleftharpoons Urocanate + Ammonia	variable	104
R01557	Maltose \rightleftharpoons alpha,alpha-Trehalose	variable	104
R02110	Amylose \rightleftharpoons Starch	variable	104
R02111	Starch + Orthophosphate \rightleftharpoons Amylose + D-Glucose 1-phosphate	variable	104
R03527	beta-D-Glucoside + H ₂ O \rightleftharpoons D-Glucose + ROH	variable	104
R00264	2,5-Dioxopentanoate + NADP ⁺ + H ₂ O \rightleftharpoons 2-Oxoglutarate + NADPH + H ⁺	variable	103
R00710	Acetaldehyde + NAD ⁺ + H ₂ O \rightleftharpoons Acetate + NADH + H ⁺	variable	103
R01986	4-Aminobutyraldehyde + NADP ⁺ + H ₂ O \rightleftharpoons 4-Aminobutanoate + NADPH + H ⁺	variable	103
R02421	ADP-glucose + Amylose \rightleftharpoons ADP + Amylose	variable	103
R02540	2-Phenylacetamide + H ₂ O \rightleftharpoons Phenylacetic acid + Ammonia	variable	103
R02678	Indole-3-acetaldehyde + NAD ⁺ + H ₂ O \rightleftharpoons Indole-3-acetate + NADH + H ⁺	variable	103
R02957	D-Glucuronolactone + NAD ⁺ + 2 H ₂ O \rightleftharpoons D-Glucarate + NADH + H ⁺	variable	103
R03096	(Indol-3-yl)acetamide + H ₂ O \rightleftharpoons Indole-3-acetate + Ammonia	variable	103
R03180	4-Guanidinobutanamide + H ₂ O \rightleftharpoons 4-Guanidinobutanoate + Ammonia	variable	103
R03283	4-Trimethylammoniobutanal + NAD ⁺ + H ₂ O \rightleftharpoons 4-Trimethylammoniobutanoate + NADH + H ⁺	variable	103
R03869	(S)-Methylmalonate semialdehyde + NAD ⁺ + H ₂ O \rightleftharpoons Methylmalonate + NADH + H ⁺	variable	103
R04065	Imidazole-4-acetaldehyde + NAD ⁺ + H ₂ O \rightleftharpoons Imidazole-4-acetate + NADH + H ⁺	variable	103
R04903	5-Hydroxyindoleacetaldehyde + NAD ⁺ + H ₂ O \rightleftharpoons 5-Hydroxyindoleacetate + H ⁺ + NADH	variable	103

Continued from previous page

R05050	N4-Acetylaminobutanal + NAD ⁺ + H ₂ O \rightleftharpoons 4-Acetamidobutanoate + NADH + H ⁺	variable	103
R00286	UDP-glucose + H ₂ O + 2 NAD ⁺ \rightleftharpoons UDP-glucuronate + 2 NADH + 2 H ⁺	variable	101
R00704	(R)-Lactate + NAD ⁺ \rightleftharpoons Pyruvate + NADH + H ⁺	variable	101
R00927	Propanoyl-CoA + Acetyl-CoA \rightleftharpoons CoA + 2-Methylacetoacetyl-CoA	variable	101
R06620	D-Glucose + Ubiquinone \rightleftharpoons D-Glucono-1,5-lactone + Ubiquinol	variable	101
R06977	L-Glutamate + L-Aspartate 4-semialdehyde \rightleftharpoons 2-Oxoglutarate + L-2,4-Diaminobutanoate	variable	101
R00610	Sarcosine + H ₂ O + Oxygen \rightleftharpoons Glycine + Formaldehyde + Hydrogen peroxide	variable	100
R00832	Succinyl-CoA + L-Arginine \rightleftharpoons CoA + N ₂ -Succinyl-L-arginine	variable	100
R04000	Acetyl-CoA + 2-Hydroxyglutarate \rightleftharpoons Acetate + 2-Hydroxyglutaryl-CoA	variable	100
R04189	N ₂ -Succinyl-L-arginine + 2 H ₂ O \rightleftharpoons N ₂ -Succinyl-L-ornithine + CO ₂ + 2 Ammonia	variable	100
R05049	N-Succinyl-L-glutamate 5-semialdehyde + NAD ⁺ + H ₂ O \rightleftharpoons N-Succinyl-L-glutamate + NADH + H ⁺	variable	100
R01361	(R)-3-Hydroxybutanoate + NAD ⁺ \rightleftharpoons Acetoacetate + NADH + H ⁺	variable	99
R06836	D-Ribose 1,5-bisphosphate + ATP \rightleftharpoons 5-Phospho-alpha-D-ribose 1-diphosphate + ADP	variable	99
R01088	L-Leucine + H ₂ O + NAD ⁺ \rightleftharpoons 4-Methyl-2-oxopentanoate + Ammonia + NADH + H ⁺	variable	98
R01434	L-Valine + H ₂ O + NAD ⁺ \rightleftharpoons 3-Methyl-2-oxobutanoic acid + Ammonia + NADH + H ⁺	variable	98
R02196	L-Isoleucine + NAD ⁺ + H ₂ O \rightleftharpoons (S)-3-Methyl-2-oxopentanoic acid + Ammonia + NADH + H ⁺	variable	98
R02288	4-Imidazolone-5-propanoate + H ₂ O \rightleftharpoons N-Formimino-L-glutamate	variable	98
R02914	4-Imidazolone-5-propanoate \rightleftharpoons Urocanate + H ₂ O	variable	98
R00997	1-Aminocyclopropane-1-carboxylate + H ₂ O \rightleftharpoons 2-Oxobutanoate + Ammonia	variable	96
R03181	4-Maleylacetoacetate \rightleftharpoons 4-Fumarylacetoacetate	variable	96
R00150	ATP + Ammonia + HCO ₃ ⁻ \rightleftharpoons ADP + Carbamoyl phosphate + H ₂ O	variable	95
R00552	L-Arginine + H ₂ O \rightleftharpoons L-Citrulline + Ammonia	variable	95
R00880	GDP-mannose + H ₂ O + 2 NAD ⁺ \rightleftharpoons GDP-D-mannuronate + 2 NADH + 2 H ⁺	variable	95
R03706	(Alginate)n \rightleftharpoons Oligouronide with 4-deoxy-alpha-L-erythro-hex-4-enopyranuronosyl group	variable	95
R08692	GDP-D-mannuronate + (beta-D-Mannuronate)n \rightleftharpoons (beta-D-Mannuronate)n + GDP	variable	95
R08693	(beta-D-Mannuronate)n \rightleftharpoons (Alginate)n	variable	95
R00013	2 Glyoxylate \rightleftharpoons 2-Hydroxy-3-oxopropanoate + CO ₂	variable	94
R00093	2 L-Glutamate + NAD ⁺ \rightleftharpoons L-Glutamine + 2-Oxoglutarate + NADH + H ⁺	variable	93
R00782	L-Cysteine + H ₂ O \rightleftharpoons Hydrogen sulfide + Pyruvate + Ammonia	variable	93
R01739	D-Gluconic acid + NADP ⁺ \rightleftharpoons 2-Keto-D-gluconic acid + NADPH + H ⁺	variable	93

Continued from previous page			
R01940	2-Oxoadipate + Enzyme N6-(lipoyl)lysine \rightleftharpoons [Dihydrolipoyllysine-residue succinyltransferase] S- glutaryl-dihydrolipoyllysine + CO ₂	variable	93
R02408	L-Cystine + H ₂ O \rightleftharpoons Pyruvate + Ammonia + Thiocysteine	variable	93
R02571	Glutaryl-CoA + Enzyme N6-(dihydrolipoyl)lysine \rightleftharpoons CoA + [Dihydrolipoyllysine-residue succinyltransferase] S-glutaryl-dihydrolipoyllysine	variable	93
R02382	Tyramine + H ₂ O + Oxygen \rightleftharpoons 4- Hydroxyphenylacetaldehyde + Ammonia + Hydrogen peroxide	variable	92
R02529	Aminoacetone + H ₂ O + Oxygen \rightleftharpoons Methylglyoxal + Ammonia + Hydrogen peroxide	variable	92
R02613	Phenethylamine + Oxygen + H ₂ O \rightleftharpoons Phenylacetaldehyde + Ammonia + Hydrogen peroxide	variable	92
R04300	Dopamine + H ₂ O + Oxygen \rightleftharpoons 3,4- Dihydroxyphenylacetaldehyde + Ammonia + Hydrogen peroxide	variable	92
R02173	Tryptamine + H ₂ O + Oxygen \rightleftharpoons Indole-3-acetaldehyde + Ammonia + Hydrogen peroxide	variable	91
R02286	N-Formimino-L-glutamate + H ₂ O \rightleftharpoons N-Formyl-L- glutamate + Ammonia	variable	91
R02532	L-Noradrenaline + H ₂ O + Oxygen \rightleftharpoons 3,4- Dihydroxymandelaldehyde + Ammonia + Hydrogen peroxide	variable	91
R02908	Serotonin + H ₂ O + Oxygen \rightleftharpoons 5- Hydroxyindoleacetaldehyde + Ammonia + Hydrogen peroxide	variable	91
R02919	L-Adrenaline + H ₂ O + Oxygen \rightleftharpoons 3,4- Dihydroxymandelaldehyde + Methylamine + Hydrogen peroxide	variable	91
R04025	N-Acetylputrescine + H ₂ O + Oxygen \rightleftharpoons N4- Acetylaminobutanal + Ammonia + Hydrogen peroxide	variable	91
R04674	N-Methylhistamine + H ₂ O + Oxygen \rightleftharpoons Methylimidazole acetaldehyde + Ammonia + Hydrogen peroxide	variable	91
R04890	3-Methoxytyramine + H ₂ O + Oxygen \rightleftharpoons 3-Methoxy-4- hydroxyphenylacetaldehyde + Hydrogen peroxide + Ammonia	variable	91
R04893	L-Normetanephrine + H ₂ O + Oxygen \rightleftharpoons 3-Methoxy-4- hydroxyphenylglycolaldehyde + Ammonia + Hydrogen peroxide	variable	91
R04894	L-Metanephrine + H ₂ O + Oxygen \rightleftharpoons 3-Methoxy-4- hydroxyphenylglycolaldehyde + Hydrogen peroxide + Methylamine	variable	91
R04907	3-Hydroxykynurenamine + Oxygen \rightleftharpoons 4,8- Dihydroxyquinoline + Ammonia + Hydrogen peroxide	variable	91
R04908	5-Hydroxykynurenamine + H ₂ O + Oxygen \rightleftharpoons 4,6- Dihydroxyquinoline + Ammonia + Hydrogen peroxide + H ₂ O	variable	91
R01051	ATP + D-Ribose \rightleftharpoons ADP + D-Ribose 5-phosphate	variable	90
R01058	D-Glyceraldehyde 3-phosphate + NADP ⁺ + H ₂ O \rightleftharpoons 3- Phospho-D-glycerate + NADPH + H ⁺	variable	90
R02750	2-Deoxy-D-ribose 5-phosphate + ADP \rightleftharpoons Deoxyribose + ATP	variable	90
R01990	4-Guanidinobutanoate + H ₂ O \rightleftharpoons 4-Aminobutanoate + Urea	variable	89

Continued from previous page			
R00026	Cellobiose + H ₂ O \rightleftharpoons 2 beta-D-Glucose	variable	88
R02108	Starch + H ₂ O \rightleftharpoons Dextrin + Starch	variable	88
R00475	Glycolate + Oxygen \rightleftharpoons Glyoxylate + Hydrogen peroxide	variable	87
R02661	2-Methylpropanoyl-CoA + Acceptor \rightleftharpoons 2-Methylprop-2-enoyl-CoA + Reduced acceptor	variable	85
R03172	(S)-2-Methylbutanoyl-CoA + Acceptor \rightleftharpoons 2-Methylbut-2-enoyl-CoA + Reduced acceptor	variable	85
R04432	Electron-transferring flavoprotein + Propanoyl-CoA \rightleftharpoons Reduced electron-transferring flavoprotein + Propenoyl-CoA	variable	85
R07599	3-Methyl-2-oxobutanoic acid + Thiamin diphosphate \rightleftharpoons 2-Methyl-1-hydroxypropyl-ThPP + CO ₂	variable	85
R07600	2-Methyl-1-hydroxypropyl-ThPP + Enzyme N6-(lipoyl)lysine \rightleftharpoons [Dihydrolipoyllysine-residue (2-methylpropanoyl)transferase] S-(2-methylpropanoyl)dihydrolipoyllysine + Thiamin diphosphate	variable	85
R07601	4-Methyl-2-oxopentanoate + Thiamin diphosphate \rightleftharpoons 3-Methyl-1-hydroxybutyl-ThPP + CO ₂	variable	85
R07602	3-Methyl-1-hydroxybutyl-ThPP + Enzyme N6-(lipoyl)lysine \rightleftharpoons [Dihydrolipoyllysine-residue (2-methylpropanoyl)transferase] S-(3-methylbutanoyl)dihydrolipoyllysine + Thiamin diphosphate	variable	85
R07603	(S)-3-Methyl-2-oxopentanoic acid + Thiamin diphosphate \rightleftharpoons 2-Methyl-1-hydroxybutyl-ThPP + CO ₂	variable	85
R07604	2-Methyl-1-hydroxybutyl-ThPP + Enzyme N6-(lipoyl)lysine \rightleftharpoons [Dihydrolipoyllysine-residue (2-methylpropanoyl)transferase] S-(2-methylbutanoyl)dihydrolipoyllysine + Thiamin diphosphate	variable	85
R10996	2-Oxobutanoate + Thiamin diphosphate \rightleftharpoons 2-(alpha-Hydroxypropyl)thiamine diphosphate + CO ₂	variable	85
R10997	2-(alpha-Hydroxypropyl)thiamine diphosphate + Enzyme N6-(lipoyl)lysine \rightleftharpoons Enzyme N6-(S-propyldihydrolipoyl)lysine + Thiamin diphosphate	variable	85
R02662	2-Methylpropanoyl-CoA + Enzyme N6-(dihydrolipoyl)lysine \rightleftharpoons CoA + [Dihydrolipoyllysine-residue (2-methylpropanoyl)transferase] S-(2-methylpropanoyl)dihydrolipoyllysine	variable	84
R03145	Pyruvate + Ubiquinone + H ₂ O \rightleftharpoons Acetate + Ubiquinol + CO ₂	variable	84
R03174	(S)-2-Methylbutanoyl-CoA + Enzyme N6-(dihydrolipoyl)lysine \rightleftharpoons CoA + [Dihydrolipoyllysine-residue (2-methylpropanoyl)transferase] S-(2-methylbutanoyl)dihydrolipoyllysine	variable	84
R03296	Hydroxyproline \rightleftharpoons cis-4-Hydroxy-D-proline	variable	84
R04097	3-Methylbutanoyl-CoA + Enzyme N6-(dihydrolipoyl)lysine \rightleftharpoons CoA + [Dihydrolipoyllysine-residue (2-methylpropanoyl)transferase] S-(3-methylbutanoyl)dihydrolipoyllysine	variable	84
R10998	Propanoyl-CoA + Enzyme N6-(dihydrolipoyl)lysine \rightleftharpoons CoA + Enzyme N6-(S-propyldihydrolipoyl)lysine	variable	84
R01357	ATP + Acetoacetate + CoA \rightleftharpoons AMP + Diphosphate + Acetoacetyl-CoA	variable	82
R01976	(S)-3-Hydroxybutanoyl-CoA + NADP ⁺ \rightleftharpoons Acetoacetyl-CoA + NADPH + H ⁺	variable	82

Continued from previous page

R02285	N-Formimino-L-glutamate + H ₂ O \rightleftharpoons L-Glutamate + Formamide	variable	82
R00196	(S)-Lactate + 2 Ferricytochrome c \rightleftharpoons Pyruvate + 2 Ferrocyclochrome c + 2 H ⁺	variable	81
R00519	Formate + NAD ⁺ \rightleftharpoons H ⁺ + CO ₂ + NADH	variable	81
R00867	ATP + D-Fructose \rightleftharpoons ADP + beta-D-Fructose 6-phosphate	variable	81
R03920	ATP + beta-D-Fructose \rightleftharpoons ADP + beta-D-Fructose 6-phosphate	variable	81
R01639	ATP + D-Xylulose \rightleftharpoons ADP + D-Xylulose 5-phosphate	variable	80
R02032	6-Phospho-D-gluconate + NAD ⁺ \rightleftharpoons 6-Phospho-2-dehydro-D-gluconate + NADH + H ⁺	variable	78
R02034	6-Phospho-D-gluconate + NADP ⁺ \rightleftharpoons 6-Phospho-2-dehydro-D-gluconate + NADPH + H ⁺	variable	78
R02658	ATP + 2-Keto-D-gluconic acid \rightleftharpoons ADP + 6-Phospho-2-dehydro-D-gluconate	variable	78
R04379	5-Carboxymethyl-2-hydroxymuconate \rightleftharpoons 5-Carboxy-2-oxohept-3-enedioate	variable	78
R00671	L-Ornithine \rightleftharpoons L-Proline + Ammonia	variable	77
R00868	Mannitol + NAD ⁺ \rightleftharpoons D-Fructose + NADH + H ⁺	variable	77
R01741	D-Gluconic acid + FAD \rightleftharpoons 2-Keto-D-gluconic acid + FADH ₂	variable	76
R04424	2-Methylcitrate \rightleftharpoons (Z)-But-2-ene-1,2,3-tricarboxylate + H ₂ O	variable	76
R02262	L-Fuculose 1-phosphate \rightleftharpoons Glycerone phosphate + (S)-Lactaldehyde	variable	74
R01424	Hippurate + H ₂ O \rightleftharpoons Benzoate + Glycine	variable	73
R01745	D-Glycerate + NAD ⁺ \rightleftharpoons 2-Hydroxy-3-oxopropanoate + NADH + H ⁺	variable	73
R01747	D-Glycerate + NADP ⁺ \rightleftharpoons 2-Hydroxy-3-oxopropanoate + NADPH + H ⁺	variable	73
R02279	5-Dehydro-4-deoxy-D-glucarate \rightleftharpoons 2,5-Dioxopentanoate + H ₂ O + CO ₂	variable	71
R03460	Phosphoenolpyruvate + Shikimate 3-phosphate \rightleftharpoons Orthophosphate + 5-O-(1-Carboxyvinyl)-3-phosphoshikimate	variable	71
R05608	D-Galactarate \rightleftharpoons 5-Dehydro-4-deoxy-D-glucarate + H ₂ O	variable	71
R01001	L-Cystathionine + H ₂ O \rightleftharpoons L-Cysteine + Ammonia + 2-Oxobutanoate	variable	70
R01152	N-Carbamoylputrescine + H ₂ O \rightleftharpoons Putrescine + CO ₂ + Ammonia	variable	70
R01171	Butanoyl-CoA + NAD ⁺ \rightleftharpoons Crotonoyl-CoA + NADH + H ⁺	variable	70
R01528	6-Phospho-D-gluconate + NADP ⁺ \rightleftharpoons D-Ribulose 5-phosphate + CO ₂ + NADPH + H ⁺	variable	70
R02752	D-Glucarate \rightleftharpoons 5-Dehydro-4-deoxy-D-glucarate + H ₂ O	variable	70
R08056	D-Glucarate \rightleftharpoons 2-Dehydro-3-deoxy-D-glucarate + H ₂ O	variable	70
R10221	6-Phospho-D-gluconate + NAD ⁺ \rightleftharpoons D-Ribulose 5-phosphate + CO ₂ + NADH + H ⁺	variable	70
R00888	GDP-mannose \rightleftharpoons GDP-4-dehydro-6-deoxy-D-mannose + H ₂ O	variable	69
R02203	L-Pipecolate + NADP ⁺ \rightleftharpoons Delta1-Piperidine-2-carboxylate + NADPH + H ⁺	variable	69
R05198	Ethanol + 2 Ferricytochrome c \rightleftharpoons 2 Ferrocyclochrome c + Acetaldehyde + 2 H ⁺	variable	69

Continued from previous page

R08197	L-Arginine + Pyruvate \rightleftharpoons 5-Guanidino-2-oxopentanoate + L-Alanine	variable	68
R09837	2-(1,2-Epoxy-1,2-dihydrophenyl)acetyl-CoA \rightleftharpoons 2-Oxepin-2(3H)-ylideneacetyl-CoA	variable	68
R09839	3-Oxo-5,6-dehydrosuberil-CoA + CoA \rightleftharpoons 5-Carboxy-2-pentenoyl-CoA + Acetyl-CoA	variable	68
R09081	Carboxyspermidine \rightleftharpoons Spermidine + CO ₂	variable	67
R09082	Carboxynorspermidine \rightleftharpoons Norspermidine + CO ₂	variable	67
R00698	L-Phenylalanine + Oxygen \rightleftharpoons 2-Phenylacetamide + CO ₂	variable	66
R01647	Succinate semialdehyde + Pyruvate \rightleftharpoons 2,4-Dihydroxyhept-2-enedioate	variable	66
R00237	(3S)-Citramalyl-CoA \rightleftharpoons Acetyl-CoA + Pyruvate	variable	64
R00420	UDP-N-acetyl-alpha-D-glucosamine \rightleftharpoons UDP-N-acetyl-D-mannosamine	variable	64
R01287	O-Acetyl-L-homoserine + Hydrogen sulfide \rightleftharpoons L-Homocysteine + Acetate	variable	64
R02491	(3S)-Citramalyl-CoA \rightleftharpoons Itaconyl-CoA + H ₂ O	variable	64
R02855	(R)-Acetoin + NAD ⁺ \rightleftharpoons Diacetyl + NADH + H ⁺	variable	64
R02946	(R,R)-Butane-2,3-diol + NAD ⁺ \rightleftharpoons (R)-Acetoin + NADH + H ⁺	variable	64
R03154	Succinyl-CoA + (S)-2-Methylmalate \rightleftharpoons Succinate + (3S)-Citramalyl-CoA	variable	62
R01602	alpha-D-Glucose \rightleftharpoons beta-D-Glucose	variable	61
R10619	D-Galactose \rightleftharpoons alpha-D-Galactose	variable	61
R01873	Quinate + PQQ \rightleftharpoons PQQH ₂ + 3-Dehydroquinate	variable	60
R02415	Shikimate + PQQ \rightleftharpoons 3-Dehydroshikimate + PQQH ₂	variable	60
R02536	Phenylacetaldehyde + NAD ⁺ + H ₂ O \rightleftharpoons Phenylacetic acid + NADH + H ⁺	variable	60
R00679	L-Tryptophan + Oxygen \rightleftharpoons (Indol-3-yl)acetamide + CO ₂ + H ₂ O	variable	59
R07658	UDP-glucuronate + NAD ⁺ \rightleftharpoons UDP-L-Ara4O + CO ₂ + NADH + H ⁺	variable	59
R07660	10-Formyltetrahydrofolate + UDP-L-Ara4N \rightleftharpoons Tetrahydrofolate + UDP-L-Ara4FN	variable	59
R07661	UDP-L-Ara4FN + di-trans,poly-cis-Undecaprenyl phosphate \rightleftharpoons Undecaprenyl phosphate alpha-L-Ara4FN + UDP	variable	59
R10841	D-Galacturonate + NAD ⁺ \rightleftharpoons D-Galactaro-1,5-lactone + NADH + H ⁺	variable	59
R04858	S-Adenosyl-L-methionine + DNA cytosine \rightleftharpoons S-Adenosyl-L-homocysteine + DNA 5-methylcytosine	variable	58
R01874	D-Cysteine + H ₂ O \rightleftharpoons Hydrogen sulfide + Ammonia + Pyruvate	variable	57
R02059	N-Acetyl-D-glucosamine 6-phosphate + H ₂ O \rightleftharpoons D-Glucosamine 6-phosphate + Acetate	variable	57
R05199	Protein N(pi)-phospho-L-histidine + N-Acetyl-D-glucosamine \rightleftharpoons Protein histidine + N-Acetyl-D-glucosamine 6-phosphate	variable	57
R09668	5'-S-Methyl-5'-thioinosine + Orthophosphate \rightleftharpoons Hypoxanthine + S-Methyl-5-thio-D-ribose 1-phosphate	variable	57
R03178	5-Guanidino-2-oxopentanoate \rightleftharpoons 4-Guanidinobutanal + CO ₂	variable	55
R03303	3,4-Dihydroxyphenylacetate + Oxygen \rightleftharpoons 2-Hydroxy-5-carboxymethylmuconate semialdehyde	variable	55

Continued from previous page

R04380	5-Carboxy-2-oxohept-3-enedioate \rightleftharpoons 2-Hydroxyhepta-2,4-dienedioate + CO ₂	variable	55
R04418	2-Hydroxy-5-carboxymethylmuconate semialdehyde + NAD ⁺ + H ₂ O \rightleftharpoons 5-Carboxymethyl-2-hydroxymuconate + NADH + H ⁺	variable	55
R01920	S-Adenosylmethioninamine + Putrescine \rightleftharpoons 5'-Methylthioadenosine + Spermidine	variable	54
R02869	S-Adenosylmethioninamine + Spermidine \rightleftharpoons 5'-Methylthioadenosine + Spermine	variable	54
R00178	S-Adenosyl-L-methionine + H ⁺ \rightleftharpoons S-Adenosylmethioninamine + CO ₂	variable	52
R01520	beta-D-Glucose + NAD ⁺ \rightleftharpoons D-Glucono-1,5-lactone + NADH + H ⁺	variable	52
R01521	beta-D-Glucose + NADP ⁺ \rightleftharpoons D-Glucono-1,5-lactone + NADPH + H ⁺	variable	52
R00891	L-Serine + Hydrogen sulfide \rightleftharpoons L-Cysteine + H ₂ O	variable	49
R01290	L-Serine + L-Homocysteine \rightleftharpoons L-Cystathionine + H ₂ O	variable	49
R00465	Glycolate + NADP ⁺ \rightleftharpoons Glyoxylate + NADPH + H ⁺	variable	48
R01286	L-Cystathionine + H ₂ O \rightleftharpoons L-Homocysteine + Ammonia + Pyruvate	variable	48
R01392	D-Glycerate + NADP ⁺ \rightleftharpoons Hydroxypyruvate + NADPH + H ⁺	variable	48
R07392	S-Methyl-5-thio-D-ribulose 1-phosphate \rightleftharpoons 2,3-Diketo-5-methylthiopentyl-1-phosphate + H ₂ O	variable	48
R00315	ATP + Acetate \rightleftharpoons ADP + Acetyl phosphate	variable	47
R01353	ATP + Propanoate \rightleftharpoons ADP + Propanoyl phosphate	variable	47
R02539	ATP + Phenylacetic acid + CoA \rightleftharpoons AMP + Diphosphate + Phenylacetyl-CoA	variable	47
R00221	D-Serine \rightleftharpoons Pyruvate + Ammonia	variable	46
R00654	L-Methionine + H ₂ O \rightleftharpoons Methanethiol + Ammonia + 2-Oxobutanoate	variable	46
R00988	Formylanthranilate + H ₂ O \rightleftharpoons Formate + Anthranilate	variable	46
R01959	L-Formylkynurenine + H ₂ O \rightleftharpoons Formate + L-Kynurenine	variable	46
R04911	5-Hydroxy-N-formylkynurenine + H ₂ O \rightleftharpoons 5-Hydroxykynurenine + Formate	variable	46
R07395	2,3-Diketo-5-methylthiopentyl-1-phosphate + H ₂ O \rightleftharpoons 1,2-Dihydroxy-5-(methylthio)pent-1-en-3-one + Orthophosphate	variable	46
R00750	Acetaldehyde + Pyruvate \rightleftharpoons 4-Hydroxy-2-oxopentanoate	variable	45
R00987	L-Kynurenine + H ₂ O \rightleftharpoons Anthranilate + L-Alanine	variable	45
R02668	3-Hydroxy-L-kynurenine + H ₂ O \rightleftharpoons 3-Hydroxyanthranilate + L-Alanine	variable	45
R03936	L-Formylkynurenine + H ₂ O \rightleftharpoons Formylanthranilate + L-Alanine	variable	45
R00678	L-Tryptophan + Oxygen \rightleftharpoons L-Formylkynurenine	variable	44
R01064	2-Dehydro-3-deoxy-6-phospho-D-galactonate \rightleftharpoons Pyruvate + D-Glyceraldehyde 3-phosphate	variable	43
R03387	ATP + 2-Dehydro-3-deoxy-D-galactonate \rightleftharpoons ADP + 2-Dehydro-3-deoxy-6-phospho-D-galactonate	variable	43
R01085	3-Fumarylpyruvate + H ₂ O \rightleftharpoons Fumarate + Pyruvate	variable	42
R01183	myo-Inositol + NAD ⁺ \rightleftharpoons 2,4,6/3,5-Pentahydroxycyclohexanone + NADH + H ⁺	variable	42

Continued from previous page

R03033	D-Galactonate \rightleftharpoons 2-Dehydro-3-deoxy-D-galactonate + H ₂ O	variable	42
R07659	UDP-L-Ara4O + L-Glutamate \rightleftharpoons UDP-L-Ara4N + 2-Oxoglutarate	variable	42
R09951	1D-chiro-Inositol + NAD ⁺ \rightleftharpoons 1-Keto-D-chiro-inositol + NADH + H ⁺	variable	42
R01541	ATP + 2-Dehydro-3-deoxy-D-gluconate \rightleftharpoons ADP + 2-Dehydro-3-deoxy-6-phospho-D-gluconate	variable	41
R08503	5-Deoxy-D-glucuronate \rightleftharpoons 2-Deoxy-5-keto-D-gluconic acid	variable	41
R00491	L-Aspartate \rightleftharpoons D-Aspartate	variable	40
R02782	2,4,6/3,5-Pentahydroxycyclohexanone \rightleftharpoons 3D-3,5/4-Trihydroxycyclohexane-1,2-dione + H ₂ O	variable	40
R03397	GDP-6-deoxy-D-mannose + NAD ⁺ \rightleftharpoons GDP-4-dehydro-6-deoxy-D-mannose + NADH + H ⁺	variable	40
R03399	GDP-6-deoxy-D-mannose + NADP ⁺ \rightleftharpoons GDP-4-dehydro-6-deoxy-D-mannose + NADPH + H ⁺	variable	40
R05661	2-Deoxy-5-keto-D-gluconic acid + ATP \rightleftharpoons 2-Deoxy-5-keto-D-gluconic acid 6-phosphate + ADP	variable	40
R08603	3D-3,5/4-Trihydroxycyclohexane-1,2-dione + H ₂ O \rightleftharpoons 5-Deoxy-D-glucuronate	variable	40
R00774	ATP + Urea + HCO ₃ ⁻ \rightleftharpoons ADP + Orthophosphate + Urea-1-carboxylate	variable	39
R03371	Phytic acid + H ₂ O \rightleftharpoons D-myo-Inositol 1,2,4,5,6-pentakisphosphate + Orthophosphate	variable	39
R08940	Acyl-[acyl-carrier protein] + S-Adenosyl-L-methionine \rightleftharpoons Acyl-carrier protein + 5'-Methylthioadenosine + N-Acyl-L-homoserine lactone	variable	39
R02527	(R)-Lactaldehyde + NAD ⁺ \rightleftharpoons Methylglyoxal + NADH + H ⁺	variable	38
R02698	4-Hydroxyphenylacetate + Oxygen + NADH + H ⁺ \rightleftharpoons 3,4-Dihydroxyphenylacetate + NAD ⁺ + H ₂ O	variable	38
R00215	(R)-Malate + NAD ⁺ \rightleftharpoons Pyruvate + CO ₂ + NADH + H ⁺	variable	36
R02545	meso-Tartaric acid + NAD ⁺ \rightleftharpoons 2-Hydroxy-3-oxosuccinate + NADH + H ⁺	variable	36
R06180	(R,R)-Tartaric acid + NAD ⁺ \rightleftharpoons 2-Hydroxy-3-oxosuccinate + NADH + H ⁺	variable	36
R00010	alpha,alpha-Trehalose + H ₂ O \rightleftharpoons 2 D-Glucose	variable	35
R09838	Phenylacetyl-CoA + Oxygen + NADPH + H ⁺ \rightleftharpoons 2-(1,2-Epoxy-1,2-dihydrophenyl)acetyl-CoA + H ₂ O + NADP ⁺	variable	35
R09952	1-Keto-D-chiro-inositol \rightleftharpoons 2,4,6/3,5-Pentahydroxycyclohexanone	variable	35
R01385	UDP-glucuronate \rightleftharpoons UDP-D-galacturonate	variable	34
R01519	D-Glucono-1,5-lactone + H ₂ O \rightleftharpoons D-Gluconic acid	variable	34
R02601	2-Hydroxy-2,4-pentadienoate + H ₂ O \rightleftharpoons 4-Hydroxy-2-oxopentanoate	variable	34
R02933	L-Gulono-1,4-lactone + H ₂ O \rightleftharpoons L-Gulonate	variable	34
R03299	3-Hydroxyphenylacetate + Oxygen + NADH + H ⁺ \rightleftharpoons 3,4-Dihydroxyphenylacetate + NAD ⁺ + H ₂ O	variable	34
R08926	6-Deoxy-L-galactose + NAD ⁺ \rightleftharpoons L-Fucono-1,5-lactone + NADH + H ⁺	variable	34
R09820	3-Oxo-5,6-dehydrosuberyl-CoA semialdehyde + NADP ⁺ + H ₂ O \rightleftharpoons 3-Oxo-5,6-dehydrosuberyl-CoA + NADPH + H ⁺	variable	34

Continued from previous page

R09836	2-Oxepin-2(3H)-ylideneacetyl-CoA + H ₂ O \rightleftharpoons 3-Oxo-5,6-dehydrosuberyl-CoA semialdehyde	variable	34
R01157	Agmatine + H ₂ O \rightleftharpoons Putrescine + Urea	variable	32
R00214	(S)-Malate + NAD ⁺ \rightleftharpoons Pyruvate + CO ₂ + NADH + H ⁺	variable	31
R01914	Spermidine + Oxygen + H ₂ O \rightleftharpoons 1,3-Diaminopropane + 4-Aminobutyraldehyde + Hydrogen peroxide	variable	31
R01915	Spermidine + Acceptor + H ₂ O \rightleftharpoons 1,3-Diaminopropane + 4-Aminobutyraldehyde + Reduced acceptor	variable	31
R03332	1-Phosphatidyl-D-myo-inositol + H ₂ O \rightleftharpoons Inositol 1-phosphate + 1,2-Diacyl-sn-glycerol	variable	31
R00317	Acetyl phosphate + H ₂ O \rightleftharpoons Acetate + Orthophosphate	variable	30
R01566	Creatine + H ₂ O \rightleftharpoons Sarcosine + Urea	variable	30
R02397	4-Trimethylammonibutanoate + 2-Oxoglutarate + Oxygen \rightleftharpoons Carnitine + Succinate + CO ₂	variable	30
R02889	UDP-glucose + Cellulose \rightleftharpoons UDP + Cellulose	variable	30
R07675	L-Galactose + NAD ⁺ \rightleftharpoons L-Galactono-1,4-lactone + NADH + H ⁺	variable	30
R09796	(R)-Lactate \rightleftharpoons Methylglyoxal + H ₂ O	variable	30
R00228	Acetaldehyde + CoA + NAD ⁺ \rightleftharpoons Acetyl-CoA + NADH + H ⁺	variable	29
R00837	H ₂ O + alpha, alpha'-Trehalose 6-phosphate \rightleftharpoons D-Glucose + D-Glucose 6-phosphate	variable	29
R00878	alpha-D-Glucose \rightleftharpoons D-Fructose	variable	29
R00919	Propanoyl-CoA + NADP ⁺ \rightleftharpoons Propenoyl-CoA + NADPH + H ⁺	variable	29
R00956	CTP + D-Glucose 1-phosphate \rightleftharpoons Diphosphate + CDP-glucose	variable	29
R01172	Butanal + CoA + NAD ⁺ \rightleftharpoons Butanoyl-CoA + NADH + H ⁺	variable	29
R01432	D-Xylose \rightleftharpoons D-Xylulose	variable	29
R02426	CDP-glucose \rightleftharpoons CDP-4-dehydro-6-deoxy-D-glucose + H ₂ O	variable	29
R02780	alpha, alpha'-Trehalose + Protein N(pi)-phospho-L-histidine \rightleftharpoons alpha, alpha'-Trehalose 6-phosphate + Protein histidine	variable	29
R00355	L-Aspartate + 2-Oxoglutarate \rightleftharpoons Oxaloacetate + L-Glutamate	variable	28
R00896	L-Cysteine + 2-Oxoglutarate \rightleftharpoons Mercaptopyruvate + Glutamate	variable	28
R01087	Maleic acid \rightleftharpoons Fumarate	variable	28
R01154	Acetyl-CoA + Putrescine \rightleftharpoons CoA + N-Acetylputrescine	variable	28
R02433	L-Cysteate + 2-Oxoglutarate \rightleftharpoons 3-Sulfopyruvate + L-Glutamate	variable	28
R02619	3-Sulfinol-L-alanine + 2-Oxoglutarate \rightleftharpoons 3-Sulfinylpyruvate + L-Glutamate	variable	28
R05052	L-erythro-4-Hydroxyglutamate + 2-Oxoglutarate \rightleftharpoons (4R)-4-Hydroxy-2-oxoglutarate + L-Glutamate	variable	28
R00999	O-Succinyl-L-homoserine + H ₂ O \rightleftharpoons 2-Oxobutanoate + Succinate + Ammonia	variable	26
R02656	2,5-Dihydroxybenzoate + Oxygen \rightleftharpoons Maleylpyruvate	variable	26
R03217	O-Acetyl-L-homoserine + L-Cysteine \rightleftharpoons L-Cystathionine + Acetate	variable	26
R03260	O-Succinyl-L-homoserine + L-Cysteine \rightleftharpoons L-Cystathionine + Succinate	variable	26

Continued from previous page			
R05772	Feruloyl-CoA + H ₂ O \rightleftharpoons 4-Hydroxy-3-methoxyphenyl-beta-hydroxypropanoyl-CoA	variable	26
R05773	4-Hydroxy-3-methoxyphenyl-beta-hydroxypropanoyl-CoA \rightleftharpoons 4-Hydroxy-3-methoxy-benzaldehyde + Acetyl-CoA	variable	26
R00802	Sucrose + H ₂ O \rightleftharpoons beta-D-Fructose + alpha-D-Glucose	variable	25
R00908	beta-Alanine + 2-Oxoglutarate \rightleftharpoons 3-Oxopropanoate + L-Glutamate	variable	25
R00005	Urea-1-carboxylate + H ₂ O \rightleftharpoons 2 CO ₂ + 2 Ammonia	variable	24
R01540	D-Altronate \rightleftharpoons 2-Dehydro-3-deoxy-D-gluconate + H ₂ O	variable	24
R03093	3-Indoleacetonitrile + 2 H ₂ O \rightleftharpoons Indole-3-acetate + Ammonia	variable	24
R06786	3-(3-Hydroxyphenyl)propanoic acid + Oxygen + NADH + H ⁺ \rightleftharpoons 3-(2,3-Dihydroxyphenyl)propanoate + H ₂ O + NAD ⁺	variable	23
R06787	trans-3-Hydroxycinnamate + Oxygen + NADH + H ⁺ \rightleftharpoons trans-2,3-Dihydroxycinnamate + H ₂ O + NAD ⁺	variable	23
R00203	Methylglyoxal + NAD ⁺ + H ₂ O \rightleftharpoons Pyruvate + NADH + H ⁺	variable	22
R00524	Formamide + H ₂ O \rightleftharpoons Formate + Ammonia	variable	22
R01066	2-Deoxy-D-ribose 5-phosphate \rightleftharpoons D-Glyceraldehyde 3-phosphate + Acetaldehyde	variable	22
R01333	Glycolaldehyde + NAD ⁺ + H ₂ O \rightleftharpoons Glycolate + NADH + H ⁺	variable	22
R01446	(S)-Lactaldehyde + NAD ⁺ + H ₂ O \rightleftharpoons (S)-Lactate + NADH + H ⁺	variable	22
R01884	Creatinine + H ₂ O \rightleftharpoons Creatine	variable	22
R01621	D-Xylulose 5-phosphate + Orthophosphate \rightleftharpoons Acetyl phosphate + D-Glyceraldehyde 3-phosphate + H ₂ O	variable	21
R03187	ATP + N-Methylhydantoin + 2 H ₂ O \rightleftharpoons ADP + Orthophosphate + N-Carbamoylsarcosine	variable	21
R00731	L-Tyrosine + Oxygen \rightleftharpoons 3,4-Dihydroxy-L-phenylalanine + H ₂ O	variable	20
R00801	Sucrose + H ₂ O \rightleftharpoons D-Fructose + D-Glucose	variable	20
R02078	3,4-Dihydroxy-L-phenylalanine + L-Tyrosine + Oxygen \rightleftharpoons Dopaquinone + 3,4-Dihydroxy-L-phenylalanine + H ₂ O	variable	20
R02410	Raffinose + H ₂ O \rightleftharpoons Melibiose + D-Fructose	variable	20
R03635	Stachyose + H ₂ O \rightleftharpoons D-Gal alpha 1-6D-Gal alpha 1-6D-Glucose + D-Fructose	variable	20
R03688	L-Fuconate \rightleftharpoons 2-Dehydro-3-deoxy-L-fuconate + H ₂ O	variable	20
R03921	Sucrose 6-phosphate + H ₂ O \rightleftharpoons D-Fructose + D-Glucose 6-phosphate	variable	20
R04188	L-3-Aminoisobutanoate + 2-Oxoglutarate \rightleftharpoons (S)-Methylmalonate semialdehyde + L-Glutamate	variable	20
R04884	2 5,6-Dihydroxyindole + Oxygen \rightleftharpoons 2 Indole-5,6-quinone + 2 H ₂ O	variable	20
R05140	D-Glucose + Levan \rightleftharpoons Sucrose + Levan	variable	20
R00342	(S)-Malate + NAD ⁺ \rightleftharpoons Oxaloacetate + NADH + H ⁺	variable	19
R00421	UDP-N-acetyl-alpha-D-glucosamine + 2 NAD ⁺ + H ₂ O \rightleftharpoons UDP-N-acetyl-2-amino-2-deoxy-D-glucuronate + 2 NADH + 2 H ⁺	variable	19
R01016	Glycerone phosphate \rightleftharpoons Methylglyoxal + Orthophosphate	variable	19

Continued from previous page

R07136	(2R)-3-Sulfolactate + NAD+ \rightleftharpoons 3-Sulfoypyruvate + NADH + H+	variable	19
R00269	2-Oxoglutaramate + H ₂ O \rightleftharpoons 2-Oxoglutarate + Ammonia	variable	18
R00348	2-Oxosuccinamate + H ₂ O \rightleftharpoons Oxaloacetate + Ammonia	variable	18
R00505	UDP-alpha-D-galactose \rightleftharpoons UDP-alpha-D-galactofuranose	variable	18
R00525	N-Formyl-L-glutamate + H ₂ O \rightleftharpoons Formate + L-Glutamate	variable	18
R01777	Succinyl-CoA + L-Homoserine \rightleftharpoons CoA + O-Succinyl-L-homoserine	variable	18
R09009	UDP-L-arabinose \rightleftharpoons UDP-L-arabinofuranose	variable	18
R00194	S-Adenosyl-L-homocysteine + H ₂ O \rightleftharpoons S-Ribosyl-L-homocysteine + Adenine	variable	17
R00922	2-Methyl-3-oxopropanoate + CoA + NAD+ \rightleftharpoons Propanoyl-CoA + CO ₂ + NADH + H+	variable	17
R01401	5'-Methylthioadenosine + H ₂ O \rightleftharpoons Adenine + 5-Methylthio-D-ribose	variable	17
R01542	2-Dehydro-3-deoxy-D-gluconate + NAD+ \rightleftharpoons (4S)-4,6-Dihydroxy-2,5-dioxohexanoate + NADH + H+	variable	17
R01544	2-Amino-2-deoxy-D-gluconate \rightleftharpoons 2-Dehydro-3-deoxy-D-gluconate + Ammonia	variable	17
R02765	(R)-Methylmalonyl-CoA \rightleftharpoons (S)-Methylmalonyl-CoA	variable	17
R04217	N ₂ -Succinyl-L-ornithine + 2-Oxoglutarate \rightleftharpoons N-Succinyl-L-glutamate 5-semialdehyde + L-Glutamate	variable	17
R07274	O-Phospho-L-serine + Hydrogen sulfide \rightleftharpoons L-Cysteine + Orthophosphate	variable	17
R09660	5'-Methylthioadenosine + H ₂ O \rightleftharpoons 5'-S-Methyl-5'-thioinosine + Ammonia	variable	17
R09979	(2S)-Ethylmalonyl-CoA \rightleftharpoons (2R)-Ethylmalonyl-CoA	variable	17
R00306	Cellobiose + H ₂ O \rightleftharpoons 2 D-Glucose	variable	16
R01859	ATP + Propanoyl-CoA + HCO ₃ - \rightleftharpoons ADP + Orthophosphate + (S)-Methylmalonyl-CoA	variable	16
R02025	L-Methionine + Thioredoxin disulfide + H ₂ O \rightleftharpoons L-Methionine S-oxide + Thioredoxin	variable	16
R02112	Starch \rightleftharpoons Dextrin + Maltose	variable	16
R09030	D-Allose 6-phosphate \rightleftharpoons D-Allulose 6-phosphate	variable	16
R09097	Propanal + NAD+ + CoA \rightleftharpoons Propanoyl-CoA + NADH + H+	variable	16
R00008	Parapyruvate \rightleftharpoons 2 Pyruvate	variable	15
R00449	L-Lysine + Oxygen \rightleftharpoons 5-Aminopentanamide + CO ₂ + H ₂ O	variable	15
R02273	5-Aminopentanamide + H ₂ O \rightleftharpoons 5-Aminopentanoate + Ammonia	variable	15
R02522	L-Arabinonate \rightleftharpoons 2-Dehydro-3-deoxy-L-arabinonate + H ₂ O	variable	15
R04020	(Indol-3-yl)acetamide \rightleftharpoons 3-Indoleacetonitrile + H ₂ O	variable	15
R06979	N(gamma)-Acetyldiaminobutyrate \rightleftharpoons H ₂ O + Ectoine	variable	15
R10851	L-Allothreonine + NADP+ \rightleftharpoons L-2-Amino-3-oxobutanoic acid + NADPH + H+	variable	15
R00028	Maltose + H ₂ O \rightleftharpoons 2 D-Glucose	variable	14
R00448	L-Lysine + Oxygen + NADPH + H+ \rightleftharpoons N6-Hydroxy-L-lysine + NADP+ + H ₂ O	variable	14
R02737	UDP-glucose + alpha-D-Glucose 6-phosphate \rightleftharpoons UDP + alpha,alpha'-Trehalose 6-phosphate	variable	14

Continued from previous page

R02778	alpha,alpha'-Trehalose 6-phosphate + H ₂ O <=> alpha,alpha'-Trehalose + Orthophosphate	variable	14
R03317	UDP-N-acetyl-D-mannosamine + 2 NAD ⁺ + H ₂ O <=> UDP-N-acetyl-D-mannosaminouronate + 2 NADH + 2 H ⁺	variable	14
R03774	L-Rhamnonate <=> 2-Dehydro-3-deoxy-L-rhamnonate + H ₂ O	variable	14
R05692	GDP-L-fucose + NADP ⁺ <=> GDP-4-dehydro-6-deoxy-D-mannose + NADPH + H ⁺	variable	14
R10140	UDP-N-acetyl-2-amino-2-deoxy-D-glucuronate + NAD ⁺ <=> UDP-2-acetamido-2-deoxy-alpha-D-ribo-hex-3-uluronate + NADH + H ⁺	variable	14
R10141	UDP-2-acetamido-2-deoxy-alpha-D-ribo-hex-3-uluronate + L-Glutamate <=> UDP-2-acetamido-3-amino-2,3-dideoxy-alpha-D-glucuronate + 2-Oxoglutarate	variable	14
R01804	N-Acetylneuraminate + Orthophosphate <=> N-Acetyl-D-mannosamine + Phosphoenolpyruvate + H ₂ O	variable	13
R02427	D-Xylonolactone + H ₂ O <=> D-Xylonate	variable	13
R03889	2-Aminomuconate semialdehyde + NAD ⁺ + H ₂ O <=> 2-Aminomuconate + H ⁺ + NADH	variable	13
R04435	N-Acetylneuraminate 9-phosphate + Orthophosphate <=> N-Acetyl-D-mannosamine 6-phosphate + Phosphoenolpyruvate + H ₂ O	variable	13
R10100	Acetyl-CoA + UDP-2-acetamido-3-amino-2,3-dideoxy-alpha-D-glucuronate <=> CoA + UDP-2,3-diacetamido-2,3-dideoxy-alpha-D-glucuronate	variable	13
R01728	Prephenate + NAD ⁺ <=> 3-(4-Hydroxyphenyl)pyruvate + CO ₂ + NADH + H ⁺	variable	12
R02280	1-Pyrroline-4-hydroxy-2-carboxylate + H ₂ O <=> 2,5-Dioxopentanoate + Ammonia	variable	12
R02886	Cellulose + H ₂ O <=> Cellulose + Cellobiose	variable	12
R09697	UDP-N-acetyl-alpha-D-glucosamine <=> UDP-2-acetamido-2,6-dideoxy-beta-L-arabino-hexos-4-ulose + H ₂ O	variable	12
R10178	4-Aminobutanoate + Pyruvate <=> Succinate semialdehyde + L-Alanine	variable	12
R06782	Phenylpropanoate + Oxygen + NADH + H ⁺ <=> cis-3-(Carboxy-ethyl)-3,5-cyclo-hexadiene-1,2-diol + NAD ⁺	variable	11
R06783	trans-Cinnamate + Oxygen + NADH + H ⁺ <=> cis-3-(3-Carboxyethenyl)-3,5-cyclohexadiene-1,2-diol + NAD ⁺	variable	11
R00339	(R,R)-Tartaric acid <=> Oxaloacetate + H ₂ O	variable	10
R00418	UDP-N-acetyl-alpha-D-glucosamine <=> UDP-N-acetyl-D-galactosamine	variable	10
R01206	Chitin + H ₂ O <=> N-Acetyl-D-glucosamine + Chitin	variable	10
R02334	Chitin + H ₂ O <=> Chitobiose + Chitin	variable	10
R02928	Galactitol + NAD ⁺ <=> D-Tagatose + NADH + H ⁺	variable	10
R10715	(S)-Lactaldehyde + NAD ⁺ <=> Methylglyoxal + NADH + H ⁺	variable	10
R10717	Propane-1,2-diol + NAD ⁺ <=> Hydroxyacetone + NADH + H ⁺	variable	10
R00197	(R)-Lactate + 2 Ferricytochrome c <=> Pyruvate + 2 Ferrocyclochrome c + 2 H ⁺	variable	9
R06978	L-2,4-Diaminobutanoate + Acetyl-CoA <=> N(gamma)-Acetyldiaminobutyrate + CoA	variable	9

Continued from previous page

R00396	L-Alanine + NAD ⁺ + H ₂ O \rightleftharpoons Pyruvate + Ammonia + NADH + H ⁺	variable	8
R00875	D-Sorbitol + NAD ⁺ \rightleftharpoons D-Fructose + NADH + H ⁺	variable	8
R01094	D-Galactose + NAD ⁺ \rightleftharpoons D-Galactono-1,4-lactone + NADH + H ⁺	variable	8
R01097	D-Galactose + NADP ⁺ \rightleftharpoons D-Galactono-1,4-lactone + NADPH + H ⁺	variable	8
R01896	Xylitol + NAD ⁺ \rightleftharpoons D-Xylulose + NADH + H ⁺	variable	8
R00551	L-Arginine + H ₂ O \rightleftharpoons L-Ornithine + Urea	variable	7
R00565	L-Arginine + Glycine \rightleftharpoons L-Ornithine + Guanidinoacetate	variable	7
R00830	Succinyl-CoA + Glycine \rightleftharpoons 5-Aminolevulinate + CoA + CO ₂	variable	7
R01167	L-Histidine \rightleftharpoons Histamine + CO ₂	variable	7
R01977	(R)-3-Hydroxybutanoyl-CoA + NADP ⁺ \rightleftharpoons Acetoacetyl-CoA + NADPH + H ⁺	variable	7
R01978	(S)-3-Hydroxy-3-methylglutaryl-CoA + CoA \rightleftharpoons Acetyl-CoA + H ₂ O + Acetoacetyl-CoA	variable	7
R01989	L-Arginine + 4-Aminobutanoate \rightleftharpoons L-Ornithine + 4-Guanidinobutanoate	variable	7
R02611	Phenylethyl alcohol + NAD ⁺ \rightleftharpoons Phenylacetaldehyde + NADH + H ⁺	variable	7
R03236	D-Tagatose 6-phosphate + ATP \rightleftharpoons D-Tagatose 1,6-bisphosphate + ADP	variable	7
R03237	CTP + D-Tagatose 6-phosphate \rightleftharpoons CDP + D-Tagatose 1,6-bisphosphate	variable	7
R03238	UTP + D-Tagatose 6-phosphate \rightleftharpoons UDP + D-Tagatose 1,6-bisphosphate	variable	7
R03239	ITP + D-Tagatose 6-phosphate \rightleftharpoons IDP + D-Tagatose 1,6-bisphosphate	variable	7
R04304	4-Hydroxyphenylethanol + NAD ⁺ \rightleftharpoons 4-Hydroxyphenylacetaldehyde + NADH + H ⁺	variable	7
R04323	2-Amino-3-carboxymuconate semialdehyde \rightleftharpoons 2-Aminomuconate semialdehyde + CO ₂	variable	7
R04779	ATP + beta-D-Fructose 6-phosphate \rightleftharpoons ADP + beta-D-Fructose 1,6-bisphosphate	variable	7
R07650	L-2,4-Diaminobutanoate \rightleftharpoons 1,3-Diaminopropane + CO ₂	variable	7
R09600	UDP-2,3-diacetamido-2,3-dideoxy-alpha-D-glucuronate \rightleftharpoons UDP-2,3-diacetamido-2,3-dideoxy-alpha-D-mannuronate	variable	7
R00685	L-Tryptophan \rightleftharpoons Tryptamine + CO ₂	variable	6
R00699	L-Phenylalanine \rightleftharpoons Phenethylamine + CO ₂	variable	6
R00736	L-Tyrosine \rightleftharpoons Tyramine + CO ₂	variable	6
R01429	D-Xylose + NAD ⁺ \rightleftharpoons D-Xylonolactone + NADH + H ⁺	variable	6
R01627	3-Dehydroshikimate \rightleftharpoons 3,4-Dihydroxybenzoate + H ₂ O	variable	6
R02080	3,4-Dihydroxy-L-phenylalanine \rightleftharpoons Dopamine + CO ₂	variable	6
R02603	2-Hydroxy-2,4-pentadienoate + Succinate \rightleftharpoons 2-Hydroxy-6-oxonona-2,4-diene-1,9-dioate + H ₂ O	variable	6
R02701	5-Hydroxy-L-tryptophan \rightleftharpoons Serotonin + CO ₂	variable	6
R04909	5-Hydroxykynurenamine + CO ₂ \rightleftharpoons 5-Hydroxykynurenine	variable	6
R06789	2-Hydroxy-6-ketonoatrienedioate + H ₂ O \rightleftharpoons 2-Hydroxy-2,4-pentadienoate + Fumarate	variable	6

Continued from previous page			
R07409	Choline + Oxygen + 2 Reduced ferredoxin + 2 H+ \rightleftharpoons Betaine aldehyde + 2 H ₂ O + 2 Oxidized ferredoxin	variable	6
R10060	2 S-Adenosyl-L-methionine + Glycine \rightleftharpoons 2 S-Adenosyl-L-homocysteine + N,N-Dimethylglycine	variable	6
R00261	L-Glutamate \rightleftharpoons 4-Aminobutanoate + CO ₂	variable	5
R00928	Acetyl-CoA + Propanoate \rightleftharpoons Acetate + Propanoyl-CoA	variable	5
R01449	Lactoyl-CoA + Propanoate \rightleftharpoons (S)-Lactate + Propanoyl-CoA	variable	5
R02454	D-Mannionate + NAD+ \rightleftharpoons D-Fructuronate + NADH + H+	variable	5
R02526	L-Arabinono-1,4-lactone + H ₂ O \rightleftharpoons L-Arabinonate	variable	5
R03372	Phytic acid + H ₂ O \rightleftharpoons Inositol 1,2,3,5,6-pentakisphosphate + Orthophosphate	variable	5
R03887	2-Aminomuconate + H ₂ O \rightleftharpoons gamma-Oxalocrotonate + Ammonia	variable	5
R03942	L-Rhamnofuranose + NAD+ \rightleftharpoons L-Rhamnono-1,4-lactone + NADH + H+	variable	5
R04376	3-(2,3-Dihydroxyphenyl)propanoate + Oxygen \rightleftharpoons 2-Hydroxy-6-oxonona-2,4-diene-1,9-dioate	variable	5
R06788	trans-2,3-Dihydroxycinnamate + Oxygen \rightleftharpoons 2-Hydroxy-6-ketnonatrienedioate	variable	5
R10788	L-Rhamnose + NADP+ \rightleftharpoons L-Rhamnono-1,4-lactone + NADPH + H+	variable	5
R10995	L-Rhamnose + NAD+ \rightleftharpoons L-Rhamnono-1,4-lactone + NADH + H+	variable	5
R01155	Putrescine + 2-Oxoglutarate \rightleftharpoons 4-Aminobutyraldehyde + L-Glutamate	variable	4
R01246	L-Proline + NAD+ \rightleftharpoons 1-Pyrroline-2-carboxylate + NADH + H+	variable	4
R01249	L-Proline + NADP+ \rightleftharpoons 1-Pyrroline-2-carboxylate + NADPH + H+	variable	4
R01402	5'-Methylthioadenosine + Orthophosphate \rightleftharpoons Adenine + S-Methyl-5-thio-D-ribose 1-phosphate	variable	4
R01934	Homoisocitrate + NAD+ \rightleftharpoons 2-Oxo adipate + CO ₂ + NADH + H+	variable	4
R01982	Pectate + H ₂ O \rightleftharpoons D-Galacturonate + Pectate	variable	4
R02201	L-Pipecolate + NAD+ \rightleftharpoons Delta1-Piperidine-2-carboxylate + NADH + H+	variable	4
R02360	Pectate + H ₂ O \rightleftharpoons Digalacturonate + Pectate	variable	4
R03013	L-Histidinol phosphate + H ₂ O \rightleftharpoons L-Histidinol + Orthophosphate	variable	4
R05606	D-Mannionate \rightleftharpoons 2-Dehydro-3-deoxy-D-gluconate + H ₂ O	variable	4
R10848	L-Gulonate + NAD+ \rightleftharpoons D-Fructuronate + NADH + H+	variable	4
R00305	D-Glucose + Quinone \rightleftharpoons D-Glucono-1,5-lactone + Hydroquinone	variable	3
R00371	Acetyl-CoA + Glycine \rightleftharpoons CoA + L-2-Amino-3-oxobutanoic acid	variable	3
R00461	L-Lysine \rightleftharpoons (3S)-3,6-Diaminohexanoate	variable	3
R00709	Isocitrate + NAD+ \rightleftharpoons 2-Oxoglutarate + CO ₂ + NADH + H+	variable	3
R01117	CTP + N-Acetylneuraminate \rightleftharpoons Diphosphate + CMP-N-acetylneuraminate	variable	3
R01644	4-Hydroxybutanoic acid + NAD+ \rightleftharpoons Succinate semialdehyde + NADH + H+	variable	3

Continued from previous page			
R02665	3-Hydroxyanthranilate + Oxygen \rightleftharpoons 2-Amino-3-carboxymuconate semialdehyde	variable	3
R03396	GDP-6-deoxy-D-talose + NAD ⁺ \rightleftharpoons GDP-4-dehydro-6-deoxy-D-mannose + NADH + H ⁺	variable	3
R03398	GDP-6-deoxy-D-talose + NADP ⁺ \rightleftharpoons GDP-4-dehydro-6-deoxy-D-mannose + NADPH + H ⁺	variable	3
R03707	(S,S)-Butane-2,3-diol + NAD ⁺ \rightleftharpoons (S)-Acetoin + NADH + H ⁺	variable	3
R04215	CTP + N-Glycoloyl-neuraminate \rightleftharpoons Diphosphate + CMP-N-glycoloylneuraminate	variable	3
R04374	trans-3-Hydroxy-L-proline \rightleftharpoons 1-Pyrroline-2-carboxylate + H ₂ O	variable	3
R09078	Diacetyl + NADH + H ⁺ \rightleftharpoons (S)-Acetoin + NAD ⁺	variable	3
R09291	(2S)-Ethylmalonyl-CoA + NADP ⁺ \rightleftharpoons Crotonoyl-CoA + CO ₂ + NADPH + H ⁺	variable	3
R09843	Pseudaminic acid + CTP \rightleftharpoons CMP-pseudaminic acid + Diphosphate	variable	3
R00233	Malonyl-CoA \rightleftharpoons Acetyl-CoA + CO ₂	variable	2
R00271	Acetyl-CoA + H ₂ O + 2-Oxoglutarate \rightleftharpoons (R)-2-Hydroxybutane-1,2,4-tricarboxylate + CoA	variable	2
R00369	L-Alanine + Glyoxylate \rightleftharpoons Pyruvate + Glycine	variable	2
R00372	Glycine + 2-Oxoglutarate \rightleftharpoons Glyoxylate + L-Glutamate	variable	2
R00397	L-Aspartate \rightleftharpoons L-Alanine + CO ₂	variable	2
R00585	L-Serine + Pyruvate \rightleftharpoons Hydroxypyruvate + L-Alanine	variable	2
R00588	L-Serine + Glyoxylate \rightleftharpoons Hydroxypyruvate + Glycine	variable	2
R00667	L-Ornithine + 2-Oxoglutarate \rightleftharpoons L-Glutamate 5-semialdehyde + L-Glutamate	variable	2
R00863	3-Sulfino-L-alanine \rightleftharpoons L-Alanine + Sulfur dioxide	variable	2
R00923	(S)-Methylmalonyl-CoA \rightleftharpoons Propanoyl-CoA + CO ₂	variable	2
R01059	ATP + D-Glyceraldehyde \rightleftharpoons ADP + D-Glyceraldehyde 3-phosphate	variable	2
R01255	L-Proline \rightleftharpoons D-Proline	variable	2
R01366	Acetoacetate \rightleftharpoons Acetone + CO ₂	variable	2
R01384	UDP-glucuronate \rightleftharpoons UDP-D-xylose + CO ₂	variable	2
R01790	Starch + H ₂ O \rightleftharpoons D-Glucose + Starch	variable	2
R01791	Dextrin + H ₂ O \rightleftharpoons D-Glucose + Dextrin	variable	2
R01895	Ribitol + NAD ⁺ \rightleftharpoons D-Ribulose + NADH + H ⁺	variable	2
R02361	Pectate \rightleftharpoons 4-(4-Deoxy- α -D-gluc-4-enuronosyl)-D-galacturonate + Pectate	variable	2
R02555	D-Altronate + NAD ⁺ \rightleftharpoons D-Tagaturonate + NADH + H ⁺	variable	2
R02925	D-Sorbitol + FAD \rightleftharpoons FADH ₂ + L-Sorbose	variable	2
R03027	(R)-3-Hydroxybutanoyl-CoA \rightleftharpoons Crotonoyl-CoA + H ₂ O	variable	2
R03130	S-Adenosyl-L-methionine + N-Acetylserotonin \rightleftharpoons S-Adenosyl-L-homocysteine + Melatonin	variable	2
R03140	D-Ribulose 1,5-bisphosphate + Oxygen \rightleftharpoons 3-Phospho-D-glycerate + 2-Phosphoglycolate	variable	2
R04905	S-Adenosyl-L-methionine + 5-Hydroxyindoleacetate \rightleftharpoons S-Adenosyl-L-homocysteine + 5-Methoxyindoleacetate	variable	2
R10460	GDP-4-amino-4,6-dideoxy- α -D-mannose + 2-Oxoglutarate \rightleftharpoons GDP-4-dehydro-6-deoxy-D-mannose + L-Glutamate	variable	2
R00262	L-threo-3-Methylaspartate \rightleftharpoons L-Glutamate	variable	1

Continued from previous page			
R00308	1,3-beta-D-Glucan + H2O \rightleftharpoons D-Glucose + 1,3-beta-D-Glucan	variable	1
R00650	S-Adenosyl-L-methionine + L-Homocysteine \rightleftharpoons S-Adenosyl-L-homocysteine + L-Methionine	variable	1
R00673	L-Tryptophan + H2O \rightleftharpoons Indole + Pyruvate + Ammonia	variable	1
R01063	D-Glyceraldehyde 3-phosphate + Orthophosphate + NADP+ \rightleftharpoons 3-Phospho-D-glyceroyl phosphate + NADPH + H+	variable	1
R01525	D-Ribitol 5-phosphate + NADP+ \rightleftharpoons D-Ribulose 5-phosphate + NADPH + H+	variable	1
R01526	ATP + D-Ribulose \rightleftharpoons ADP + D-Ribulose 5-phosphate	variable	1
R01974	Indolepyruvate \rightleftharpoons Indole-3-acetaldehyde + CO2	variable	1
R02429	D-Xylonate \rightleftharpoons 2-Dehydro-3-deoxy-D-xylonate + H2O	variable	1
R02749	2-Deoxy-D-ribose 1-phosphate \rightleftharpoons 2-Deoxy-D-ribose 5-phosphate	variable	1
R02754	5-Dehydro-4-deoxy-D-glucarate \rightleftharpoons Pyruvate + 2-Hydroxy-3-oxopropanoate	variable	1
R02755	meso-2,6-Diaminoheptanedioate + NADP+ + H2O \rightleftharpoons L-2-Amino-6-oxoheptanedioate + Ammonia + NADPH + H+	variable	1
R02833	Chitosan + H2O \rightleftharpoons D-Glucosaminide + Chitosan	variable	1
R02921	CTP + D-Ribitol 5-phosphate \rightleftharpoons Diphosphate + CDP-ribitol	variable	1
R02948	(S)-2-Acetolactate \rightleftharpoons (R)-Acetoin + CO2	variable	1
R03028	Glutaconyl-1-CoA \rightleftharpoons Crotonoyl-CoA + CO2	variable	1
R03115	1,3-beta-D-Glucan + H2O \rightleftharpoons 1,3-beta-D-Glucan + alpha-D-Glucose	variable	1
R03168	Acetyl-CoA + N6-Hydroxy-L-lysine \rightleftharpoons CoA + N6-Acetyl-N6-hydroxy-L-lysine	variable	1
R03277	2-Hydroxy-3-oxopropanoate + Pyruvate \rightleftharpoons 2-Dehydro-3-deoxy-D-glucarate	variable	1
R03629	Melatonin + [Reduced NADPH---hemoprotein reductase] + Oxygen \rightleftharpoons 6-Hydroxymelatonin + [Oxidized NADPH---hemoprotein reductase] + H2O	variable	1
R03868	Maleylpyruvate \rightleftharpoons 3-Fumarylpyruvate	variable	1
R05133	Arbutin 6-phosphate + H2O \rightleftharpoons Hydroquinone + beta-D-Glucose 6-phosphate	variable	1
R05134	Salicin 6-phosphate + H2O \rightleftharpoons Salicyl alcohol + beta-D-Glucose 6-phosphate	variable	1
R07613	LL-2,6-Diaminoheptanedioate + 2-Oxoglutarate \rightleftharpoons 2,3,4,5-Tetrahydrodipicolinate + L-Glutamate + H2O	variable	1
R09293	(2S)-Methylsuccinyl-CoA + Electron-transferring flavoprotein \rightleftharpoons 2-Methylfumaryl-CoA + Reduced electron-transferring flavoprotein	variable	1
R10846	D-Galactaro-1,5-lactone \rightleftharpoons D-Galactaro-1,4-lactone	variable	1

Appendix E: List of conserved reactions related to energy metabolism

Reaction Id	Reaction Equation	Type
R00238	2 Acetyl-CoA \rightleftharpoons CoA + Acetoacetyl-CoA	conserved
R00253	ATP + L-Glutamate + Ammonia \rightleftharpoons ADP + Orthophosphate + L-Glutamine	conserved
R00267	Isocitrate + NADP+ \rightleftharpoons 2-Oxoglutarate + CO ₂ + NADPH + H+	conserved
R00344	ATP + Pyruvate + HCO ₃ ⁻ \rightleftharpoons ADP + Orthophosphate + Oxaloacetate	conserved
R00405	ATP + Succinate + CoA \rightleftharpoons ADP + Orthophosphate + Succinyl-CoA	conserved
R00508	3'-Phosphoadenylyl sulfate + H ₂ O \rightleftharpoons Adenylyl sulfate + Orthophosphate	conserved
R00509	ATP + Adenylyl sulfate \rightleftharpoons ADP + 3'-Phosphoadenylyl sulfate	conserved
R00529	ATP + Sulfate \rightleftharpoons Diphosphate + Adenylyl sulfate	conserved
R00586	L-Serine + Acetyl-CoA \rightleftharpoons O-Acetyl-L-serine + CoA	conserved
R00658	2-Phospho-D-glycerate \rightleftharpoons Phosphoenolpyruvate + H ₂ O	conserved
R00742	ATP + Acetyl-CoA + HCO ₃ ⁻ \rightleftharpoons ADP + Orthophosphate + Malonyl-CoA	conserved
R00762	D-Fructose 1,6-bisphosphate + H ₂ O \rightleftharpoons D-Fructose 6-phosphate + Orthophosphate	conserved
R00858	Hydrogen sulfide + 3 NADP+ + 3 H ₂ O \rightleftharpoons Sulfite + 3 NADPH + 3 H+	conserved
R00945	5,10-Methylenetetrahydrofolate + Glycine + H ₂ O \rightleftharpoons Tetrahydrofolate + L-Serine	conserved
R01015	D-Glyceraldehyde 3-phosphate \rightleftharpoons Glycerone phosphate	conserved
R01056	D-Ribose 5-phosphate \rightleftharpoons D-Ribulose 5-phosphate	conserved
R01061	D-Glyceraldehyde 3-phosphate + Orthophosphate + NAD+ \rightleftharpoons 3-Phospho-D-glyceroyl phosphate + NADH + H+	conserved
R01082	(S)-Malate \rightleftharpoons Fumarate + H ₂ O	conserved
R01325	Citrate \rightleftharpoons cis-Aconitate + H ₂ O	conserved
R01395	ATP + Carbamate \rightleftharpoons ADP + Carbamoyl phosphate	conserved
R01518	2-Phospho-D-glycerate \rightleftharpoons 3-Phospho-D-glycerate	conserved
R01529	D-Ribulose 5-phosphate \rightleftharpoons D-Xylulose 5-phosphate	conserved
R01900	Isocitrate \rightleftharpoons cis-Aconitate + H ₂ O	conserved
R01931	Thiosulfate + Cyanide ion \rightleftharpoons Sulfite + Thiocyanate	conserved
R01975	(S)-3-Hydroxybutanoyl-CoA + NAD+ \rightleftharpoons Acetoacetyl-CoA + NADH + H+	conserved
R02021	Thioredoxin + 3'-Phosphoadenylyl sulfate \rightleftharpoons Thioredoxin disulfide + Sulfite + Adenosine 3',5'-bisphosphate	conserved
R02164	Quinone + Succinate \rightleftharpoons Hydroquinone + Fumarate	conserved
R03026	(S)-3-Hydroxybutanoyl-CoA \rightleftharpoons Crotonoyl-CoA + H ₂ O	conserved
R03045	3-Hydroxypropionyl-CoA \rightleftharpoons Propenoyl-CoA + H ₂ O	conserved
R04173	O-Phospho-L-serine + 2-Oxoglutarate \rightleftharpoons 3-Phosphonooxypyruvate + L-Glutamate	conserved
R07168	5-Methyltetrahydrofolate + NAD+ \rightleftharpoons 5,10-Methylenetetrahydrofolate + NADH + H+	conserved
R09099	L-Serine + 5,6,7,8-Tetrahydromethanopterin \rightleftharpoons 5,10-Methylenetetrahydromethanopterin + Glycine + H ₂ O	conserved
R10092	HCO ₃ ⁻ + H+ \rightleftharpoons CO ₂ + H ₂ O	conserved

Appendix F: List of variable reactions related to energy metabolism

Reaction Id	Reaction Equation	Type	No. of organisms
R00114	2 L-Glutamate + NADP+ \rightleftharpoons L-Glutamine + 2-Oxoglutarate + NADPH + H+	variable	110
R00235	ATP + Acetate + CoA \rightleftharpoons AMP + Diphosphate + Acetyl-CoA	variable	110
R00248	L-Glutamate + NADP+ + H ₂ O \rightleftharpoons 2-Oxoglutarate + Ammonia + NADPH + H+	variable	110
R00258	L-Alanine + 2-Oxoglutarate \rightleftharpoons Pyruvate + L-Glutamate	variable	110
R00345	Orthophosphate + Oxaloacetate \rightleftharpoons H ₂ O + Phosphoenolpyruvate + CO ₂	variable	110
R00527	S-Formylglutathione + H ₂ O \rightleftharpoons Formate + Glutathione	variable	110
R00582	O-Phospho-L-serine + H ₂ O \rightleftharpoons L-Serine + Orthophosphate	variable	110
R00897	O-Acetyl-L-serine + Hydrogen sulfide \rightleftharpoons L-Cysteine + Acetate	variable	110
R01067	D-Fructose 6-phosphate + D-Glyceraldehyde 3-phosphate \rightleftharpoons D-Erythrose 4-phosphate + D-Xylulose 5-phosphate	variable	110
R01068	D-Fructose 1,6-bisphosphate \rightleftharpoons Glycerone phosphate + D-Glyceraldehyde 3-phosphate	variable	110
R01220	5,10-Methylenetetrahydrofolate + NADP+ \rightleftharpoons 5,10-Methenyltetrahydrofolate + NADPH	variable	110
R01388	D-Glycerate + NAD+ \rightleftharpoons Hydroxypyruvate + NADH + H+	variable	110
R01512	ATP + 3-Phospho-D-glycerate \rightleftharpoons ADP + 3-Phospho-D-glyceroyl phosphate	variable	110
R01513	3-Phospho-D-glycerate + NAD+ \rightleftharpoons 3-Phosphonooxypyruvate + NADH + H+	variable	110
R01641	Sedoheptulose 7-phosphate + D-Glyceraldehyde 3-phosphate \rightleftharpoons D-Ribose 5-phosphate + D-Xylulose 5-phosphate	variable	110
R01655	5,10-Methenyltetrahydrofolate + H ₂ O \rightleftharpoons 10-Formyltetrahydrofolate + H+	variable	110
R01829	Sedoheptulose 1,7-bisphosphate \rightleftharpoons Glycerone phosphate + D-Erythrose 4-phosphate	variable	110
R06983	S-(Hydroxymethyl)glutathione + NAD+ \rightleftharpoons S-Formylglutathione + NADH + H+	variable	110
R08572	D-Glycerate + ATP \rightleftharpoons 2-Phospho-D-glycerate + ADP	variable	110
R00199	ATP + Pyruvate + H ₂ O \rightleftharpoons AMP + Phosphoenolpyruvate + Orthophosphate	variable	109
R00243	L-Glutamate + NAD+ + H ₂ O \rightleftharpoons 2-Oxoglutarate + Ammonia + NADH + H+	variable	109
R00341	ATP + Oxaloacetate \rightleftharpoons ADP + Phosphoenolpyruvate + CO ₂	variable	109
R01288	O-Succinyl-L-homoserine + Hydrogen sulfide \rightleftharpoons L-Homocysteine + Succinate	variable	109
R00025	Ethylnitronate + Oxygen + Reduced FMN \rightleftharpoons Acetaldehyde + Nitrite + FMN + H ₂ O	variable	108
R00216	(S)-Malate + NADP+ \rightleftharpoons Pyruvate + CO ₂ + NADPH + H+	variable	107
R00230	Acetyl-CoA + Orthophosphate \rightleftharpoons CoA + Acetyl phosphate	variable	107
R00787	Ammonia + 3 NAD+ + 2 H ₂ O \rightleftharpoons Nitrite + 3 NADH + 3 H+	variable	104
R07210	Alkanesulfonate + Reduced FMN + Oxygen \rightleftharpoons Aldehyde + FMN + Sulfite + H ₂ O	variable	102

Continued from previous page

R10206	Methanesulfonic acid + Reduced FMN + Oxygen \rightleftharpoons FMN + Sulfite + H ₂ O + Formaldehyde	variable	102
R00604	Formaldehyde + NAD ⁺ + H ₂ O \rightleftharpoons Formate + NADH + H ⁺	variable	100
R05320	Taurine + 2-Oxoglutarate + Oxygen \rightleftharpoons Sulfite + Aminoacetaldehyde + Succinate + CO ₂	variable	95
R00093	2 L-Glutamate + NAD ⁺ \rightleftharpoons L-Glutamine + 2-Oxoglutarate + NADH + H ⁺	variable	93
R00519	Formate + NAD ⁺ \rightleftharpoons H ⁺ + CO ₂ + NADH	variable	81
R00798	Nitrite + Acceptor + H ₂ O \rightleftharpoons Nitrate + Reduced acceptor	variable	76
R03546	Cyanate + H ⁺ + HCO ₃ ⁻ \rightleftharpoons CO ₂ + Carbamate	variable	76
R00237	(3S)-Citramalyl-CoA \rightleftharpoons Acetyl-CoA + Pyruvate	variable	64
R00783	Nitric oxide + H ₂ O + Ferricytochrome c \rightleftharpoons Nitrite + Ferrocycytochrome c + H ⁺	variable	54
R00785	Nitric oxide + H ₂ O + Oxidized azurin \rightleftharpoons Nitrite + Reduced azurin + H ⁺	variable	54
R09497	Nitrate + Hydroquinone \rightleftharpoons Nitrite + Quinone + H ₂ O	variable	50
R00315	ATP + Acetate \rightleftharpoons ADP + Acetyl phosphate	variable	47
R00294	2 Nitric oxide + 2 Ferrocycytochrome c + 2 H ⁺ \rightleftharpoons Nitrous oxide + 2 Ferricytochrome c + H ₂ O	variable	46
R02804	Nitrogen + 2 Ferricytochrome c + H ₂ O \rightleftharpoons Nitrous oxide + 2 Ferrocycytochrome c + 2 H ⁺	variable	42
R00143	Ammonia + NAD ⁺ + H ₂ O \rightleftharpoons Hydroxylamine + NADH + H ⁺	variable	40
R00214	(S)-Malate + NAD ⁺ \rightleftharpoons Pyruvate + CO ₂ + NADH + H ⁺	variable	31
R00919	Propanoyl-CoA + NADP ⁺ \rightleftharpoons Propenoyl-CoA + NADPH + H ⁺	variable	29
R00355	L-Aspartate + 2-Oxoglutarate \rightleftharpoons Oxaloacetate + L-Glutamate	variable	28
R09289	3-Hydroxypropanoate + NADP ⁺ \rightleftharpoons 3-Oxopropanoate + NADPH + H ⁺	variable	25
R00540	Nitrile + 2 H ₂ O \rightleftharpoons Carboxylate + Ammonia	variable	24
R00524	Formamide + H ₂ O \rightleftharpoons Formate + Ammonia	variable	22
R00761	D-Fructose 6-phosphate + Orthophosphate \rightleftharpoons Acetyl phosphate + D-Erythrose 4-phosphate + H ₂ O	variable	21
R01621	D-Xylulose 5-phosphate + Orthophosphate \rightleftharpoons Acetyl phosphate + D-Glyceraldehyde 3-phosphate + H ₂ O	variable	21
R00342	(S)-Malate + NAD ⁺ \rightleftharpoons Oxaloacetate + NADH + H ⁺	variable	19
R07136	(2R)-3-Sulfolactate + NAD ⁺ \rightleftharpoons 3-Sulfopyruvate + NADH + H ⁺	variable	19
R01777	Succinyl-CoA + L-Homoserine \rightleftharpoons CoA + O-Succinyl-L-homoserine	variable	18
R05623	Trimethylamine + NADPH + H ⁺ + Oxygen \rightleftharpoons Trimethylamine N-oxide + NADP ⁺ + H ₂ O	variable	18
R10820	ATP + 3-(Methylthio)propanoate + CoA \rightleftharpoons AMP + Diphosphate + 3-(Methylthio)propanoyl-CoA	variable	18
R02765	(R)-Methylmalonyl-CoA \rightleftharpoons (S)-Methylmalonyl-CoA	variable	17
R01859	ATP + Propanoyl-CoA + HCO ₃ ⁻ \rightleftharpoons ADP + Orthophosphate + (S)-Methylmalonyl-CoA	variable	16
R00606	Methylamine + H ₂ O + 2 Amicyanin \rightleftharpoons Formaldehyde + Ammonia + 2 Reduced amicyanin + 2 H ⁺	variable	14
R01011	ATP + Glycerone \rightleftharpoons ADP + Glycerone phosphate	variable	14
R08058	5,6,7,8-Tetrahydromethanopterin + Formaldehyde \rightleftharpoons 5,10-Methylenetetrahydromethanopterin + H ₂ O	variable	10

Continued from previous page			
R06982	Glutathione + Formaldehyde \rightleftharpoons S-(Hydroxymethyl)glutathione	variable	8
R00736	L-Tyrosine \rightleftharpoons Tyramine + CO ₂	variable	6
R01586	Methylamine + L-Glutamate \rightleftharpoons Ammonia + N-Methyl-L-glutamate	variable	6
R10936	3-(Methylthio)acryloyl-CoA + 2 H ₂ O \rightleftharpoons Acetaldehyde + Methanethiol + CoA + CO ₂	variable	6
R01283	L-Homocysteine + H ₂ O \rightleftharpoons Hydrogen sulfide + Ammonia + 2-Oxobutanoate	variable	3
R01588	Dimethylamine + H ₂ O + Electron-transferring flavoprotein \rightleftharpoons Methylamine + Formaldehyde + Reduced electron-transferring flavoprotein	variable	3
R02511	Trimethylamine + H ₂ O + Electron-transferring flavoprotein \rightleftharpoons Dimethylamine + Formaldehyde + Reduced electron-transferring flavoprotein	variable	3
R00024	D-Ribulose 1,5-bisphosphate + CO ₂ + H ₂ O \rightleftharpoons 2 3-Phospho-D-glycerate	variable	2
R00588	L-Serine + Glyoxylate \rightleftharpoons Hydroxypyruvate + Glycine	variable	2
R05185	16 ATP + Nitrogen + 8 Reduced ferredoxin + 8 H ⁺ + 16 H ₂ O \rightleftharpoons 16 Orthophosphate + 16 ADP + 8 Oxidized ferredoxin + 2 Ammonia + Hydrogen	variable	2
R01063	D-Glyceraldehyde 3-phosphate + Orthophosphate + NADP ⁺ \rightleftharpoons 3-Phospho-D-glyceroyl phosphate + NADPH + H ⁺	variable	1
R02560	Trimethylamine + 2 Ferri cytochrome c + H ₂ O \rightleftharpoons Trimethylamine N-oxide + 2 Ferrocytochrome c + 2 H ⁺	variable	1
R03025	Coenzyme F420 + Hydrogen \rightleftharpoons Reduced coenzyme F420	variable	1
R09282	3-Methylfumaryl-CoA + H ₂ O \rightleftharpoons (3S)-Citramalyl-CoA	variable	1
R09283	2-Methylfumaryl-CoA \rightleftharpoons 3-Methylfumaryl-CoA	variable	1
R09513	Methanesulfonic acid + NADH + H ⁺ + Oxygen \rightleftharpoons Formaldehyde + NAD ⁺ + Sulfite + H ₂ O	variable	1

Appendix G: Number of pathways predicted from Amino Acids to Ammonia in *Pseudomonas*

Organism Name	Amino Acid	Pathway Length						Total
		1	2	3	4	5	6	
<i>Candidatus Pseudomonas adelgestsugas</i> (<i>Adelges tsugae</i>)	L-Proline	-	-	6	-	-	-	7
	L-Glutamate	1	-	-	-	-	-	
<i>Pseudomonas aeruginosa</i> B136-33	L-Proline	-	-	12	-	30	-	52
	L-Glutamate	2	-	5	-	-	-	
<i>-Pseudomonas aeruginosa</i> c7447m	L-Glutamine	-	3	-	-	-	-	52
	L-Proline	-	-	12	-	30	-	
	L-Glutamate	2	-	5	-	-	-	
<i>Pseudomonas aeruginosa</i> DK2	L-Glutamine	-	3	-	-	-	-	52
	L-Proline	-	-	12	-	30	-	
	L-Glutamate	2	-	5	-	-	-	
<i>Pseudomonas aeruginosa</i> LES431	L-Glutamine	-	3	-	-	-	-	52
	L-Proline	-	-	12	-	30	-	
	L-Glutamate	2	-	5	-	-	-	
<i>Pseudomonas aeruginosa</i> LESB58	L-Glutamine	-	3	-	-	-	-	52
	L-Proline	-	-	12	-	30	-	
	L-Glutamate	2	-	5	-	-	-	
<i>Pseudomonas aeruginosa</i> M18	L-Glutamine	-	3	-	-	-	-	52
	L-Proline	-	-	12	-	30	-	
	L-Glutamate	2	-	5	-	-	-	
<i>Pseudomonas aeruginosa</i> MTB-1	L-Glutamine	-	3	-	-	-	-	52
	L-Proline	-	-	12	-	30	-	
	L-Glutamate	2	-	5	-	-	-	
<i>Pseudomonas aeruginosa</i> NCGM 1900	L-Glutamine	-	3	-	-	-	-	52
	L-Proline	-	-	12	-	30	-	
	L-Glutamate	2	-	5	-	-	-	
<i>Pseudomonas aeruginosa</i> NCGM2.S1	L-Glutamine	-	3	-	-	-	-	52
	L-Proline	-	-	12	-	30	-	
	L-Glutamate	2	-	5	-	-	-	
<i>Pseudomonas aeruginosa</i> PA1	L-Glutamine	-	3	-	-	-	-	52
	L-Proline	-	-	12	-	30	-	
	L-Glutamate	2	-	5	-	-	-	
<i>Pseudomonas aeruginosa</i> PA1R	L-Glutamine	-	3	-	-	-	-	52
	L-Proline	-	-	12	-	30	-	
	L-Glutamate	2	-	5	-	-	-	
<i>Pseudomonas aeruginosa</i> PA38182	L-Glutamine	-	3	-	-	-	-	52
	L-Proline	-	-	12	-	30	-	
	L-Glutamate	2	-	5	-	-	-	
<i>Pseudomonas aeruginosa</i> PA7	L-Glutamine	-	3	-	-	-	-	52
	L-Proline	-	-	12	-	30	-	
	L-Glutamate	2	-	5	-	-	-	
<i>Pseudomonas aeruginosa</i> PAO1	L-Glutamine	-	3	-	-	-	-	52
	L-Proline	-	-	12	-	30	-	
	L-Glutamate	2	-	5	-	-	-	

Continued from previous page							
<i>Pseudomonas aeruginosa</i> PAO1-VE13	L-Proline	-	-	12	30	-	52
	L-Glutamate	2	-	5	-	-	
	L-Glutamine	-	3	-	-	-	
<i>Pseudomonas aeruginosa</i> PAO1-VE2	L-Proline	-	-	12	30	-	52
	L-Glutamate	2	-	5	-	-	
	L-Glutamine	-	3	-	-	-	
<i>Pseudomonas aeruginosa</i> PAO581	L-Proline	-	-	12	30	-	52
	L-Glutamate	2	-	5	-	-	
	L-Glutamine	-	3	-	-	-	
<i>Pseudomonas aeruginosa</i> RP73	L-Proline	-	-	12	30	-	52
	L-Glutamate	2	-	5	-	-	
	L-Glutamine	-	3	-	-	-	
<i>Pseudomonas aeruginosa</i> SCV20265	L-Proline	-	-	12	30	-	52
	L-Glutamate	2	-	5	-	-	
	L-Glutamine	-	3	-	-	-	
<i>Pseudomonas aeruginosa</i> UCBPP-PA14	L-Proline	-	-	12	30	-	52
	L-Glutamate	2	-	5	-	-	
	L-Glutamine	-	3	-	-	-	
<i>Pseudomonas aeruginosa</i> YL84	L-Proline	-	-	12	30	-	52
	L-Glutamate	2	-	5	-	-	
	L-Glutamine	-	3	-	-	-	
<i>Pseudomonas alcaligenes</i>	L-Proline	-	-	12	30	-	52
	L-Glutamate	2	-	5	-	-	
	L-Glutamine	-	3	-	-	-	
<i>Pseudomonas alkylphenolica</i>	L-Proline	-	-	12	24	-	43
	L-Glutamate	2	-	4	-	-	
	L-Glutamine	-	1	-	-	-	
<i>Pseudomonas amygdali</i>	L-Proline	-	-	12	-	-	16
	L-Glutamate	2	-	-	-	-	
	L-Glutamine	-	2	-	-	-	
<i>Pseudomonas antarctica</i>	L-Proline	-	-	12	-	24	69
	L-Glutamate	2	-	4	-	-	
	L-Glutamine	-	3	-	-	-	
	L-Arginine	-	-	-	-	24	
<i>Pseudomonas avellanae</i>	L-Proline	-	-	12	-	-	16
	L-Glutamate	2	-	-	-	-	
	L-Glutamine	-	2	-	-	-	
<i>Pseudomonas azotoformans</i>	L-Proline	-	-	6	1	24	42
	L-Glutamate	1	-	4	-	-	
	L-Glutamine	-	2	-	-	-	
<i>Pseudomonas balearica</i>	L-Proline	-	-	12	-	30	52
	L-Glutamate	2	-	5	-	-	
	L-Glutamine	-	3	-	-	-	
<i>Pseudomonas brassicacearum</i> DF41	L-Proline	-	-	16	-	40	66
	L-Glutamate	2	-	5	-	-	
	L-Glutamine	-	3	-	-	-	

Continued from previous page							
<i>Pseudomonas brassicacearum</i> subsp. <i>brassicacearum</i> NFM421	L-Proline	-	-	12	30	-	52
	L-Glutamate	2	-	5	-	-	
	L-Glutamine	-	3	-	-	-	
<i>Pseudomonas chlororaphis</i> PA23	L-Proline	-	-	12	30	-	52
	L-Glutamate	2	-	5	-	-	
	L-Glutamine	-	3	-	-	-	
<i>Pseudomonas chlororaphis</i> PCL1606	L-Proline	-	-	12	-	30	76
	L-Glutamate	2	-	5	-	-	
	L-Glutamine	-	3	-	-	-	
	L-Arginine	-	-	-	-	24	
<i>Pseudomonas chlororaphis</i> subsp. <i>aurantiaca</i>	L-Proline	-	-	12	-	24	43
	L-Glutamate	2	-	4	-	-	
	L-Glutamine	-	1	-	-	-	
<i>Pseudomonas cichorii</i>	L-Proline	-	-	12	-	-	16
	L-Glutamate	2	-	-	-	-	
	L-Glutamine	-	2	-	-	-	
<i>Pseudomonas citronellolis</i>	L-Proline	-	-	12	-	24	45
	L-Glutamate	2	-	4	-	-	
	L-Glutamine	-	3	-	-	-	
<i>Pseudomonas corrugata</i>	L-Proline	-	-	16	-	32	57
	L-Glutamate	2	-	4	-	-	
	L-Glutamine	-	3	-	-	-	
<i>Pseudomonas cremoricolorata</i>	L-Proline	-	-	12	-	-	16
	L-Glutamate	2	-	-	-	-	
	L-Glutamine	-	2	-	-	-	
<i>Pseudomonas entomophila</i>	L-Proline	-	-	12	-	30	64
	L-Glutamate	2	-	5	-	-	
	L-Glutamine	-	3	-	-	-	
	L-Arginine	-	-	-	-	12	
<i>Pseudomonas fluorescens</i> A506	L-Proline	-	-	16	-	40	66
	L-Glutamate	2	-	5	-	-	
	L-Glutamine	-	3	-	-	-	
<i>Pseudomonas fluorescens</i> F113	L-Proline	-	-	12	-	30	52
	L-Glutamate	2	-	5	-	-	
	L-Glutamine	-	3	-	-	-	
<i>Pseudomonas fluorescens</i> Pf0-1	L-Proline	-	-	12	-	30	52
	L-Glutamate	2	-	5	-	-	
	L-Glutamine	-	3	-	-	-	
<i>Pseudomonas fluorescens</i> SBW25	L-Proline	-	-	12	-	30	64
	L-Glutamate	2	-	5	-	-	
	L-Glutamine	-	3	-	-	-	
	L-Arginine	-	-	-	-	12	
<i>Pseudomonas fluorescens</i> UK4	L-Proline	-	-	12	-	30	52
	L-Glutamate	2	-	5	-	-	
	L-Glutamine	-	3	-	-	-	

Continued from previous page							
<i>Pseudomonas fragi</i>	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas frederiksbergensis</i>	L-Proline	-	-	12	-	24	-
	L-Glutamate	2	-	4	-	-	45
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas fulva</i>	L-Proline	-	-	12	-	-	-
	L-Glutamate	2	-	-	-	-	16
	L-Glutamine	-	2	-	-	-	-
<i>Pseudomonas knackmussii</i>	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas koreensis</i>	L-Proline	-	-	12	-	24	-
	L-Glutamate	2	-	4	-	-	45
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas lundensis</i>	L-Proline	-	-	6	-	24	-
	L-Glutamate	1	-	4	-	-	37
	L-Glutamine	-	2	-	-	-	-
<i>Pseudomonas lurida</i>	L-Proline	-	-	12	2	24	4
	L-Glutamate	2	-	4	-	-	51
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas mandelii</i>	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas mendocina</i> NK-01	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas mendocina</i> ymp	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas monteilii</i> SB3078	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas monteilii</i> SB3101	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas orientalis</i>	L-Proline	-	-	12	-	24	-
	L-Glutamate	2	-	4	-	-	45
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas oryzihabitans</i>	L-Proline	-	-	12	-	-	-
	L-Glutamate	2	-	-	-	-	16
	L-Glutamine	-	2	-	-	-	-
<i>Pseudomonas parafulva</i>	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-

Continued from previous page							
<i>Pseudomonas plecoglossicida</i>	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas poae</i>	L-Proline	-	-	12	-	-	-
	L-Glutamate	2	-	-	-	-	-
	L-Glutamine	-	2	-	-	-	28
	L-Arginine	-	-	-	-	12	-
<i>Pseudomonas protegens</i> Cab57	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas protegens</i> CHA0	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas protegens</i> Pf-5	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas pseudoalcaligenes</i>	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas psychrotolerans</i>	L-Proline	-	-	12	-	-	-
	L-Glutamate	2	-	-	-	-	16
	L-Glutamine		2	-	-	-	-
<i>Pseudomonas putida</i> BIRD-1	L-Proline	-	-	12		30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas putida</i> DLL-E4	L-Proline	-	-	12		30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas putida</i> DOT-T1E	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas putida</i> F1	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas putida</i> GB-1	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas putida</i> H8234	L-Proline	-	-	12		30	-
	L-Glutamate	2	-	5	-	-	-
	L-Glutamine	-	3	-	-	-	200
	L-Alanine	-	-	-	-	-	118
	L-Cysteine	-	-	-	-	-	30
<i>Pseudomonas putida</i> HB3267	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	--	-	-	-

Continued from previous page							
<i>Pseudomonas putida</i> KT2440	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas putida</i> NBRC 14164	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas putida</i> ND6	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas putida</i> S16	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas putida</i> W619	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas resinovorans</i>	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas rhizosphaerae</i>	L-Proline	-	-	12	-	-	-
	L-Glutamate	2	-	-	-	-	16
	L-Glutamine	-	2	-	-	-	-
<i>Pseudomonas savastanoi</i> pv. <i>phaseolicola</i> 1448A	L-Proline	-	-	12	-	-	-
	L-Glutamate	2	-	-	-	-	16
	L-Glutamine	-	2	-	-	-	-
<i>Pseudomonas silesiensis</i>	L-Proline	-	-	12	-	24	-
	L-Glutamate	2	-	4	-	-	45
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas simiae</i> PCL1751	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas simiae</i> PICF7	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas soli</i>	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas sp.</i> ATCC 13867	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas sp.</i> CCOS 191	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	-
	L-Glutamine	-	3	-	-	-	64
	L-Arginine	-	-	-	-	12	-
<i>Pseudomonas sp.</i> MRSN12121	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	-
	L-Glutamine	-	3	-	-	-	76
	L-Arginine	-	-	-	-	24	-

Continued from previous page							
<i>Pseudomonas</i> sp. Os17	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas</i> sp. R2A2	L-Proline	-	-	12	-	18	-
	L-Glutamate	2	-	3	-	-	38
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas</i> sp. StFLB209	L-Proline	-	-	12	-	-	-
	L-Glutamate	2	-	-	-	-	16
	L-Glutamine	-	2	-	-	-	-
<i>Pseudomonas</i> sp. TCU-HL1	L-Proline	-	-	12	-	24	-
	L-Glutamate	2	-	4	-	-	45
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas</i> sp. TKP	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas</i> sp. UW4	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas</i> sp. VLB120	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas stutzeri</i> 19SMN4	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas stutzeri</i> 28a24	L-Proline	-	-	12	-	-	-
	L-Glutamate	2	-	-	-	-	16
	L-Glutamine	-	2	-	-	-	-
<i>Pseudomonas stutzeri</i> A1501	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas stutzeri</i> ATCC 17588	L-Proline	-	-	12	-	-	-
	L-Glutamate	2	-	-	-	-	16
	L-Glutamine	-	2	-	-	-	-
<i>Pseudomonas stutzeri</i> CCUG 29243	L-Proline	-	-	12	-	-	-
	L-Glutamate	2	-	-	-	-	16
	L-Glutamine	-	2	-	-	-	-
<i>Pseudomonas stutzeri</i> DSM 10701	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas stutzeri</i> DSM 4166	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas stutzeri</i> RCH2	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas synxantha</i> LBUM223	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-

Continued from previous page							
<i>Pseudomonas syringae</i> CC1557	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas syringae</i> pv. <i>syringae</i> B728a	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas syringae</i> pv. tomato DC3000	L-Proline	-	-	12	-	-	-
	L-Glutamate	2	-	-	-	-	16
	L-Glutamine	-	2	-	-	-	-
<i>Pseudomonas trivialis</i>	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas veronii</i>	L-Proline	-	-	12	-	24	-
	L-Glutamate	2	-	4	-	-	45
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas versuta</i>	L-Proline	-	-	12	-	30	-
	L-Glutamate	2	-	5	-	-	52
	L-Glutamine	-	3	-	-	-	-
<i>Pseudomonas yamanorum</i>	L-Proline	-	-	12	-	24	-
	L-Glutamate	2	-	4	-	-	45
	L-Glutamine	-	3	-	-	-	-

Appendix H: Number of unique pathways predicted from amino acids to Ammonia

Organism Name	Amino Acid	Pathway Length						Total
		1	2	3	4	5	6	
<i>Pseudomonas putida</i> H8234	L-Alanine	-	-	-	-	-	118	148
	L-Cysteine	-	-	-	-	-	30	
<i>Pseudomonas lurida</i>	L-Proline	-	-	-	1	-	-	1

Appendix I: Number of pathways predicted from amino acids to TCA Cycle intermediates in *Pseudomonas*

Organism Name	Amino Acid	Pathway Length						Total
		1	2	3	4	5	6	
<i>Candidatus Pseudomonas adelgestugas</i> (<i>Adelges tsugae</i>)	L-Glutamate	6	-	-	-	-	-	42
	L-Proline	-	-	36	-	-	-	
<i>Pseudomonas aeruginosa</i> B136-33	Glycine	-	-	-	1	8	7	330
	L-Alanine	-	4	-	28	24	24	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	7	-	6	6	
	L-Glutamate	18	-	-	-	-	-	
	L-Glutamine	-	19	-	36	-	-	
	L-Methionine	-	-	-	-	1	7	
	L-Proline	-	-	114	-	-	-	
	L-Tryptophan	-	-	-	1	8	8	
	Glycine	-	-	-	1	8	7	
<i>Pseudomonas aeruginosa</i> c7447m	L-Alanine	-	4	28	-	24	24	340
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	7	-	6	6	
	L-Glutamate	19	-	-	-	-	-	
	L-Glutamine	-	20	-	38	-	-	
	L-Methionine	-	-	-	-	1	7	
	L-Proline	-	-	120	-	-	-	
	L-Tryptophan	-	-	-	1	8	8	
	Glycine	-	-	-	1	5	4	
<i>Pseudomonas aeruginosa</i> DK2	L-Alanine	-	4	16	-	24	24	299
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	4	-	6	6	
	L-Glutamate	18	-	-	-	-	-	
	L-Glutamine	-	19	-	36	-	-	
	L-Methionine	-	-	-	-	1	4	
	L-Proline	-	-	114	-	-	-	
	L-Tryptophan	-	-	-	1	5	4	
	Glycine	-	-	-	1	8	7	
<i>Pseudomonas aeruginosa</i> LES431	L-Alanine	-	4	28	-	24	24	330
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	7	-	6	6	
	L-Glutamate	18	-	-	-	-	-	
	L-Glutamine	-	19	-	36	-	-	
	L-Methionine	-	-	-	-	1	7	
	L-Proline	-	-	114	-	-	-	
	Glycine	-	-	-	1	8	7	

<i>Pseudomonas aeruginosa</i> LESB58	L-Tryptophan	-	-	-	1	8	8	330
	Glycine	-	-	-	1	8	7	
	L-Alanine	-	4	28	-	24	24	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	7	-	6	6	
	L-Glutamate	18	-	-	-	-	-	
	L-Glutamine	-	19	-	36	-	-	
	L-Homoserine	-	-	-	-	-	-	
	L-Methionine	-	-	-	-	1	7	
	L-Proline	-	-	114	-	-	-	
	L-Tryptophan	-	-	-	1	8	8	
<i>Pseudomonas aeruginosa</i> M18	Glycine	-	-	-	1	8	7	330
	L-Alanine	-	4	28	-	24	24	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	7	-	6	6	
	L-Glutamate	18	-	-	-	-	-	
	L-Glutamine	-	19	-	36	-	-	
	L-Methionine	-	-	-	-	1	7	
	L-Proline	-	-	114	-	-	-	
	L-Tryptophan	-	-	-	1	8	8	
<i>Pseudomonas aeruginosa</i> MTB-1	Glycine	-	-	-	1	8	7	340
	L-Alanine	-	4	28	-	24	24	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	7	-	6	6	
	L-Glutamate	19	-	-	-	-	-	
	L-Glutamine	-	20	-	38	-	-	
	L-Methionine	-	-	-	-	1	7	
	L-Proline	-	-	120	-	-	-	
	L-Tryptophan	-	-	-	1	8	8	
<i>Pseudomonas aeruginosa</i> NCGM 1900	Glycine	-	-	-	1	8	7	330
	L-Alanine	-	4	28	-	24	24	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	7	-	6	6	
	L-Glutamate	18	-	-	-	-	-	
	L-Glutamine	-	19	-	36	-	-	
	L-Methionine	-	-	-	-	1	7	
	L-Proline	-	-	114	-	-	-	
	L-Tryptophan	-	-	-	1	8	8	
<i>Pseudomonas aeruginosa</i> NCGM2.S1	Glycine	-	-	-	1	8	7	330
	L-Alanine	-	4	28	-	24	24	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	

	L-Cysteine	-	1	7	-	6	6	
	L-Glutamate	18	-	-	-	-	-	
	L-Glutamine	-	19	-	36	-	-	
	L-Methionine	-	-	-	-	1	7	
	L-Proline	-	-	114	-	-	-	
	L-Tryptophan	-	-	-	1	8	8	
<i>Pseudomonas aeruginosa</i> PA1	Glycine	-	-	-	1	8	7	330
	L-Alanine	-	4	28	-	24	24	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	7	-	6	6	
	L-Glutamate	18	-	-	-	-	-	
	L-Glutamine	-	19	-	36	-	-	
	L-Methionine	-	-	-	-	1	7	
	L-Proline	-	-	114	-	-	-	
	L-Tryptophan	-	-	-	1	8	8	
<i>Pseudomonas aeruginosa</i> PA1R	Glycine	-	-	-	1	8	7	330
	L-Alanine	-	4	28	-	24	24	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	7	-	6	6	
	L-Glutamate	18	-	-	-	-	-	
	L-Glutamine	-	19	-	36	-	-	
	L-Methionine	-	-	-	-	1	7	
	L-Proline	-	-	114	-	-	-	
	L-Tryptophan	-	-	-	1	8	8	
<i>Pseudomonas aeruginosa</i> PA38182	Glycine	-	-	-	1	8	7	330
	L-Alanine	-	4	28	-	24	24	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	7	-	6	6	
	L-Glutamate	18	-	-	-	-	-	
	L-Glutamine	-	19	-	36	-	-	
	L-Methionine	-	-	-	-	1	7	
	L-Proline	-	-	114	-	-	-	
	L-Tryptophan	-	-	-	1	8	8	
<i>Pseudomonas aeruginosa</i> PA7	Glycine	-	-	-	1	8	7	330
	L-Alanine	-	4	28	-	24	24	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	7	-	6	6	
	L-Glutamate	18	-	-	-	-	-	
	L-Glutamine	-	19	-	36	-	-	
	L-Methionine	-	-	-	-	1	7	
	L-Proline	-	-	114	-	-	-	
	L-Tryptophan	-	-	-	1	8	8	

<i>Pseudomonas aeruginosa</i> PAO1	Glycine	-	-	-	1	8	7	320
	L-Alanine	-	4	28	-	24	24	
	L-Asparagine	-	1	-	-	-	-	
	L-Cysteine	-	1	7	-	6	6	
	L-Glutamine	-	20	-	38	-	-	
	L-Methionine	-	-	-	-	1	7	
	L-Proline	-	-	120	-	-	-	
	L-Tryptophan	-	-	-	1	8	8	
<i>Pseudomonas aeruginosa</i> PAO1-VE13	Glycine	-	-	-	1	8	7	320
	L-Alanine	-	4	28	-	24	24	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	-	1	7	-	6	6	
	L-Glutamine	-	20	-	38	-	-	
	L-Methionine	-	-	-	-	1	7	
	L-Proline	-	-	120	-	-	-	
	L-Tryptophan	-	-	-	1	8	8	
<i>Pseudomonas aeruginosa</i> PAO1-VE2	Glycine	-	-	-	1	8	7	320
	L-Alanine	-	4	28	-	24	24	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	-	1	7	-	6	6	
	L-Glutamine	-	20	-	38	-	-	
	L-Methionine	-	-	-	-	1	7	
	L-Proline	-	-	120	-	-	-	
	L-Tryptophan	-	-	-	1	8	8	
<i>Pseudomonas aeruginosa</i> PAO581	Glycine	-	-	-	1	8	7	330
	L-Alanine	-	4	28	-	24	24	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	7	-	6	6	
	L-Glutamate	19	-	-	-	-	-	
	L-Glutamine	-	19	-	36	-	-	
	L-Methionine	-	-	-	-	1	7	
	L-Proline	-	-	114	-	-	-	
	L-Tryptophan	-	-	-	1	8	7	
<i>Pseudomonas aeruginosa</i> RP73	Glycine	-	-	-	1	8	7	330
	L-Alanine	-	4	28	-	24	24	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	7	-	6	6	
	L-Glutamate	18	-	-	-	-	-	
	L-Glutamine	-	19	-	36	-	-	
	L-Methionine	-	-	-	-	1	7	
	L-Proline	-	-	114	-	-	-	
	L-Tryptophan	-	-	-	1	8	8	
<i>Pseudomonas aeruginosa</i> SCV20265	Glycine	-	-	-	1	8	7	330
	L-Alanine	-	4	28	-	24	24	

	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	7	-	6	6	
	L-Glutamate	18	-	-	-	-	-	
	L-Glutamine	-	19	-	36	-	-	
	L-Methionine	-	-	-	-	1	7	
	L-Proline	-	-	114	-	-	-	
	L-Tryptophan	-	-	-	1	8	8	
<i>Pseudomonas aeruginosa</i> UCBPP-PA14	Glycine	-	-	-	1	8	7	312
	L-Alanine	-	4	28	-	24	24	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	7	-	6	6	
	L-Glutamine	-	19	-	36	-	-	
	L-Methionine	-	-	-	-	1	7	
	L-Proline	-	-	114	-	-	-	
	L-Tryptophan	-	-	-	1	8	8	
<i>Pseudomonas aeruginosa</i> YL84	Glycine	-	-	-	1	8	7	330
	L-Alanine	-	4	28	-	24	24	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	7	-	6	6	
	L-Glutamate	18	-	-	-	-	-	
	L-Glutamine	-	19	-	36	-	-	
	L-Methionine	-	-	-	-	1	7	
	L-Proline	-	-	114	-	-	-	
	L-Tryptophan	-	-	-	1	8	8	
<i>Pseudomonas alcaligenes</i>	Glycine	-	-	-	-	-	-	202
	L-Alanine	-	4	20	-	-	8	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Glutamate	17	-	-	-	-	-	
	L-Glutamine	-	17	-	32	-	-	
	L-Proline	-	-	102	-	-	-	
	L-Tryptophan	-	-	-	-	-	-	
<i>Pseudomonas alkylphenolica</i>	Glycine	-	-	-	1	7	6	246
	L-Alanine	-	4	24	-	-	12	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	6	-	-	3	
	L-Glutamate	17	-	-	-	-	-	
	L-Glutamine	-	-	-	34	-	-	
	L-Methionine	-	-	-	-	1	6	
	L-Proline	-	-	108	-	-	-	
	L-Tryptophan	-	-	-	1	7	7	
<i>Pseudomonas amygdalli</i>	L-Alanine	-	4	28	-	24	16	261
	L-Asparagine	-	2	-	-	-	-	

	L-Aspartate	2	-	-	-	-	-	
	L-Cysteine	-	1	7	-	6	4	
	L-Glutamate	20	-	-	-	-	-	
	L-Glutamine	-	21	-	-	-	-	
	L-Proline	-	-	126	-	-	-	
<i>Pseudomonas antarctica</i>	Glycine	-	-	-	1	6	5	540
	L-Alanine	-	4	20		24	24	
	L-Arginine	-	-	-	-	228	-	
	L-Asparagine	-	2	-	-	-	-	
	L-Aspartate	2	-	-	-	-	-	
	L-Cysteine	-	1	5	-	6	6	
	L-Glutamate	18	-	-	-	-	-	
	L-Glutamine	-	19	-	36	-	-	
	L-Methionine	-	-	-	-	1	5	
	L-Proline	-	-	114	-	-	-	
	L-Tryptophan	-	-	-	1	6	6	
<i>Pseudomonas avellanae</i>	L-Alanine	-	4	28		24	16	261
	L-Asparagine	-	2	-	-	-	-	
	L-Aspartate	2	-	-	-	-	-	
	L-Cysteine	-	1	7	-	6	4	
	L-Glutamate	20	-	-	-	-	-	
	L-Glutamine	-	21	-	-	-	-	
	L-Proline	-	-	126	-			
<i>Pseudomonas azotoformans</i>	Glycine	-	-	-	1	6	5	302
	L-Alanine	-	4	20	-	16	16	
	L-Asparagine	-	2	-	-	-	-	
	L-Aspartate	2	-	-	-	-	-	
	L-Cysteine	-	1	5	-	4	4	
	L-Glutamate	19	-	-	-	-	-	
	L-Glutamine	-	20	-	19	-	-	
	L-Methionine	-	-	-	-	1	5	
	L-Proline	-	-	120	19	-	-	
	L-Tryptophan	-	-	-	1	6	6	
<i>Pseudomonas balearica</i>	L-Alanine	-	4	20	-	-	8	192
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Glutamate	15	-	-	-	-	-	
	L-Glutamine	-	16	-	30	-	-	
	L-Proline	-	-	96	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
<i>Pseudomonas brassicacearum</i> DF41	L-Alanine	-	5	25	-	30	20	354
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	5	-	6	4	
	L-Glutamate	20	-	-	-	-	-	
	L-Glutamine	-	21	-	40	-	-	

<i>Pseudomonas brassicacearum</i> subsp. <i>brassicacearum</i> NFM421	L-Homoserine	-	-	-	-	1	5	306
	L-Proline	-	-	168	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
	L-Alanine	-	5	25	-	30	20	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	5	-	6	4	
	L-Glutamate	20	-	-	-	-	-	
	L-Glutamine	-	21	-	40	-	-	
	L-Proline	-	-	126	-	-	-	
<i>Pseudomonas chlororaphis</i> PA23	L-Tryptophan	-	-	-	-	-	1	280
	Glycine	-	-	-	1	6	5	
	L-Alanine	-	4	20	-	-	8	
	L-Arginine	-	-	-	-	-	-	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	5	-	-	2	
	L-Glutamate	20	-	-	-	-	-	
	L-Glutamine	-	21	-	40	-	-	
	L-Methionine	-	-	-	-	1	5	
<i>Pseudomonas chlororaphis</i> PCL1606	L-Proline	-	-	126	-	-	-	528
	L-Tryptophan	-	-	-	1	6	6	
	Glycine	-	-	-	1	6	5	
	L-Alanine	-	4	20	-	24	16	
	L-Arginine	-	-	-	-	228	-	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	5	-	6	4	
	L-Glutamate	18	-	-	-	-	-	
	L-Glutamine	-	19	-	36	-	-	
<i>Pseudomonas chlororaphis</i> subsp. <i>aurantiaca</i>	L-Methionine	-	-	-	-	1	5	240
	L-Proline	-	-	114	-	-	-	
	L-Tryptophan	-	-	-	1	6	6	
	Glycine	-	-	-	1	6	5	
	L-Alanine	-	4	20	-	-	8	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	5	-	-	2	
	L-Glutamate	18	-	-	-	-	-	
	L-Glutamine	-	-	-	36	-	-	
	L-Methionine	-	-	-	-	1	5	
<i>Pseudomonas cichorii</i>	L-Proline	-	-	114	-	-	-	230
	L-Tryptophan	-	-	-	1	6	6	
	Glycine	-	-	-	-	-	-	
	L-Alanine	-	4	28	-	-	8	
	L-Arginine	-	-	-	-	-	-	
	L-Asparagine	-	2	-	-	-	-	

	L-Aspartate	2	-	-	-	-	-	
	L-Cysteine	-	1	7	-	-	2	
	L-Glutamate	21	-	-	-	-	-	
	L-Glutamine	-	22	-	-	-	-	
	L-Proline	-	-	132	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
<i>Pseudomonas citronellolis</i>	Glycine	-	-	-	1	8	7	269
	L-Alanine	-	4	28	-	-	8	
	L-Arginine	-	-	-	-	-	-	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	7	-	-	2	
	L-Glutamate	16	-	-	-	-	-	
	L-Glutamine	-	18	-	34	-	-	
	L-Methionine	-	-	-	-	1	7	
	L-Proline	-	-	108	-	-	-	
	L-Tryptophan	-	-	-	1	8	8	
<i>Pseudomonas corrugata</i>	L-Alanine	-	4	20	-	24	16	332
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine		1	5	-	6	4	
	L-Glutamate	20	-	-	-	-	-	
	L-Glutamine	-	21	-	40	-	-	
	L-Proline	-	-	168	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
<i>Pseudomonas cremoricolorata</i>	L-Alanine	-	4	20	-	24	16	234
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	5	-	6	4	
	L-Glutamate	18		-	-	-	-	
	L-Glutamine	-	19	-	-	-	-	
	L-Proline	-	-	114	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
<i>Pseudomonas entomophila</i>	L-Alanine	-	4	24	-	-	12	362
	L-Arginine	-		-	-	120	-	
	L-Asparagine	-	2	-	-	-	-	
	L-Aspartate	2	-	-	-	-	-	
	L-Glutamate	19	-	-	-	-	-	
	L-Glutamine	-	20	-	38	-	-	
	L-Proline	-	-	120	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
<i>Pseudomonas fluorescens</i> A506	Glycine	-	-	-	1	6	5	364
	L-Alanine	-	4	20	-	24	16	
	L-Asparagine	-	2	-	-	-	-	
	L-Aspartate	2	-	-	-	-	-	
	L-Cysteine	-	1	5	-	6	4	

	L-Glutamate	20	-	-	-	-	-	
	L-Glutamine	-	21	-	40	-	-	
	L-Methionine	-	-	-	-	1	5	
	L-Proline	-	-	168	-	-	-	
	L-Tryptophan	-	-	-	1	6	6	
<i>Pseudomonas fluorescens</i> F113	L-Alanine	-	4	20	-	24	24	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	5	-	6	6	
	L-Glutamate	19	-	-	-	-	-	296
	L-Glutamine	-	20	-	38	-	-	
	L-Homoserine	-	-	-	-	1	5	
	L-Proline	-	-	120	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
	Glycine	-	-	-	1	4	3	
<i>Pseudomonas fluorescens</i> Pf0-1	L-Alanine	-	4	12	-	24	16	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	3	-	6	4	281
	L-Glutamate	19	-	-	-	-	-	
	L-Glutamine	-	19	-	36	-	-	
	L-Methionine	-	-	-	-	1	3	
	L-Proline	-	-	114	-	-	-	
	L-Tryptophan	-	-	-	1	4	4	
	Glycine	-	-	-	1	5	-	
<i>Pseudomonas fluorescens</i> SBW25	L-Alanine	-	4	20	-	24	16	
	L-Arginine	-	-	-	-	132	-	
	L-Asparagine	-	2	-	-	-	-	
	L-Aspartate	2	-	-	-	-	-	
	L-Cysteine	-	1	5	-	6	4	446
	L-Glutamate	21	-	-	-	-	-	
	L-Glutamine	-	22	-	42	-	-	
	L-Proline	-	-	132	-	-	-	
	L-Tryptophan	-	-	-	1	5	1	
	L-Alanine	-	5	25	-	-	10	
<i>Pseudomonas fluorescens</i> UK4	L-Arginine	-	-	-	-	-	-	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	5	-	-	2	238
	L-Glutamate	18	-	-	-	-	-	
	L-Glutamine	-	19	-	36	-	-	
	L-Proline	-	-	114	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
	L-Alanine	-	4	20	-	24	32	
	L-Asparagine	-	2	-	-	-	-	292
<i>Pseudomonas fragi</i>	L-Aspartate	2	-	-	-	-	-	

	L-Cysteine	-	1	5	-	6	8	
	L-Glutamate	19	-	-	-	-	-	
	L-Glutamine	-	19	-	36	-	-	
	L-Proline	-	-	114	-	-	-	
<i>Pseudomonas frederiksbergensis</i>	Glycine	-	-	-	1	6	5	240
	L-Alanine	-	4	20	-	-	8	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	5	-	-	2	
	L-Glutamate	16	-	-	-	-	-	
	L-Glutamine	-	17	-	32	-	-	
	L-Methionine	-	-	-	-	1	5	
	L-Proline	-	-	102	-	-	-	
	L-Tryptophan	-	-	-	1	6	6	
<i>Pseudomonas fulva</i>	L-Alanine	-	4	20	-	16	24	225
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Glutamate	19	-	-	-	-	-	
	L-Glutamine	-	20	-	-	-	-	
	L-Proline	-	-	120	-	-	-	
<i>Pseudomonas knackmussii</i>	Glycine	-	-	-	1	6	5	257
	L-Alanine	-	4	20	-	-	8	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	5	-	-	2	
	L-Glutamate	18	-	-	-	-	-	
	L-Glutamine	-	18	-	34	-	-	
	L-Homoserine	-	-	-	-	1	5	
	L-Methionine	-	-	-	-	1	5	
	L-Proline	-	-	108	-	-	-	
	L-Tryptophan	-	-	-	1	6	6	
<i>Pseudomonas koreensis</i>	Glycine	-	-	-	1	6	5	296
	L-Alanine	-	5	25	-	30	20	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	5	-	6	4	
	L-Glutamate	16	-	-	-	-	-	
	L-Glutamine	-	17	-	32	-	-	
	L-Methionine	-	-	-	-	1	5	
	L-Proline	-	-	102	-	-	-	
	L-Tryptophan	-	-	-	1	6	6	
<i>Pseudomonas lundensis</i>	L-Alanine	-	4	20	-	24	24	234
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	5	-	6	6	
	L-Glutamate	15	-	-	-	-	-	

	L-Glutamine	-	16	-	15	-	-	
	L-Proline	-	-	96	-	-	-	
<i>Pseudomonas lurida</i>	Glycine	-	-	-	1	6	5	307
	L-Alanine	-	4	20	-	24	16	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	5	-	6	4	
	L-Glutamate	17	-	-	-	-	-	
	L-Glutamine	-	18	-	34	-	-	
	L-Methionine	-	-	-	-	1	5	
	L-Proline	-	-	108	17	-	-	
	L-Tryptophan	-	-	-	1	6	6	
	Glycine	-	-	-	1	6	5	
<i>Pseudomonas mandelii</i>	L-Alanine	-	5	25	-	30	30	328
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	5	-	6	6	
	L-Glutamate	18	-	-	-	-	-	
	L-Glutamine	-	19	-	36	-	-	
	L-Methionine	-	-	-	-	1	5	
	L-Proline	-	-	114	-	-	-	
	L-Tryptophan	-	-	-	1	6	6	
	L-Alanine	-	5	25	-	30	30	
<i>Pseudomonas mendocina</i> NK-01	L-Asparagine	-	1	-	-	-	-	269
	L-Aspartate	1	-	-	-	-	-	
	L-Glutamate	17	-	-	-	-	-	
	L-Glutamine	-	18	-	34	-	-	
	L-Proline	-	-	108	-	-	-	
<i>Pseudomonas mendocina</i> ymp	L-Alanine	-	5	25	-	20	30	269
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Glutamate	18	-	-	-	-	-	
	L-Glutamine	-	19	-	36	-	-	
	L-Proline	-	-	114	-	-	-	
<i>Pseudomonas monteilii</i> SB3078	L-Alanine	-	4	24	-	36	48	327
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	6	-	9	12	
	L-Glutamate	17	-	-	-	-	-	
	L-Glutamine	-	18	-	34	-	-	
	L-Homoserine	-	-	-	-	1	6	
	L-Proline	-	-	108	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
<i>Pseudomonas monteilii</i> SB3101	L-Alanine	-	4	24	-	36	48	327
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	

	L-Cysteine	-	1	6	-	9	12	
	L-Glutamate	17		-	-	-	-	
	L-Glutamine	-	18	-	34	-	-	
	L-Homoserine	-	-	-	-	1	6	
	L-Proline	-	-	108	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
<i>Pseudomonas orientalis</i>	Glycine	-	-	-	1	6	5	
	L-Alanine	-	4	20	-	24	16	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	5	-	6	4	
	L-Glutamate	18	-	-	-	-	-	300
	L-Glutamine	-	19	-	36	-	-	
	L-Methionine	-	-	-	-	1	5	
	L-Proline	-	-	114	-	-	-	
	L-Tryptophan	-	-	-	1	6	6	
<i>Pseudomonas oryzihabitans</i>	L-Alanine	-	4	20	-	24	16	
	L-Asparagine	-	2	-	-	-	-	
	L-Aspartate	2	-	-	-	-	-	
	L-Glutamate	20	-	-	-	-	-	235
	L-Glutamine	-	21	-	-	-	-	
	L-Proline	-	-	126	-	-	-	
<i>Pseudomonas parafulva</i>	L-Alanine	-	4	20	--	24	16	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	5	-	6	4	
	L-Glutamate	18	-	-	-	-	-	270
	L-Glutamine	-	19	-	36	-	-	
	L-Proline	-	-	114	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
<i>Pseudomonas plecoglossicida</i>	L-Alanine	-	5	30	-	45	60	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	6	-	9	12	
	L-Glutamate	17	-	-	-	-	-	355
	L-Glutamine	-	18	-	34	-	-	
	L-Homoserine	-	-	-	-	1	6	
	L-Proline	-	-	108	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
<i>Pseudomonas poae</i>	Glycine	-	-	-	1	6	5	
	L-Alanine	-	4	20	-	24	16	
	L-Arginine	-	-	-	-	120	-	
	L-Asparagine	-	1	-	-	-	-	392
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	5	-	6	4	
	L-Glutamate	19	-	-	-	-	-	

	L-Glutamine	-	20	-	-	-	-	
	L-Methionine	-	-	-	-	1	5	
	L-Proline	-	-	120	-	-	-	
	L-Tryptophan	-	-	-	1	6	6	
<i>Pseudomonas protegens</i> Cab57	L-Alanine	-	4	20	-	24	16	292
	L-Asparagine	-	2	-	-	-	-	
	L-Aspartate	2	-	-	-	-	-	
	L-Cysteine	-	1	5	-	6	4	
	L-Glutamate	20	-	-	-	-	-	
	L-Glutamine	-	21	-	40	-	-	
	L-Proline	-	-	126	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
<i>Pseudomonas protegens</i> CHA0	L-Alanine	-	4	20	-	24	16	292
	L-Asparagine	-	2	-	-	-	-	
	L-Aspartate	2	-	-	-	-	-	
	L-Cysteine	-	1	5	-	6	4	
	L-Glutamate	20	-	-	-	-	-	
	L-Glutamine	-	21	-	40	-	-	
	L-Proline	-	-	126	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
<i>Pseudomonas protegens</i> Pf-5	L-Alanine	-	4	12	-	24	16	282
	L-Asparagine	-	2	-	-	-	-	
	L-Aspartate	2	-	-	-	-	-	
	L-Cysteine	-	1	3	-	6	4	
	L-Glutamate	20	-	-	-	-	-	
	L-Glutamine	-	21	-	40	-	-	
	L-Proline	-	-	126	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
<i>Pseudomonas pseudoalcaligenes</i>	L-Alanine	-	5	25	-	-	10	220
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Glutamate	18	-	-	-	-	-	
	L-Glutamine	-	18	-	34	-	-	
	L-Proline	-	-	108	-	-	-	
<i>Pseudomonas psychrotolerans</i>	L-Alanine	-	5	25	-	40	20	279
	L-Asparagine	-	2	-	-	-	-	
	L-Aspartate	2	-	-	-	-	-	
	L-Cysteine	-	1	5	-	8	4	
	L-Glutamate	20	-	-	-	-	-	
	L-Glutamine	-	21	-	-	-	-	
	L-Proline	-	-	126	-	-	-	
<i>Pseudomonas putida</i> BIRD-1	L-Alanine	-	4	24	-	-	12	237
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	6	-	-	3	
	L-Glutamate	17	-	-	-	-	-	

	L-Glutamine	-	18	-	34	-	-	
	L-Homoserine	-	-	-	-	1	6	
	L-Proline	-	-	108	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
<i>Pseudomonas putida</i> DLL-E4	L-Alanine	-	4	24	-	36	24	307
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	6	-	9	6	
	L-Glutamate	18	-	-	-	-	-	
	L-Glutamine	-	19	-	36	-	-	
	L-Homoserine	-	-	-	-	1	6	
	L-Proline	-	-	114	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
<i>Pseudomonas putida</i> DOT-T1E	L-Alanine	-	4	24	-	24	48	334
	L-Asparagine	-	2	-	-	-	-	
	L-Aspartate	2	-	-	-	-	-	
	L-Cysteine	-	1	6	-	6	12	
	L-Glutamate	19	-	-	-	-	-	
	L-Glutamine	-	20	-	38	-	-	
	L-Homoserine	-	-	-	-	1	6	
	L-Proline	-	-	120	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
<i>Pseudomonas putida</i> F1	L-Alanine	-	4	24	-	24	48	312
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	6	-	6	12	
	L-Glutamate	17	-	-	-	-	-	
	L-Glutamine	-	18	-	34	-	-	
	L-Homoserine	-	-	-	-	1	6	
	L-Proline	-	-	108	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
<i>Pseudomonas putida</i> GB-1	L-Alanine	-	4	24	-	24	24	275
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	6	-	6	6	
	L-Glutamate	17	-	-	-	-	-	
	L-Glutamine	-	18	-	34	-	-	
	L-Proline	-	-	108	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
<i>Pseudomonas putida</i> H8234	L-Alanine	-	4	24	4	36	52	352
	L-Asparagine	-	2	-	-	-	-	
	L-Aspartate	2	-	-	-	-	-	
	L-Cysteine	-	1	6	1	9	13	
	L-Glutamate	19	-	-	-	-	-	
	L-Glutamine	-	20	-	38	-	-	
	L-Proline	-	-	120	-	-	-	

	L-Tryptophan	-	-	-	-	-	1	
<i>Pseudomonas putida</i> HB3267	L-Alanine	-	5	30	-	-	15	257
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	6	-	-	3	
	L-Glutamate	18	-	-	-	-	-	
	L-Glutamine	-	19	-	36	-	-	
	L-Homoserine	-	-	-	-	1	6	
	L-Proline	-	-	114	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
<i>Pseudomonas putida</i> KT2440	L-Alanine	-	5	15	-	-	15	225
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	3	-	-	3	
	L-Glutamate	17	-	-	-	-	-	
	L-Glutamine	-	18	-	34	-	-	
	L-Homoserine	-	-	-	-	1	3	
	L-Proline	-	-	108	-	-	-	
<i>Pseudomonas putida</i> NBRC 14164	L-Alanine	-	4	24	-	36	48	342
	L-Asparagine	-	2	-	-	-	-	
	L-Aspartate	2	-	-	-	-	-	
	L-Cysteine	-	1	6	-	9	12	
	L-Glutamate	19	-	-	-	-	-	
	L-Glutamine	-	20	-	38	-	-	
	L-Proline	-	-	120	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
<i>Pseudomonas putida</i> ND6	L-Alanine	-	4	24	-	24	24	282
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	6	-	6	6	
	L-Glutamate	17	-	-	-	-	-	
	L-Glutamine	-	18	-	34	-	-	
	L-Homoserine	-	-	-	-	1	6	
	L-Proline	-	-	108	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
<i>Pseudomonas putida</i> S16	L-Alanine	-	4	24	-	36	24	307
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	6	-	9	6	
	L-Glutamate	18	-	-	-	-	-	
	L-Glutamine	-	19	-	36	-	-	
	L-Homoserine	-	-	-	-	1	6	
	L-Proline	-	-	114	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
<i>Pseudomonas putida</i> W619	L-Alanine	-	4	20	-	16	32	270
	L-Asparagine	-	1	-	-	-	-	

	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	5	-	4	8	
	L-Glutamate	17	-	-	-	-	-	
	L-Glutamine	-	18	-	34	-	-	
	L-Proline	-	-	108	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
<i>Pseudomonas resinovorans</i>	Glycine	-	-	-	1	7	6	
	L-Alanine	-	6	36	-	-	18	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	6	-	-	3	
	L-Glutamate	17	-	-	-	-	-	
	L-Glutamine	-	18	-	34	-	-	
	L-Methionine	-	-	-	-	1	6	
	L-Proline	-	-	108	-	-	-	
	L-Tryptophan	-	-	-	1	7	7	
<i>Pseudomonas rhizosphaerae</i>	Glycine	-	-	4	4	12	4	
	L-Alanine	-	5	35	-	30	20	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	7	-	6	4	
	L-Glutamate	20	-	-	-	-	-	
	L-Glutamine	-	20	-	-	-	-	
	L-Proline	-	-	120	-	-	-	
	L-Tryptophan	-	-	-	-	4	8	
<i>Pseudomonas savastanoi</i> pv. <i>phaseolicola</i> 1448A	L-Alanine	-	4	28		24	16	
	L-Asparagine	-	2	-	-	-	-	
	L-Aspartate	2	-	-	-	-	-	
	L-Cysteine	-	1	7		6	4	
	L-Glutamate	22	-	-	-	-	-	
	L-Glutamine	-	23	-	-	-	-	
	L-Proline	-	-	138	-	-	-	
<i>Pseudomonas silesiensis</i>	Glycine	-	-	-	1	6	5	
	L-Alanine	-	5	25	-	-	10	
	L-Asparagine	-	2	-	-	-	-	
	L-Aspartate	2	-	-	-	-	-	
	L-Cysteine	-	1	5	-	-	2	
	L-Glutamate	18	-	-	-	-	-	
	L-Glutamine	-	19	-	36	-	-	
	L-Homoserine	-	-	-	-	2	10	
	L-Methionine	-	-	-	-	1	5	
	L-Proline	-	-	114	-	-	-	
	L-Tryptophan	-	-	-	1	6	6	
<i>Pseudomonas simiae</i> PCL1751	L-Alanine	-	4	20	-	24	16	
	L-Asparagine	-	2	-	-	-	-	
	L-Aspartate	2	-	-	-	-	-	

	L-Cysteine	-	1	5	-	6	4	
	L-Glutamate	19	-	-	-	-	-	
	L-Glutamine	-	20	-	38	-	-	
	L-Proline	-	-	120	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
<i>Pseudomonas simiae</i> PICF7	L-Alanine	-	4	20	-	24	16	282
	L-Asparagine	-	2	-	-	-	-	
	L-Aspartate	2	-	-	-	-	-	
	L-Cysteine	-	1	5	-	6	4	
	L-Glutamate	19	-	-	-	-	-	
	L-Glutamine	-	20	-	38	-	-	
	L-Proline	-	-	120	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
<i>Pseudomonas soli</i>	L-Alanine	-	5	25	-	30	40	322
	L-Asparagine	-	2	-	-	-	-	
	L-Aspartate	2	-	-	-	-	-	
	L-Cysteine	-	1	5	-	6	8	
	L-Glutamate	19	-	-	-	-	-	
	L-Glutamine	-	20	-	38	-	-	
	L-Proline	-	-	120	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
<i>Pseudomonas sp.</i> ATCC 13867	Glycine	-	-	-	1	6	5	270
	L-Alanine	-	4	20	-	-	8	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	5	-	-	2	
	L-Glutamate	19	-	-	-	-	-	
	L-Glutamine	-	20	-	38	-	-	
	L-Methionine	-	-	-	-	1	5	
	L-Proline	-	-	120	-	-	-	
	L-Tryptophan	-	-	-	1	6	6	
<i>Pseudomonas sp.</i> CCOS 191	L-Alanine	-	5	25	-	-	10	336
	L-Arginine	-	-	-	-	108	-	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	5	-	-	2	
	L-Glutamate	17	-	-	-	-	-	
	L-Glutamine	-	18	-	34	-	-	
	L-Proline	-	-	108	-	-	-	
<i>Pseudomonas sp.</i> MRSN12121	L-Tryptophan	-	-	-	-	-	1	528
	Glycine	-	-	-	1	6	5	
	L-Alanine	-	4	20	-	24	16	
	L-Arginine	-	-	-	-	228	-	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	5	-	6	4	

	L-Glutamate	18	-	-	-	-	-	
	L-Glutamine	-	19	-	36	-	-	
	L-Methionine	-	-	-	-	1	5	
	L-Proline	-	-	114	-	-	-	
	L-Tryptophan	-	-	-	1	6	6	
<i>Pseudomonas sp. Os17</i>	L-Alanine	-	4	20	-	-	8	262
	L-Asparagine	-	2	-	-	-	-	
	L-Aspartate	2	-	-	-	-	-	
	L-Cysteine	-	1	5	-	-	2	
	L-Glutamate	21	-	-	-	-	-	
	L-Glutamine	-	22	-	42	-	-	
	L-Proline	-	-	132	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
<i>Pseudomonas sp. R2A2</i>	L-Alanine	-	4	20	-	24	24	241
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Glutamate	15	-	-	-	-	-	
	L-Glutamine	-	17	-	32	-	-	
	L-Proline	-	-	102	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
<i>Pseudomonas sp. StFLB209</i>	L-Alanine	-	4	28	-	24	32	282
	L-Asparagine	-	2	-	-	-	-	
	L-Aspartate	2	-	-	-	-	-	
	L-Cysteine	-	1	7	-	6	8	
	L-Glutamate	20	-	-	-	-	-	
	L-Glutamine	-	21	-	-	-	-	
	L-Proline	-	-	126	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
<i>Pseudomonas sp. TCU-HL1</i>	Glycine	-	-	-	1	7	6	279
	L-Alanine	-	5	30	-	-	15	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	6	-	-	3	
	L-Glutamate	16	-	-	-	-	-	
	L-Glutamine	-	17	-	32	-	-	
	L-Homoserine	-	-	-	-	2	12	
	L-Methionine	-	-	-	-	1	6	
	L-Proline	-	-	102	-	-	-	
	L-Tryptophan	-	-	-	1	7	7	
<i>Pseudomonas sp. TKP</i>	Glycine	-	-	-	1	6	5	364
	L-Alanine	-	6	30	-	36	24	
	L-Asparagine	-	2	-	-	-	-	
	L-Aspartate	2	-	-	-	-	-	
	L-Cysteine	-	1	5	-	6	4	
	L-Glutamate	21	-	-	-	-	-	
	L-Glutamine	-	22	-	42	-	-	

<i>Pseudomonas sp.</i> UW4	L-Methionine	-	-	-	-	1	5	279
	L-Proline	-	-	132	-	-	-	
	L-Tryptophan	-	-	-	1	6	6	
	Glycine	-	-	-	1	6	5	
	L-Alanine	-	4	20	-	-	8	
	L-Asparagine	-	2	-	-	-	-	
	L-Aspartate	2	-	-	-	-	-	
	L-Cysteine	-	1	5	-	-	2	
	L-Glutamate	20	-	-	-	-	-	
	L-Glutamine	-	20	-	38	-	-	
	L-Homoserine	-	-	-	-	1	5	
	L-Methionine	-	-	-	-	1	5	
	L-Proline	-	-	120	-	-	-	
	L-Tryptophan	-	-	-	1	6	6	
	L-Alanine	-	4	24	-	-	12	
	L-Asparagine	-	1	-	-	-	-	
<i>Pseudomonas sp.</i> VLB120	L-Aspartate	1	-	-	-	-	-	230
	L-Cysteine	-	1	6	-	-	3	
	L-Glutamate	17	-	-	-	-	-	
	L-Glutamine	-	18	-	34	-	-	
	L-Proline	-	-	108	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
	L-Alanine	-	4	20	-	-	8	
	L-Asparagine	-	1	-	-	-	-	
<i>Pseudomonas stutzeri</i> 19SMN4	L-Aspartate	1	-	-	-	-	-	212
	L-Glutamate	17	-	-	-	-	-	
	L-Glutamine	-	18	-	34	-	-	
	L-Proline	-	-	108	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
	Glycine	-	-	4	2	6	8	
	L-Alanine	-	5	25	-	30	30	
<i>Pseudomonas stutzeri</i> 28a24	L-Asparagine	-	1	-	-	-	-	273
	L-Aspartate	1	-	-	-	-	-	
	L-Glutamate	18	-	-	-	-	-	
	L-Glutamine	-	19	-	-	-	-	
	L-Proline	-	-	114	-	-	-	
	L-Tryptophan	-	-	-	-	4	6	
	L-Alanine	-	4	20	-	24	24	
<i>Pseudomonas stutzeri</i> A1501	L-Asparagine	-	1	-	-	-	-	252
	L-Aspartate	1	-	-	-	-	-	
	L-Glutamate	17	-	-	-	-	-	
	L-Glutamine	-	18	-	34	-	-	
	L-Proline	-	-	108	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
	L-Alanine	-	4	20	-	24	24	
<i>Pseudomonas stutzeri</i> ATCC 17588	L-Asparagine	-	1	-	-	-	-	218

	L-Aspartate	1	-	-	-	-	-	
	L-Glutamate	17	-	-	-	-	-	
	L-Glutamine	-	18	-	-	-	-	
	L-Proline	-	-	108	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
<i>Pseudomonas stutzeri</i> CCUG 29243	L-Alanine	-	4	20	-	-	8	178
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Glutamate	17	-	-	-	-	-	
	L-Glutamine	-	18	-	-	-	-	
	L-Proline	-	-	108	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
<i>Pseudomonas stutzeri</i> DSM 10701	L-Alanine	-	4	20	-	24	24	251
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Glutamate	17	-	-	-	-	-	
	L-Glutamine	-	18	-	34	-	-	
	L-Proline	-	-	108	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
<i>Pseudomonas stutzeri</i> DSM 4166	L-Alanine	-	4	20	-	24	24	252
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Glutamate	17	-	-	-	-	-	
	L-Glutamine	-	18	-	34	-	-	
	L-Proline	-	-	108	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
<i>Pseudomonas stutzeri</i> RCH2	L-Alanine	-	4	20	-	-	8	212
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Glutamate	17	-	-	-	-	-	
	L-Glutamine	-	18	-	34	-	-	
	L-Proline	-	-	108	-	-	-	
	L-Tryptophan	-	-	-	-	-	1	
<i>Pseudomonas synxantha</i> LBUM223	Glycine	-	-	-	1	6	5	290
	L-Alanine	-	4	20	-	24	16	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	5	-	6	4	
	L-Glutamate	17	-	-	-	-	-	
	L-Glutamine	-	18	-	34	-	-	
	L-Methionine	-	-	-	-	1	5	
	L-Proline	-	-	108	-	-	-	
	L-Tryptophan	-	-	-	1	6	6	
<i>Pseudomonas syringae</i> CC1557	L-Alanine	-	4	28	-	24	16	299
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	7	-	6	4	

<i>Pseudomonas syringae</i> pv. <i>syringae</i> B728a	L-Glutamate	20	-	-	-	-	-	299
	L-Glutamine	-	21	-	40	-	-	
	L-Proline	-	-	126	-	-	-	
	L-Alanine	-	4	28	-	24	16	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	7	-	6	4	
	L-Glutamate	20	-	-	-	-	-	
	L-Glutamine	-	21	-	40	-	-	
	L-Proline	-	-	126	-	-	-	
<i>Pseudomonas syringae</i> pv. tomato DC3000	L-Alanine	-	4	20	-	24	24	269
	L-Asparagine	-	2	-	-	-	-	
	L-Aspartate	2	-	-	-	-	-	
	L-Cysteine	-	1	5	-	6	6	
	L-Glutamate	21	-	-	-	-	-	
	L-Glutamine	-	22	-	-	-	-	
	L-Proline	-	-	132	-	-	-	
<i>Pseudomonas trivialis</i>	Glycine	-	-	-	1	6	5	310
	L-Alanine	-	4	20	-	24	16	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	5	-	6	4	
	L-Glutamate	19	-	-	-	-	-	
	L-Glutamine	-	20	-	38	-	-	
	L-Methionine	-	-	-	-	1	5	
	L-Proline	-	-	120	-	-	-	
	L-Tryptophan	-	-	-	1	6	6	
<i>Pseudomonas veronii</i>	Glycine	-	-	-	1	6	5	340
	L-Alanine	-	4	20	-	24	40	
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	5	-	6	10	
	L-Glutamate	19	-	-	-	-	-	
	L-Glutamine	-	20	-	38	-	-	
	L-Methionine	-	-	-	-	1	5	
	L-Proline	-	-	120	-	-	-	
	L-Tryptophan	-	-	-	1	6	6	
<i>Pseudomonas versuta</i>	L-Alanine	-	4	16	-	24	32	265
	L-Asparagine	-	1	-	-	-	-	
	L-Aspartate	1	-	-	-	-	-	
	L-Cysteine	-	1	4	-	6	8	
	L-Glutamate	17	-	-	-	-	-	
	L-Glutamine	-	17	-	32	-	-	
	L-Proline	-	-	102	-	-	-	
<i>Pseudomonas yamanorum</i>	Glycine	-	-	-	1	6	5	302
	L-Alanine	-	4	20	-	24	16	

L-Asparagine	-	2	-	-	-	-
L-Aspartate	2	-	-	-	-	-
L-Cysteine	-	1	5	-	6	4
L-Glutamate	18	-	-	-	-	-
L-Glutamine	-	19	-	36	-	-
L-Methionine	-	-	-	-	1	5
L-Proline	-	-	114	-	-	-
L-Tryptophan	-	-	-	1	6	6

Appendix J: Number of unique pathways predicted from different amino acids to four metabolites of TCA Cycle in *Pseudomonas*

Organism Name	Amino Acid	Pathway Length						Total
		1	2	3	4	5	6	
<i>Pseudomonas antarctica</i>	L-Arginine	-	-	-	-	24	-	24
	Glycine	-	-	-	2	6	-	
<i>Pseudomonas rhizosphaerae</i>	L-Alanine	-	-	2	-	-	-	12
	L-Tryptophan	-	-	-	-	-	2	
<i>Pseudomonas putida</i> H8234	L-Alanine	-	-	-	4	-	4	10
	L-Cysteine	-	-	-	1	-	1	
	L-Glutamate	1	-	-	-	-	-	
<i>Pseudomonas orientalis</i>	L-Glutamine	-	-	-	2	-	-	9
	L-Proline	-	-	6	-	-	-	
<i>Pseudomonas stutzeri</i> 28a24	Glycine	-	-	-	-	-	4	6
	L-Alanine	-	-	-	-	-	2	
<i>Pseudomonas plecoglossicida</i>	L-Alanine	-	-	-	-	3	3	6
<i>Pseudomonas psychrotolerans</i>	L-Alanine	-	-	-	-	2	-	4
	L-Cysteine	-	-	-	-	2	-	
<i>Pseudomonas fluorescens</i> A506	L-Proline	-	-	4	-	-	-	4
<i>Pseudomonas lundensis</i>	L-Cysteine	-	-	-	-	-	4	4
<i>Pseudomonas azotoformans</i>	L-Proline	-	-	-	3	-	-	3
<i>Pseudomonas lurida</i>	L-Glutamine	-	-	-	1	-	-	2
	L-Proline	-	-	-	1	-	-	
<i>Pseudomonas mandelii</i>	L-Alanine	-	-	-	-	-	2	2
<i>Pseudomonas corrugata</i>	L-Proline	-	-	2	-	-	-	2
<i>Pseudomonas veronii</i>	L-Cysteine	-	-	-	-	-	2	2
<i>Pseudomonas sp.</i> TCU-HL1	L-Homoserine	-	-	-	-	-	1	1

Appendix K: List of organisms used as representative set in KAAS for assigning metabolic annotations to mango genes

KEGG Organism Code	Organism Name
cic	Citrus clementina (mandarin orange)
cit	Citrus sinensis (Valencia orange)
crb	Capsella rubella
dzi	Durio zibethinus (durian)
egr	Eucalyptus grandis (rose gum)
fve	Fragaria vesca (woodland strawberry)
gab	Gossypium arboreum
ghi	Gossypium hirsutum (upland cotton)
gmh	Glycine max (soybean)
gra	Gossypium raimondii
gsj	Glycine soja (wild soybean)
han	Helianthus annuus (common sunflower)
hbr	Hevea brasiliensis (rubber tree)
jre	Juglans regia (English walnut)
lja	Lotus japonicus
lsv	Lactuca sativa (garden lettuce)
mdm	Malus domestica (apple)
mtr	Medicago truncatula (barrel medic)
mus	Musa acuminata (wild Malaysian banana)
oeu	Olea europaea var. sylvestris (wild olive)
peu	Populus euphratica (Euphrates poplar)
pop	Populus trichocarpa (black cottonwood)
qsu	Quercus suber (cork oak)
rcn	Rosa chinensis (China rose)
sly	Solanum lycopersicum (tomato)
tcc	Theobroma cacao (cacao)
thj	Tarenaya hassleriana (spider flower)
vun	Vigna unguiculata (cowpea)
vvi	Vitis vinifera (wine grape)

Appendix L: List of enzymes predicted to be upregulated in ‘Sindhri’

KEGG Orthology Number	Gene	Enzyme	KEGG Pathway Map
K00021	HMGCR	hydroxymethylglutaryl-CoA reductase (NADPH) (EC 1.1.1.34)	Terpenoid backbone biosynthesis
K00059	fabG OAR1	3-oxoacyl-(acyl-carrier protein) reductase (EC 1.1.1.100)	Biotin metabolism, Prodigiosin biosynthesis Fatty acid biosynthesis
K00264	GLT1	glutamate synthase (NADH) (EC 1.4.1.14)	Alanine, aspartate and glutamate metabolism Nitrogen metabolism
K00278	nadB	L-aspartate oxidase (EC 1.4.3.16)	Alanine, aspartate and glutamate metabolism Nicotinate and nicotinamide metabolism
K00366	nirA	ferredoxin-nitrite reductase (EC 1.7.7.1)	Nitrogen metabolism
K00434	E1.11.1.11	L-ascorbate peroxidase (EC 1.11.1.11)	Ascorbate and aldarate metabolism Glutathione metabolism
K00454	LOX2S	lipoygenase (EC 1.13.11.12)	Linoleic acid metabolism alpha-Linolenic acid metabolism
K00512	CYP17A	steroid 17alpha-monooxygenase / 17alpha-hydroxyprogesterone deacetylase (EC 1.14.14.19 1.14.14.32)	Steroid hormone biosynthesis
K00640	cysE	serine O-acetyltransferase (EC 2.3.1.30)	Cysteine and methionine metabolism Sulfur metabolism
K00660	CHS	chalcone synthase (EC 2.3.1.74)	Flavonoid biosynthesis
K00737	MGAT3	beta-1,4-mannosyl-glycoprotein beta-1,4-N-acetylglucosaminyltransferase (EC 2.4.1.144)	N-Glycan biosynthesis
K00799	GST gst	glutathione S-transferase (EC 2.5.1.18)	Glutathione metabolism Metabolism of xenobiotics by cytochrome P450
K00826	E2.6.1.42 ilvE	branched-chain amino acid aminotransferase (EC 2.6.1.42)	Cysteine and methionine metabolism Glucosinolate biosynthesis Pantothenate and CoA biosynthesis Valine, leucine and isoleucine biosynthesis
K00873	PK pyk	pyruvate kinase (EC 2.7.1.40)	Glycolysis / Gluconeogenesis Pyruvate metabolism Purine metabolism
K00891	E2.7.1.71 aroK aroL	shikimate kinase (EC 2.7.1.71)	Phenylalanine, tyrosine and tryptophan biosynthesis
K00901	dgkA DGK	diacylglycerol kinase (ATP) (EC 2.7.1.107)	Glycerolipid metabolism, Glycerophospholipid metabolism
K00913	ITPK1	inositol-1,3,4-trisphosphate 5/6-kinase / inositol-tetrakisphosphate 1-kinase (EC 2.7.1.159 2.7.1.134)	Inositol phosphate metabolism

Continued from previous page

K00942	E2.7.4.8 gmk	guanylate kinase (EC 2.7.4.8)	Purine metabolism
K01052	LIPA	lysosomal acid lipase/cholesteryl ester hydrolase (EC 3.1.1.13)	Steroid biosynthesis
K01054	MGLL	acylglycerol lipase (EC 3.1.1.23)	Glycerolipid metabolism
K01082	cysQ MET22 BPNT1	3'(2'), 5'-bisphosphate nucleotidase (EC 3.1.3.7)	Sulfur metabolism
K01087	otsB	trehalose 6-phosphate phosphatase (EC 3.1.3.12)	Starch and sucrose metabolism
K01115	PLD1_2	phospholipase D1/2 (EC 3.1.4.4)	Ether lipid metabolism Glycerophospholipid metabolism
K01126	E3.1.4.46 glpQ ugpQ AMY	glycerophosphoryl diester phosphodiesterase (EC 3.1.4.46)	Glycerophospholipid metabolism
K01176	amyA malS	alpha-amylase (EC 3.2.1.1)	Starch and sucrose metabolism
K01177	E3.2.1.2	beta-amylase (EC 3.2.1.2)	Starch and sucrose metabolism
K01568	PDC pdc	pyruvate decarboxylase (EC 4.1.1.1)	Glycolysis / Gluconeogenesis
K01601	rbcL	ribulose-bisphosphate carboxylase large chain (EC 4.1.1.39)	Carbon fixation in photosynthetic organisms Glyoxylate and dicarboxylate metabolism
K01620	ltaE	threonine aldolase (EC 4.1.2.48)	Glycine, serine and threonine metabolism
K01772	hemH FECH	protoporphyrin/coproporphyrin ferrochelatase (EC 4.99.1.1)	Porphyrin and chlorophyll metabolism
K01807	rpiA	ribose 5-phosphate isomerase A (EC 5.3.1.6)	Carbon fixation in photosynthetic organisms Pentose phosphate pathway
K01897	ACSL fadD	long-chain acyl-CoA synthetase (EC 6.2.1.3)	Fatty acid biosynthesis and degradation
K01915	glnA GLUL	glutamine synthetase (EC 6.3.1.2)	Alanine, aspartate and glutamate metabolism Nitrogen metabolism Arginine biosynthesis Glyoxylate and dicarboxylate metabolism
K02108	ATPF0A atpB	F-type H ⁺ -transporting ATPase subunit a	Oxidative phosphorylation Photosynthesis
K02109	ATPF0B atpF	F-type H ⁺ -transporting ATPase subunit b	Oxidative phosphorylation Photosynthesis
K02126	ATPeF0A MTATP6 ATP6	F-type H ⁺ -transporting ATPase subunit a	Oxidative phosphorylation
K02132	ATPeF1A ATP5A1 ATP1	F-type H ⁺ -transporting ATPase subunit alpha	Oxidative phosphorylation
K02256	COX1	cytochrome c oxidase subunit 1 (EC 7.1.1.9)	Oxidative phosphorylation
K02262	COX3	cytochrome c oxidase subunit 3	Oxidative phosphorylation
K02291	crtB	15-cis-phytoene synthase (EC 2.5.1.32)	Carotenoid biosynthesis

Continued from previous page

K02575	NRT narK nrtP nasA	MFS transporter, NNP family, nitrate/nitrite transporter	Nitrogen metabolism
K02634	petA	apocytochrome f	Photosynthesis
K02635	petB	cytochrome b6	Photosynthesis
K02704	psbB	photosystem II CP47 chlorophyll apoprotein	Photosynthesis
K02707	psbE	photosystem II cytochrome b559 subunit alpha	Photosynthesis
K02709	psbH	photosystem II PsbH protein	Photosynthesis
K03517	nadA	quinolinate synthase (EC 2.5.1.72)	Nicotinate and nicotinamide metabolism
K03715	MGD	1,2-diacylglycerol 3-beta- galactosyltransferase (EC 2.4.1.46)	Glycerolipid metabolism
K03881	ND4	NADH-ubiquinone oxidoreductase chain 4 (EC 7.1.1.2)	Oxidative phosphorylation
K03884	ND6	NADH-ubiquinone oxidoreductase chain 6 (EC 7.1.1.2)	Oxidative phosphorylation
K03935	NDUFS2	NADH dehydrogenase (ubiquinone) Fe-S protein 2 (EC 7.1.1.2 1.6.99.3)	Oxidative phosphorylation
K04125	E1.14.11.1 3	gibberellin 2beta-dioxygenase (EC 1.14.11.13)	Diterpenoid biosynthesis
K05578	ndhG	NAD(P)H-quinone oxidoreductase subunit 6 (EC 7.1.1.2)	Oxidative phosphorylation
K05894	OPR	12-oxophytodienoic acid reductase (EC 1.3.1.42)	alpha-Linolenic acid metabolism
K05907	APR	adenylyl-sulfate reductase (glutathione) (EC 1.8.4.9)	Sulfur metabolism
K05929	E2.1.1.103 NMT	phosphoethanolamine N- methyltransferase (EC 2.1.1.103)	Glycerophospholipid metabolism
K07513	ACAA1	acetyl-CoA acyltransferase 1 (EC 2.3.1.16)	Biosynthesis of unsaturated fatty acids Fatty acid degradation Valine, leucine and isoleucine degradation alpha-Linolenic acid metabolism
K08081	TR1	tropinone reductase I (EC 1.1.1.206)	Tropane, piperidine and pyridine alkaloid biosynthesis
K08099	3.1.1.14	chlorophyllase (EC 3.1.1.14)	Porphyrin and chlorophyll metabolism
K08247	E2.1.1.12	methionine S-methyltransferase (EC 2.1.1.12)	Selenocompound metabolism
K09753	CCR	cinnamoyl-CoA reductase (EC 1.2.1.44)	Phenylpropanoid biosynthesis
K09840	NCED	9-cis-epoxycarotenoid dioxygenase (EC 1.13.11.51)	Carotenoid biosynthesis
K10534	NR	nitrate reductase (NAD(P)H) (EC 1.7.1.1 1.7.1.2 1.7.1.3)	Nitrogen metabolism
K12448	UXE uxe	UDP-arabinose 4-epimerase (EC 5.1.3.5)	Amino sugar and nucleotide sugar metabolism
K12450	RHM	UDP-glucose 4,6-dehydratase (EC 4.2.1.76)	Amino sugar and nucleotide sugar metabolism

Continued from previous page			
K12930	BZ1	anthocyanidin 3-O-glucosyltransferase (EC 2.4.1.115)	Anthocyanin biosynthesis
K13051	ASRGL1 iaaA	L-asparaginase / beta-aspartyl-peptidase (EC 3.5.1.1 3.4.19.5)	Alanine, aspartate and glutamate metabolism Cyanoamino acid metabolism
K13260	CYP81E1_7	isoflavone/4'-methoxyisoflavone 2'-hydroxylase (EC 1.14.14.90 1.14.14.89)	Isoflavonoid biosynthesis
K13379	RGP UTM	reversibly glycosylated polypeptide / UDP-arabinopyranose mutase (EC 2.4.1.- 5.4.99.30)	Amino sugar and nucleotide sugar metabolism
K13600	CAO	chlorophyllide a oxygenase (EC 1.14.13.122)	Porphyrin and chlorophyll metabolism
K13679	WAXY	granule-bound starch synthase (EC 2.4.1.242)	Starch and sucrose metabolism
K14156	CHK	choline/ethanolamine kinase (EC 2.7.1.32 2.7.1.82)	Glycerophospholipid metabolism
K14190	VTC2_5	GDP-L-galactose phosphorylase (EC 2.7.7.69)	Ascorbate and aldarate metabolism
K14595	AOG	abscisate beta-glucosyltransferase (EC 2.4.1.263)	Carotenoid biosynthesis
K15398	CYP86A4S	fatty acid omega-hydroxylase (EC 1.14.-.-)	Cutin, suberine and wax biosynthesis
K15718	LOX1_5	linoleate 9S-lipoxygenase (EC 1.13.11.58)	Linoleic acid metabolism
K15889	PCME	prenylcysteine alpha-carboxyl methyltransferase (EC 3.1.1.-)	Terpenoid backbone biosynthesis
K19562	BIO3-BIO1	bifunctional dethiobiotin synthetase / adenosylmethionine--8-amino-7-oxononanoate aminotransferase (EC 6.3.3.3 2.6.1.62)	Biotin metabolism
K20623	CYP92A6	typhasterol/6-deoxytyphasterol 2alpha-hydroxylase	Brassinosteroid biosynthesis
K22133	AAE3	oxalate---CoA ligase (EC 6.2.1.8)	Glyoxylate and dicarboxylate metabolism
K22911	TH2	thiamine phosphate phosphatase / amino-HMP aminohydrolase (EC 3.1.3.100 3.5.99.-)	Thiamine metabolism

Appendix M: List of enzymes predicted to be downregulated in ‘Sindhri’

KEGG Orthology Number	Gene(s)	Enzyme	Pathway
K00133	asd	aspartate-semialdehyde dehydrogenase (EC 1.2.1.11)	Cysteine and methionine metabolism Lysine biosynthesis Monobactam biosynthesis Glycine, serine and threonine metabolism
K00145	argC	N-acetyl-gamma-glutamyl-phosphate reductase (EC 1.2.1.38)	Arginine biosynthesis
K00166	BCKDHA bkdA1	2-oxoisovalerate dehydrogenase E1 component alpha subunit (EC 1.2.4.4)	Propanoate metabolism Valine, leucine and isoleucine degradation
K00222	TM7SF2 ERG24	Delta14-sterol reductase (EC 1.3.1.70)	Steroid biosynthesis
K00227	SC5DL ERG3	Delta7-sterol 5-desaturase (EC 1.14.19.20)	Steroid biosynthesis
K00326	E1.6.2.2	cytochrome-b5 reductase (EC 1.6.2.2)	Amino sugar and nucleotide sugar metabolism
K00411	UQCRFS1 RIP1 petA	ubiquinol-cytochrome c reductase iron-sulfur subunit (EC 7.1.1.8)	Oxidative phosphorylation
K00511	SQLE ERG1	squalene monooxygenase (EC 1.14.14.17)	Sesquiterpenoid and triterpenoid biosynthesis Steroid biosynthesis
K00547	mmuM BHMT2	homocysteine S-methyltransferase (EC 2.1.1.10)	Cysteine and methionine metabolism
K00736	MGAT2	alpha-1,6-mannosyl-glycoprotein beta-1,2-N-acetylglucosaminyltransferase (EC 2.4.1.143)	N-Glycan biosynthesis Various types of N-glycan biosynthesis
K00827	AGXT2	alanine-glyoxylate transaminase / (R)-3-amino-2-methylpropionate-pyruvate transaminase (EC 2.6.1.44 2.6.1.40)	Alanine, aspartate and glutamate metabolism Cysteine and methionine metabolism Glycine, serine and threonine metabolism Valine, leucine and isoleucine degradation
K00915	IPMK IPK2 E2.7.7.41	inositol-polyphosphate multikinase (EC 2.7.1.140 2.7.1.151)	Inositol phosphate metabolism
K00981	CDS1 CDS2 cdsA	phosphatidate cytidyltransferase (EC 2.7.7.41)	Glycerophospholipid metabolism
K01074	PPT	palmitoyl-protein thioesterase (EC 3.1.2.22)	Fatty acid elongation
K01210	E3.2.1.58	glucan 1,3-beta-glucosidase (EC 3.2.1.58)	Starch and sucrose metabolism
K01213	E3.2.1.67	galacturan 1,4-alpha-galacturonidase (EC 3.2.1.67)	Pentose and glucuronate interconversions
K01214	ISA treX	isoamylase (EC 3.2.1.68)	Starch and sucrose metabolism
K01512	acyP	acylphosphatase (EC 3.6.1.7)	Aminobenzoate degradation Pyruvate metabolism

Continued from previous page

K01641	E2.3.3.10	hydroxymethylglutaryl-CoA synthase (EC 2.3.3.10)	Butanoate metabolism Valine, leucine and isoleucine degradation Synthesis and degradation of ketone bodies Terpenoid backbone biosynthesis
K01714	dapA	4-hydroxy-tetrahydrodipicolinate synthase (EC 4.3.3.7)	Lysine biosynthesis Monobactam biosynthesis
K01739	metB	cystathionine gamma-synthase (EC 2.5.1.48)	Cysteine and methionine metabolism Sulfur metabolism Selenocompound metabolism
K01749	hemC HMBS	hydroxymethylbilane synthase (EC 2.5.1.61)	Porphyrin and chlorophyll metabolism
K01783	rpe RPE	ribulose-phosphate 3-epimerase (EC 5.1.3.1)	Carbon fixation in photosynthetic organisms Pentose and glucuronate interconversions Pentose phosphate pathway
K01805	xylA	xylose isomerase (EC 5.3.1.5)	Fructose and mannose metabolism Pentose and glucuronate interconversions
K01937	pyrG CTPS	CTP synthase (EC 6.3.4.2)	Pyrimidine metabolism
K01962	accA	acetyl-CoA carboxylase carboxyl transferase subunit alpha (EC 6.4.1.2 2.1.3.15)	Pyruvate metabolism Fatty acid biosynthesis Propanoate metabolism
K02155	ATPeV0C ATP6L	V-type H ⁺ -transporting ATPase 16kDa proteolipid subunit	Oxidative phosphorylation
K02699	psaL	photosystem I subunit XI	Photosynthesis
K03428	bchM chlM	magnesium-protoporphyrin O-methyltransferase (EC 2.1.1.11)	Porphyrin and chlorophyll metabolism
K03859	PIGC GPI2	phosphatidylinositol N-acetylglucosaminyltransferase subunit C	Glycosylphosphatidylinositol (GPI)-anchor biosynthesis
K04122	GA3 CYP701	ent-kaurene oxidase (EC 1.14.14.86)	Diterpenoid biosynthesis
K04718	SPHK	sphingosine kinase (EC 2.7.1.91)	Sphingolipid metabolism
K05283	PIGW	glucosaminylphosphatidylinositol acyltransferase (EC 2.3.-.-)	Glycosylphosphatidylinositol (GPI)-anchor biosynthesis
K05291	PIGS	GPI-anchor transamidase subunit S	Glycosylphosphatidylinositol (GPI)-anchor biosynthesis
K05349	bglX	beta-glucosidase (EC 3.2.1.21)	Cyanoamino acid metabolism Starch and sucrose metabolism Phenylpropanoid biosynthesis
K05350	bglB	beta-glucosidase (EC 3.2.1.21)	Cyanoamino acid metabolism Phenylpropanoid biosynthesis Starch and sucrose metabolism
K05857	PLCD	phosphatidylinositol phospholipase C, delta (EC 3.1.4.11)	Inositol phosphate metabolism
K08679	E5.1.3.6	UDP-glucuronate 4-epimerase (EC 5.1.3.6)	Amino sugar and nucleotide sugar metabolism

Continued from previous page

K08908	LHCA2	light-harvesting complex I chlorophyll a/b binding protein 2	Photosynthesis - antenna proteins
K09588	CYP90A1 CPD	cytochrome P450 family 90 subfamily A polypeptide 1 (EC 1.14.-.-)	Brassinosteroid biosynthesis
K09754	CYP98A C3'H	5-O-(4-coumaroyl)-D-quinic 3'-monooxygenase (EC 1.14.14.96)	Flavonoid biosynthesis Stilbenoid, diarylheptanoid and gingerol biosynthesis Phenylpropanoid biosynthesis
K09903	pyrH	uridylate kinase (EC 2.7.4.22)	Pyrimidine metabolism
K10251	HSD17B12 KAR IFA38	17beta-estradiol 17-dehydrogenase / very-long-chain 3-oxoacyl-CoA reductase (EC 1.1.1.62 1.1.1.330)	Biosynthesis of unsaturated fatty acids Steroid hormone biosynthesis Fatty acid elongation
K11420	EHMT	(histone H3)-lysine9 N-trimethyltransferase EHMT (EC 2.1.1.355)	Lysine degradation
K12345	SRD5A3	3-oxo-5-alpha-steroid 4-dehydrogenase 3 / polyprenol reductase (EC 1.3.1.22 1.3.1.94)	N-Glycan biosynthesis Steroid hormone biosynthesis
K12638	CYP90D1	3-epi-6-deoxocathasterone 23-monooxygenase (EC 1.14.14.147)	Brassinosteroid biosynthesis
K13034	ATCYSC1	L-3-cyanoalanine synthase/ cysteine synthase (EC 2.5.1.47 4.4.1.9)	Cyanoamino acid metabolism Sulfur metabolism Cysteine and methionine metabolism
K13082	DFR	bifunctional dihydroflavonol 4-reductase/flavanone 4-reductase (EC 1.1.1.219 1.1.1.234)	Flavonoid biosynthesis
K13427	NOA1	nitric-oxide synthase, plant (EC 1.14.13.39)	Arginine and proline metabolism Arginine biosynthesis
K13511	TAZ	monolysocardiolipin acyltransferase (EC 2.3.1.-)	Glycerophospholipid metabolism
K13832	aroDE DHQ-SDH	3-dehydroquinate dehydratase / shikimate dehydrogenase (EC 4.2.1.10 1.1.1.25)	Phenylalanine, tyrosine and tryptophan biosynthesis
K14157	AASS	alpha-aminoacidic semialdehyde synthase (EC 1.5.1.8 1.5.1.9)	Lysine degradation
K14423	SMO1	plant 4,4-dimethylsterol C-4alpha-methyl-monooxygenase (EC 1.14.18.10)	Steroid biosynthesis
K15227	TYRAAT	arogenate dehydrogenase (NADP+), plant (EC 1.3.1.78)	Phenylalanine, tyrosine and tryptophan biosynthesis
K15404	K15404 CER1	aldehyde decarbonylase (EC 4.1.99.5)	Cutin, suberine and wax biosynthesis
K16903	TAA1	L-tryptophan---pyruvate aminotransferase (EC 2.6.1.99)	Tryptophan metabolism
K18826	CAMKMT	calmodulin-lysine N-methyltransferase (EC 2.1.1.60)	Lysine degradation
K19355	MAN	mannan endo-1,4-beta-mannosidase (EC 3.2.1.78)	Fructose and mannose metabolism
K19891	GN1_2_3	glucan endo-1,3-beta-glucosidase 1/2/3 (EC 3.2.1.39)	Starch and sucrose metabolism
K19893	GN5_6	glucan endo-1,3-beta-glucosidase 5/6 (EC 3.2.1.39)	Starch and sucrose metabolism

Continued from previous page			
K20896	TENA_E	formylaminopyrimidine deformylase / aminopyrimidine aminohydrolase (EC 3.5.1.- 3.5.99.-)	Thiamine metabolism
K22845	PGT1	phlorizin synthase (EC 2.4.1.357)	Flavonoid biosynthesis
K22849	DGAT3	diacylglycerol O-acyltransferase 3, plant (EC 2.3.1.20)	Glycerolipid metabolism