Thesis for the degree of doctor of philosophy

# Advancing Metabolic Engineering through Combination of Systems Biology and Adaptive Evolution

KUK-KI HONG


Department of Chemical and Biological Engineering,

Chalmers University of Technology

Göteborg, Sweden 2012

Advancing Metabolic Engineering through Combination of Systems Biology and Adaptive Evolution

KUK-KI HONG



ISBN: 978-91-7385-735-2

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 3416
ISSN: 0346-718X

PhD Thesis
Systems and Synthetic Biology
Department of Chemical and Biological Engineering
Chalmers University of Technology
SE-412 96 Göteborg
Sweden
Telephone +46 (0) 31-772 1000

**Thesis Supervisor: Prof. Jens B Nielsen**

Primary Funding Source:
Bioscience Research Center, CJ CheilJedang Corp.
92-1 Gayang-dong, Gangseo-gu, Seoul
Republic of Korea

Cover illustration: evolutionary strategies of yeast for improving galactose utilization; for more details, refer to Fig. 3-13.

Printed by Chalmers Reproservice
Göteborg, Sweden 2012

To my family

My wife, Min-Jin

My angels, Jin-Seo and Jin-Ha

A thunderstorm can be viewed as a consequence of Zeus' anger or of a difference of potential between the clouds and the earth. A disease can be seen as the result of a spell cast on the patient or of an infection by a virus. In all cases, however, one watches the visible effect of some hidden cause related to the whole set of invisible forces that are supposed to run the world.

- François Jacob (Evolution and Tinkering, Science, 1977)

# PREFACE

This dissertation is submitted for the partial fulfillment of the degree of doctor of philosophy at the department of chemical and biological engineering, Chalmers University of Technology, Sweden. The doctoral research is on the application of systems biology for the characterization of adaptively evolved mutants. The process of systems biology approach enhances the understanding of evolutionary strategies that may contribute to advance metabolic engineering. This advance is likely useful to improve biological engineering, which provides one of the possible solutions to substitute petroleum based chemical production. This research was funded by the doctoral fellowship program of CJ CheilJedang (Korea), the Chalmers Foundation, the Knut and Alice Wallenberg Foundation, the European Union funded projects UNICELLSYS (Contract 201142), SYSINBIO (Contract 212766), European Research Council Grant 247013 and the Novo Nordisk Foundation.

Kuk-Ki Hong

August 2012

Advancing Metabolic Engineering through Combination of Systems Biology and Adaptive Evolution

KUK-KI HONG
Systems and Synthetic Biology Group
Department of Chemical and Biological Engineering, Chalmers University of Technology

# ABSTRACT

Understanding evolutionary strategies of microorganisms may provide opportunities for advanced strain development with the aim to produce valuable bio-products from renewable biomass resources. Through evolutionary processes, microorganisms can attain new traits associated with genetic changes that may be useful for the construction of improved strains. Therefore, the characterization of evolutionary strategies may result in identification of the molecular and genetic changes underlying newly obtained traits, and can hereby become an essential step in strain development. However, so far the depth of analysis has limited the range of comprehension. This thesis applied genome-wide analyses such as transcriptome, metabolome and whole-genome sequencing to investigate the evolutionary strategies of the yeast *Saccharomyces cerevisiae*. Three evolved mutants were independently generated by adaptive evolution on galactose minimal media to obtain the trait of improved galactose utilization by yeast. Those strains expressed higher galactose utilization rates than a reference strain in terms of both maximum specific growth rate and specific galactose uptake rate. Application of the genome-scale comparative analyses employing engineered strains as controls elucidated unique changes obtained by adaptive evolution. Molecular bases referred from the changes of transcriptome and metabolome were located around galactose metabolism, while genetic bases from whole-genome sequencing showed no mutations in those changes. Common mutations among the evolved mutants were identified in the Ras/PKA signaling pathway. Those mutations were placed on the reference strain background and their effects were evaluated by comparison with the evolved mutants. One of the site-directed mutants showed even higher specific galactose uptake rate than the evolved mutants, and just few number of genetic and molecular changes were enough to recover complete the adaptive phenotype. These results indicate that identification of key mutations provide new strategies for further metabolic engineering of strains. In addition, the pleiotropy of obtained phenotype that is improved galactose availability was tested. When the galactose-evolved mutants were cultured on glucose that is the most favorite carbon source of yeast, those mutants showed reduction of glucose utilization. Genome-wide analyses and site-directed mutagenesis were applied again to understand underlying molecular and genetic bases of this trade-off in carbon utilization. The results indicated that loosening of tight glucose regulation was likely the reason of increased galactose availability. The implications of evolutionary strategies and the impact of genome-scale analyses on characterization of evolved mutants are discussed.

Key words: metabolic engineering, evolutionary engineering, systems biology, galactose utilization, Ras/PKA signaling pathway, pleiotropy of evolutionary strategies

# LIST OF PUBLICATIONS

This thesis is based on the following publications, referred to as Paper I to IV in the text:

**I.** **Unravelling evolutionary strategies of yeast for improving galactose utilization through integrated systems level analysis**
Kuk-Ki Hong, Wanwipa Vongsangnak, Goutham N. Vemuri and Jens Nielsen
Proc. Natl. Acad. Sci. USA. 2011 Jul 19; 108(29):12179–84.

**II.** **Recovery of phenotypes obtained by adaptive evolution through inverse metabolic engineering**
Kuk-Ki Hong and Jens Nielsen
Accepted in Appl. Environ. Microbiol. 2012

**III.** **Adaptively evolved yeast mutants on galactose show trade-offs in carbon utilization on glucose**
Kuk-Ki Hong and Jens Nielsen
Submitted for publication

**IV.** **Metabolic engineering of *Saccharomyces cerevisiae*: a key cell factory platform for future biorefineries (Review)**
Kuk-Ki Hong and Jens Nielsen
Cell Mol Life Sci. 2012 Aug; 69(16):2671-90. Epub 2012 Mar 3.

Additional publications during doctoral research not included in this thesis

**V.** **Dynamic (13) C-labeling experiments prove important differences in protein turnover rate between two *Saccharomyces cerevisiae* strains.**
Kuk-Ki Hong, Jin Hou, Saeed Shoaie, Jens Nielsen and Sergio Bordel
FEMS Yeast Res. 2012 Jun 20. doi: 10.1111/j.1567-1364.2012.00823

**VI.** **Quantitative analysis of glycerol accumulation under hyper-osmotic stress and its various links to glycolysis**
Elzbieta Petelenz-Kurdziel, Clemens Kuehn, Bodil Nordlander, Dagmara Klein, Kuk-Ki Hong, Therese Jacobson, Peter Dahl, Joerg Schaber, Jens Nielsen, Stefan Hohmann, Edda Klipp.
Submitted for publication

# CONSTRIBUTION SUMMARY

A summary of contribution of Kuk-Ki Hong to each of the publications

I. Designed research; performed research; analyzed data; wrote the paper.

II. Designed research; performed research; analyzed data; wrote the paper.

III. Designed research; performed research; analyzed data; wrote the paper.

IV. Designed review; analyzed data; wrote the paper.

V. Designed research; performed research; analyzed data; wrote the paper.

VI. Performed research; analyzed data.

# TABLE OF CONTENTS

# LIST OF FIGURES AND TABLES

# ABBREVIATIONS AND SYMBOLS

βGal: β-D-galactose
αGal: α-D-galactose
αGal-1P (Gal 1P, Galactose 1P, Galacose-1-P): α-D-galactose-1-phosphate
αGlu-1P (Glu 1P, Glucose 1P, Glucose-1-P): α-D-glucose-1-phosphate
αGlu-6P (Glu 6P, Glucose 6P, Glucose-6-P): α-D-glucose-6-phosphate
UDP-Glu (UDP-Glucose): Uridine diphosphate glucose
UDP-Gal (UDP-Galactose): Uridine diphosphate galactose
gal: galactose
glu: glucose
Trehalose 6P: Trehalose-6-phosphate
REF: Reference strain
KEGG: Kyoto Encyclopedia of Genes and Genomes
GO term: Gene Ontology term
PCA: Principal component analysis
PC: Principal component
SNPs: Single-nucleotide polymorphism
gTME: Global transcription machinery engineering
MAGE: Multiplex automated genome engineering
TRMR: Trackable multiplex recombineering

*Yeast nomenclature*

Gene name consists of three letters and up to three numbers, ex. *GAL4, MIG1, ura3*

Wild-type gene name is written with capital letters in italic, ex. *PGM2, RAS2, UGP1*

Recessive mutant gene name is written with small letters in italics, ex. *mig1, gal80, gal6*

Mutant alleles are named with a dash and a number, ex. *ura3-52, cdc28-2*

Deleted gene with the genetic marker is used for deletion, ex. *tps1Δ::HIS3*

The gene product, a protein, is written with a capital letter at the first letter and not in italics; often a ”p” is added at the end, ex. Pgm2p, Ugp1p

Some genes, which are only found by systematic sequencing and their functions are not determined, get a landmark name, ex. YNL200C, YHL042W, YLR278C
(Y, yeast; the second letter, the chromosome (D=IV, M=XIII....); L or R, left or right chromosome arm; the three-digit number; the ORF counted from the centromere; C or W, Crick or Watson, i.e. direction of the ORF)

Exceptional case, ex. *HO, MATa, MATα*

Amino acid sequence change is described by gene name and changed amino acid with its position, ex. $RAS2^{Lys\ 77}$, $ERG5^{\ Pro\ 370}$

# 1. INTRODUCTION

## 1.1. Yeast *Saccharomyces cerevisiae* for future biorefineries

Even before recognizing the presence of microorganisms, mankind has used microbial fermentation to produce beverages and foods. Since 1920 industrial microbial fermentation has been used to manufacture organic acids, amino acids and vitamins (Kinoshita, *et al.*, 1957, Nakayama, *et al.*, 1961, Demain, 2000). The advent of genetic engineering in the 1970s led to the use of microbial fermentation for the production of pharmaceutical proteins such as human insulin and human growth hormone (Goeddel, *et al.*, 1979, Johnson, 1983). Currently, the world is confronting serious challenges such as climate changes due to greenhouse gas emission and the depletion of petroleum oil causing limitation of energy and chemical resources. Microbial fermentation is considered as one of the possible solutions to these grand challenges, because it uses renewable biomass that can also absorb carbon dioxide during growth, and produce fuels and chemicals in eco-friendly processes (Lipinsky, 1981, Werpy & Petersen, 2004, Vennestrom, *et al.*, 2011). There are already several successful industrial trials to produce chemicals from biomass by microbial fermentations (Table 1-1).

**Table 1-1.** Chemicals are recently produced from biomass, including major players and host strains

| Chemicals | Products/Uses | Major players | Host strains |
|---|---|---|---|
| succinic acid | plastics, chemical intermediates, solvents, polyurethanes, plasticizers | BASF/Purac(CSM) | *Basfia succiniciproducens (from Bovin rumen, Gram-negative)* |
| 3-hydroxypropionic acid | acrylic acid: plastics, fiber, coatings, paints, super-absorbent diapers | Novozymes/Cargill | *Escherichia coli Saccharomyces cerevisiae* |
| isoprene | synthetic rubber | Genencor(Danisco) /Goodyear | *Bacillus subtilis, Escherichia coli, Pantoea citrea, Trichoderma reesei, Yarrowia lipolytica* |
| lactic acid | plastics, synthetic fibers | Cargill | *Kluyveromyces marxianus* |

| | | | |
|---|---|---|---|
| lactic acid | plastics, synthetic fibers | Purac(CSM)/Arkema | thermophilic *Bacillus,* thermophilic *Geobacillus* |
| 1,3-propanediol | engine coolant, cosmetics, surfactants, emulsifiers, preservatives, polymers | Dupont/Tate&Lyle | *Escherichia coli* |
| propylene | Thermoplastic | Braskem/Novozymes | *Propionibacterium acidipropionici* |

Strain development is a pre-requisite to materialize bio-based chemical production, as it is directly related to not only improving yield, titer, and productivity of products, but also utilizing cheap raw materials efficiently (Tyo, *et al.*, 2007, Patnaik, 2008, Elkins, *et al.*, 2010). The yeast *Saccharomyces cerevisiae* has been used for the production of a wide range of industrial products due to its tolerance to industrial conditions and the vast amount of knowledge about its physiology, biochemistry, genetics, and long history of fermentation (Pronk, 2002, van Maris, *et al.*, 2006, Nevoigt, 2008, Nielsen & Jewett, 2008, Krivoruchko, *et al.*, 2011). Thus, its products range and available current technologies are quite broad (Table 1-2).

**Table 1-2.** Examples of products and strains of *S.cerevisiae*

(More detailed explanation is in **Paper IV**.)

| Categories | Products | Strains | References |
|---|---|---|---|
| Biofuels | Ethanol | *CEN.PK102-3A (MATα ura3 leu2)* | (Guadalupe Medina*, et al.*, 2010) |
| | Biobutanol | *CEN.PK 2-1C (MATα leu2-3, 112 his3-Δ1 ura3-52 trp1-289 MAL2-8(Con) MAL3 SUC3)* | (Chen*, et al.*, 2011) |
| | Biodiesels | *YPH499 (MATα ura3-52 lys2-801_amber ade2-101_ochre trp1-D63 his3-D200 leu2-D1)* | (Yu*, et al.*, 2012) |
| | Bisabolene | *BY4742 (MATα his3D1 leu2D0 lys2D0 ura3D0)* | (Peralta-Yahya*, et al.*, 2011) |
| Bulk chemicals | 1,2-propanediol | *NOY386αA (MATα ura3-52 lys2-801 trp1-Δ63 his3-Δ200 leu2-Δ1), BWG1-7a (MATa ade1-100 his4-519 leu2-3,112 ura3-52 GAL$^{+}$)* | (Lee & Dasilva, 2006) |
| | L-Lactic acid | *CEN. PK2-1C (MATa ura3-52 trp1-289 leu2-3,112 his3Ä1 MAL2-8C SUC2)* | (Zhao*, et al.*, 2011) |
| | Polyhydroxy-alkanoates | *BY4743 (MATa/α his3Δ1/his3Δ1 leu2Δ0/leu2Δ0 ura3Δ0/ura3Δ0)* | (Zhang*, et al.*, 2006) |

| | | | |
|---|---|---|---|
| | Pyruvic acid | *CEN.PK113-7D (MATa MAL2-8C, SUC2)* | (van Maris*, et al.*, 2004) |
| | Succinic acid | *AH22ura3 (MATa ura3Δ leu2-3 leu2-112 his4-519 can1)* | (Raab*, et al.*, 2010) |
| Fine chemicals | β-amyrin | *CEN.PK113-7D (MATa MAL2-8C SUC2)* | (Madsen*, et al.*, 2011) |
| | β-carotene | *CEN.PK113-7D (MATa MAL2-8C SUC2)* | (Verwaal*, et al.*, 2007) |
| | Amorpha-4, 11-diene | *CEN.PK2-1C (MATa ura3-52 trp1-289 leu2-3,112 his3Ä1 MAL2-8C SUC2), CEN.PK2-1D (MATα ura3-52 trp1-289 leu2-3,112 his3Ä1 MAL2-8C SUC2)* | (Westfall*, et al.*, 2012) |
| | Cinnamoyl anthranilates | *BY4742 (MATα his3D1 leu2D0 lys2D0 ura3D0)* | (Eudes*, et al.*, 2011) |
| | Cubebol | *CEN.PK113-5D (MATa MAL2-8c SUC2 ura3-52 )* | (Asadollahi*, et al.*, 2010) |
| | Eicosapentaenoic acid (EPA) | *CEN.PK113-5D (MATa MAL2-8c SUC2 ura3-52 )* | (Tavares*, et al.*, 2011) |
| | Linalool | *BQS252 (MATa ura3-52 (derivative of FY1679))* | (Rico*, et al.*, 2010) |
| | Methylmalonyl-coenzyme A | *InvSC1 (MATa, his3delta1, leu2, trp1-289, ura3-52 (Invitrogen, Carlsbad, CA, USA)) BJ5464 (MATα, ura3-52, trp1, leu2-delta1, his3-delta200, pep4::HIS3, prb1-delta1.6R, can1, GAL).* | (Mutka*, et al.*, 2006) |
| | Patchoulol | *CEN.PK113-13D and CEN.PK113-5D* | (Albertsen*, et al.*, 2011) |
| | Resveratrol | *FY23 (MATa ura3-52 trplA63 leu2A1)* | (Becker*, et al.*, 2003) |
| | Vanillin | *X2180-1A ( MATa his3D1 leu2D0 met15D0 ura3D0 adh6::LEU2 bgl1::KanMX4 PTPI1::3DSD [AurC]::HsOMT [NatMX]::ACAR [HphMX])* | (Brochado*, et al.*, 2010) |
| | Se-methylseleno-cysteine | *CEN.PK113-7D (MATa MAL2-8C SUC2)* | (Mapelli*, et al.*, 2011) |
| | Non-ribosomal peptides | *CEN.PK113-11C (MAT a MAL2-8c SUC2 ura3-52 his3-D1)* | (Siewers*, et al.*, 2010) |
| Protein drugs | Insulin-like growth factor 1 (fhlGF-1) | *GcP3 (MAT a pep4-3 prb1-1122 ura3-52 leu2 gal2 cir°)* | (Vai*, et al.*, 2000) |
| | Glucagon | *SY107 (MATα YPS1 Δtpi::LEU2 pep4-3 leu2 Δura3 cir$^{+}$)* | (Egel-Mitani*, et al.*, 2000) |

| | | |
|---|---|---|
| single-chain antibodies (scFv) | *BJ5464 (a ura3-52 trp1 leu2D1 his3D200 pep40HIS3 prb1D1.6R can1 GAL)* | (Hackel*, et al.*, 2006) |
| Hepatitis surface antigen (HBsAg) | *INVSc1 (MATa his3D1 leu2 trp1-289 ura3-52)* | (Vellanki*, et al.*, 2007) |
| Parvovirus B19 VP2 | *HT393 (MATa leu2-3 leu2-112 ura3Δ5 prb1-1 prc1-1 pra1-1 pre1-1)* | (Lowin*, et al.*, 2005) |
| Epidermal Growth factor (EGF) | *W303-1A (MATa leu2-3,112 his3-11,15 ade2-1 ura3-1 trp1-1 can1-100), W303-1B (MATα leu2-3,112 his3-11,15 ade2-1 ura3-1 trp1-1 can1-100)* | (Chigira*, et al.*, 2008) |
| Immunoglobulin G | *BJ5464a (MATα ura3-52 leu2~1 his3~200 pep4::HIS3 prb1~1.6Rcan1 GAL)* | (Rakestraw*, et al.*, 2009) |
| Hepatitis B virus surface antigen (HBsAg) | *S.cerevisiae 2805 (MATα pep4::HIS3 prb-Δ1.6 his3 ura3-52 gal2 can1)* | (Kim*, et al.*, 2009) |
| L1 protein of human papillomavirus (HPV) type16 | *S.cerevisiae 2805 (MATα pep4::HIS3 prb-Δ1.6 his3 ura3-52 gal2 can1)* | (Kim*, et al.*, 2010) |

There is also extensive research on extending substrate range of this yeast. Resources for traditional fermentations have been derived from food crops like corn, wheat and sugar cane, but to replace the large amounts of fuels and chemicals currently derived from mineral oil, the use of abundant and renewable non-food resources such as switchgrass, corn-cob, bagasse, cheese whey and algae is necessary. These biomass resources are composed of diverse kinds of carbon structure: polymers (cellulose, starch, xylan), dimers (cellobiose, melibiose, lactose) and monomers (glucose, fructose, galactose, arabinose, xylose). Except the hexoses (glucose, fructose, galactose) and a few dimers (sucrose, maltose), most of these carbon compounds are not endogenously metabolized by *S. cerevisiae*. Even among the hexoses there are broad differences in uptake rate, for example the uptake rate of galactose is much lower than for the other hexoses. Therefore, the extension of substrate range of *S. cerevisiae* provides an excellent opportunity to enhance its suitability for biofuels and biochemicals production (van Maris, *et al.*, 2006, Hahn-Hagerdal, *et al.*, 2007, Nevoigt, 2008, Marie, *et al.*, 2009) (Fig. 1-1).

(For details, refer to **Paper IV**.).

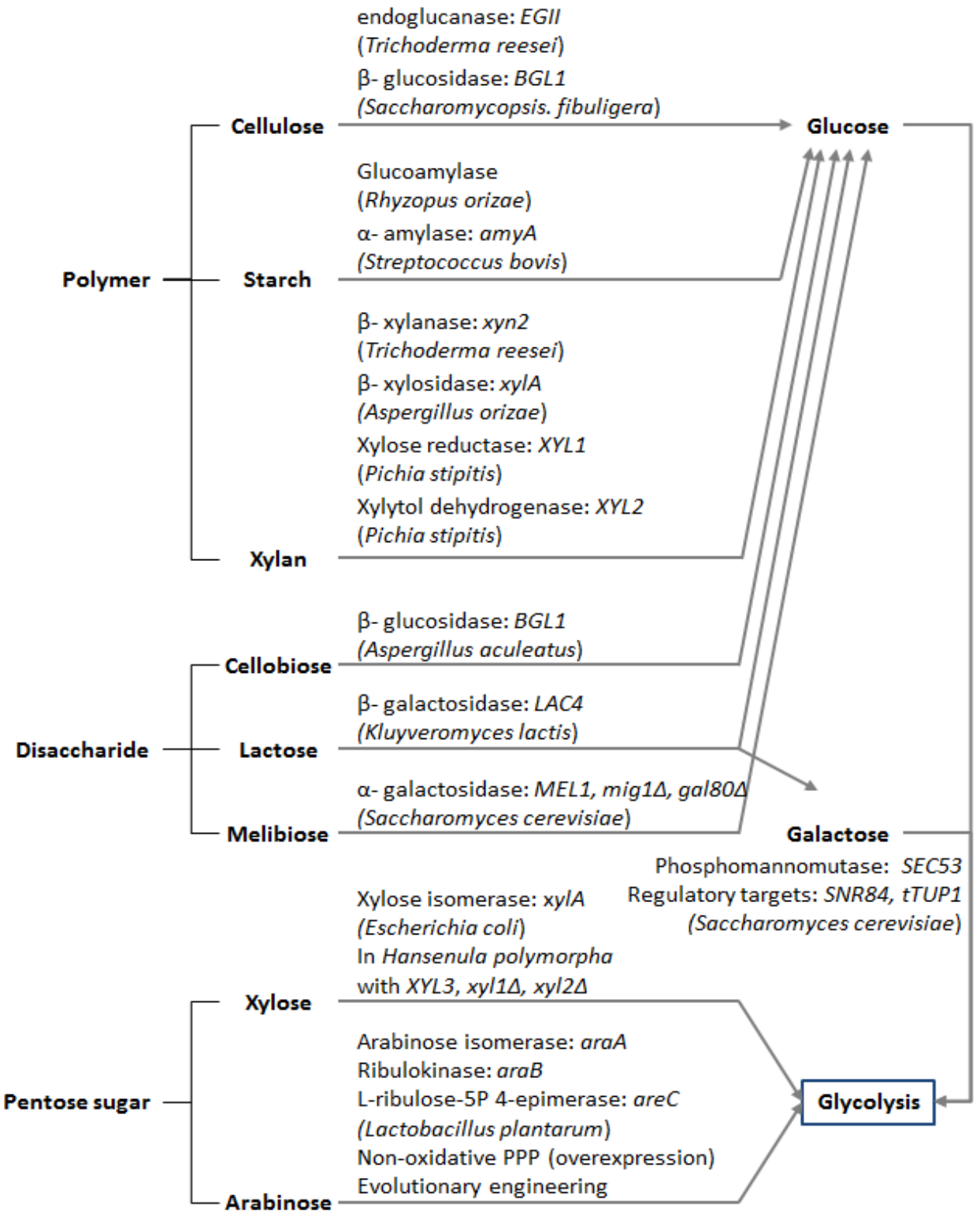


**Fig. 1-1.** Overview of relevant carbon sources for yeast fermentation. Heterologous enzymes that are currently introduced are summarized for non-utilizable carbon sources (polymer, disaccharide and pentose sugar) and non-preferred one (galactose) in *S. cerevisiae.*

## 1.2. Evolutionary approaches in strain development

### 1.2.1. Evolutionary engineering

Evolutionary engineering has been traditionally employed for strain development in industry, since it can generate specific traits relatively quickly at some level, even though governing biological principle may not be evident (Sauer, 2001, Zhang, *et al.*, 2002). The term of evolutionary engineering is composed of evolution and engineering. The evolution is a strategy of life to adapt to changed environments by natural selection. It is operated through iterative process of creating variation in population and selecting proper individuals, consequently a specific trait in the population is enriched. This nature's algorithms can be engineered to make biotechnological relevant traits by adjusting the rate of variant generation or defining new selection pressures. Therefore, evolutionary engineering is the application of suitable mutagenesis and artificially designed selection procedures based on evolutionary mechanisms for strain development. Success of evolutionary engineering is, hence, dependent on the ability to design mutagenesis and selection conditions (Sauer, 2001, Sonderegger & Sauer, 2003). On the one hand random mutagenesis by treatment of mutagens for adjusting mutation rate can be used; it generates a broader distribution of mutations in the genome, whereas it makes it more difficult to identify beneficial mutations (Sauer, 2001, Ikeda, *et al.*, 2006). On the other hand, the methods that can generate traceable mutations in specific regions have been developed, which is called genome engineering such as gTME, MAGE and TRMR (Santos & Stephanopoulos, 2008, Boyle & Gill, 2012). Another important consideration is to understand the underlying evolution mechanisms. Basically, evolutionary engineering relies on evolutionary mechanisms such as natural selection or natural preservation (Darwin regretted using selection more frequently than preservation). Since natural preservation is the fundamental evolutionary mechanism, this concept is used in the design stage of evolutionary engineering. It is also important to be aware of other relevant evolutionary mechanisms for strain development such as clonal interference, trade-offs in traits, negative epistasis (Elena & Lenski, 2003).

### 1.2.2. Inverse metabolic engineering

Developed strains based on evolutionary engineering can be directly used for industrial application. In most cases, only a few specific traits from the mutant strains are needed; however

industrial strains retain the combination of several non-necessary traits simultaneously (Ohnishi, *et al.*, 2002, Ikeda, *et al.*, 2006, Ikeda, *et al.*, 2009, Warner, *et al.*, 2009). Therefore, additionally the concept of inverse metabolic engineering has been used to make evolutionary engineering more useful. Inverse metabolic engineering starts from the identification of genetic basis of obtained phenotypes, and completed by transfer of that specific genotype(s) to an industrial strain (Bailey, *et al.*, 2002, Ikeda, *et al.*, 2006). The important part in this engineering is the identification of the genetic basis; not only in order to enable the transfer of genetic changes related to the gained trait, but also the specific trait(s) may not easily be reached to its optimum stage because of evolutionary constraints such as negative epistasis, clonal interference; therefore, additional engineering based on the identified genetic changes is sometimes required (Warner, *et al.*, 2009). For these reasons, the identification of the genetic bases of selected traits is crucial. Recently, analytical capabilities that can scan molecular or genetic alteration at genome scale have been developed such as omics tools and next generation whole genome sequencing. The integration of data generated from those tools is expected to facilitate identification of molecular and genetic changes in a more comprehensive fashion (Bro & Nielsen, 2004, Heinemann & Sauer, 2010, Oud, *et al.*, 2012).

### 1.2.3. Adaptive evolution

Adaptive evolution is often confused with several similar terms such as adaptive laboratory evolution, experimental evolution, and even evolutionary engineering (Sauer, 2001, Elena & Lenski, 2003, Conrad, *et al.*, 2011, Portnoy, *et al.*, 2011, Dettman, *et al.*, 2012). Adaptive evolution has been used to explain adaptation process of life in biology. When this process can be imitated in a laboratory to understand evolution mechanisms or applied to strain development, derivative words have been generated. Therefore, adaptive evolution includes both natural processes in basic science and a tool in biotechnology. It generates mutations spontaneously based on the cell's endogenous system, and finds a phenotype that have improved fitness to a given environment than an ancestor strain, simply by continuous exposure of a population to the given environment over a period of time.

## 1.3. Characterization of evolved mutants by genome-scale analysis

Evolutionary engineering has made commercially successful stories in strain development, while the identification of molecular or genetic basis that is involved in phenotypic changes has remained an enduring challenge (Ohnishi, et al., 2002,Warner, et al., 2009). There have been several efforts to find genetic changes. If biological information about an obtained phenotype is present at the pathway level or its regulation, one can check molecular changes in that specific pathway. For example, the strain producing high concentration of lysine was characterized based on analysis of specific amino acid production pathways (Ikeda, et al., 2006). Key mutations that were likely related to release of allosteric regulation were detected, and partial contribution of the mutations on the overall phenotype was confirmed. Technological advance has led to accumulation of huge amount of knowledge about the biological reactions and regulations, facilitating better predictability of relative molecular changes. However, since the changes are happened at the whole genome level, and the complexity of biological reactions and regulations are still beyond full comprehension, advanced analytical tools that can scan overall molecular changes in a system level of a cell are required. During the last decade, omics techniques have been developed for genome-wide analysis (Bro & Nielsen, 2004, Herrgard, et al., 2008, Petranovic & Vemuri, 2009, Snyder & Gallagher, 2009), and omics approaches established a new field in life science, so called Systems Biology that aims to understand a cell in an holistic view by using high-throughput omics data and mathematical models. Systems Biology has been implemented by quantifying each level of molecules through whole-genome sequencing, transcriptome, proteome, metabolome. Their usefulness and limitation especially for the characterization of evolved stains has been recently reviewed (Oud, et al., 2012).

### 1.3.1. Transcriptome analysis for the characterization of evolved strains

Transcriptome analysis has been routinely used in the last decade because of standardization of techniques and data with the support of bioinformatics and models (Bro, et al., 2005, Patil & Nielsen, 2005, Bengtsson, et al., 2008, Reimand, et al., 2011). Not only technical maturation, but also it has the best coverage among other omics tools (Herrgard, et al., 2008, Reimand, et al., 2011). The effect of environmental or genetic perturbation can be checked easily by counting the number of significantly changed genes. Identified differentially expressed genes between evolved mutants and a reference strain are routinely analyzed to find altered pathways, metabolisms, and

regulation circuits based on several gene enrichment methods. Therefore, the transcriptome analysis can enumerate all possible transcriptional changes that are related to obtained phenotypes. Although there are several restrictions such as mixing of transcriptional changes between cause and consequence or the desired phenotype related and the experimental condition related and so on, transcriptome analysis is essentially useful as a first scan of molecular changes in evolved mutants. Additionally, comparison of multi strains or combination with other omics data has identified key molecular changes in different mutants (Ideker, et al., 2001, Bro, et al., 2005, Bengtsson, et al., 2008, Vijayendran, et al., 2008, Hazelwood, et al., 2009).

### 1.3.2. Metabolome analysis for the characterization of evolved strains

Metabolites play important roles as intermediates of biochemical reactions, which means their concentration is a key factor for controlling the reaction rate and they further are involved in regulation of the metabolic network through allosteric regulation. Thus, the level of metabolites represents integrative information of the cellular function; they can give critical clues to define the phenotype in evolved mutants (Zaldivar, et al., 2002, Kummel, et al., 2010). However, since metabolites have very diverse molecular kinds, it is almost impossible to analyze and quantify all metabolites in a cell simultaneously unlike the transcriptome. Practically targeted metabolome that analyze and quantify selected metabolites therefore has been more frequently used than metabolite profiling that tries to increase the number of covering metabolites. Targeted metabolomics can get clues from transcriptome data in selected metabolites of interest; and these metabolites data can be used to provide additional proof about the link between a desired phenotype and molecular changes.

### 1.3.3. Whole-genome sequencing for the characterization of evolved strains

A genetic change is the first and direct origin of a phenotypic change. Other molecular alterations are reflections of the genetic change. Therefore identification of driving genetic changes is crucial in inverse metabolic engineering. The importance of the identification of genetic changes was mentioned by Bailey et al. in 1996, The power of the technology for deciphering the genetic basis for a given phenotype is a critical determinant of the feasibility of inverse metabolic engineering (Bailey, et al., 2002). At that time the main limitation was in the technical part, since whole genome sequencing was time consuming and had a high cost.

However, with next-generation sequencing there has been a revolution in genome sequencing technologies and this has reduced the costs many folds (Herring, et al., 2006, Mardis, 2008, Shendure & Ji, 2008, Le Crom, et al., 2009, MacLean, et al., 2009, Metzker, 2010, Oud, et al., 2012). These techniques show the possibility of substantial reduction of the time and cost of genome sequencing such that it can be used for routine application similar to transcriptome analysis. As an example, the results of sequencing three yeast evolved mutants that were used in this thesis are explained in Table 3 and 4.

**Table 1-3.** The price of the next-generation sequencing (Illumina/Solexa, at 13th January 2010)

| **Description** | **Quntity** | **Unit Price** (€) | **Total Price** (€) |
|---|---|---|---|
| Sample preparation for Genome Analysis, Genomic Shotgun | 3 (strains) | 544 | 1632 |
| Sample preparation with bar-coded adapters | 3 | 34 | 102 |
| Sequencing on the Genome Analyzer GAIIx, 1 paired-ends channel 2x38 bp | 1 | 3,808 | 3,808 |
| Additional bar-coded sample in the same channel, paired-ends | 2 | 136 | 272 |
| Bioinformatics analyses | 0 | 340 | 0 |
| | | **Total** | 5,814 |

**Table 1-4.** The performance of overall genome sequencing results

| **Sequencing Parameters** | **Mutant A** | **Mutant B** | **Mutant C** |
|---|---|---|---|
| No. of Reads | 5,605,504 | 18,203,846 | 5,239,106 |
| Total Bases (bp) * | 213,009,152 | 691,746,148 | 199,086,028 |
| Coverage Fold | 17 | 55 | 16 |
| Undetermined Base | 158,723 | 86,791 | 171,362 |
| Genome percent reference coverage (%)† | 98.7 | 99.3 | 98.6 |
| No. of supercontigs | 17 | 17 | 17 |
| Chromosomes | 16 | 16 | 16 |
| Mitochondria | 1 | 1 | 1 |

* 38 bases per sequence read for 2 cycles

† Based on genome consensus sequence length of CEN.PK113-7D of 12,155,742 base pairs

In addition, whole genome sequencing of evolved mutants can give genetic proofs for evolutionary theories or related questions as Dettman et al. mentioned in 2012, *How many mutations underlie adaptive evolution, and how are they distributed across the genome and through time? Are there general rules or principles governing which genes contribute to adaptation, and are certain kinds of genes (e.g. regulatory vs. structural) more likely to be targets than others? How common is epistasis among adaptive mutations, and what, if anything, does this reveal about the variety of genetic routes to adaptation? How common is parallel evolution, where the same mutations evolve repeatedly and independently in response to similar selective pressures?* (Dettman, et al., 2012) Phenotypic results of mutations are constrained by evolutionary genetic context such as epistasis, pleiotropy, hitch-hiking of negative mutation with beneficial ones, and so on. Therefore, the whole genome sequencing can give vast amount of possibility for increasing our understanding about evolution itself, and prediction our ability to use evolutionary strategies in setting the further design in strain development.

## 1.4. Improving galactose utilization in *Saccharomyces cerevisiae*

Galactose metabolism in *S. cerevisiae* was selected to generate evolutionary strategies and explore them through genome-scale analyses in this study. The galactose regulon of *S. cerevisiae* has been extensively investigated, since it has very strict gene expression control properties and it is a model system for human disease, galactosemia (Lai, *et al.*, 2009). In addition, yeast strains retaining higher galactose utilization ability have been developed for industrial application, because galactose is one of the abundant renewable carbon sources (Panesar, *et al.*, 2007, Wi, *et al.*, 2009, Kim, *et al.*, 2012). In previous studies, direct genetic engineering approach in galactose metabolism showed successful results (Ostergaard, *et al.*, 2000, Bro, *et al.*, 2005, Garcia Sanchez, *et al.*, 2010, Lee, *et al.*, 2011). For the next turn in the metabolic engineering cycle, new strategies are required.

### 1.4.1. Galactose metabolism in *Saccharomyces cerevisiae*

#### 1.4.1.1. Leloir Pathway

Even though the molecular structure of galactose is very similar with glucose, more enzymatic reactions for galactose utilization are needed to reach glucose-6-phosphate, a precursor of glycolysis. And the number of transporter specialized for galactose is just one, while there are at least 20 transporters for glucose (Boles & Hollenberg, 1997, Ozcan & Johnston, 1999, Wieczorke, *et al.*, 1999). Galactose is metabolized through the Leloir pathway, after the Nobel Prize laureate, biochemist Louis Leloir (Cabib, 1970). This pathway is composed of 5 enzymes: galactose mutarotase (*GAL10*), galactokinase (*GAL1*), galactose-1-phosphate uridyltransferase (*GAL7*), UDP-galactose 4-epimerase (*GAL10*) and phosphoglucomutase (*PGM1/PGM2*), and expression of those enzymes is controlled by very tight regulatory system consisting of 3 regulators, Gal3p, Gal4p and Gal80p (Timson, 2007). Further regulation is mediated by Mig1p, i.e. glucose is present, Mig1p is de-phosphorylated resulting in its transfer into the nucleus where it inhibits expression of *GAL1* and *GAL4* by binding to upstream repression site (URS) of those genes (Timson, 2007) (Fig. 2). Galactose enters the cells mainly through Gal2p, a specific galactose transporter. Intracellular galactose is structurally changed to alpha-D-galactose from beta-D-galactose by Gal10p, and phosphorylated to galactose-1-phosphate with ATP by galactose kinase, Gal1p (Holden, *et al.*, 2003). Galactose-1-phosphate reacts with UDP-glucose resulting in

glucose-1-phosphate and UDP-galactose in a reaction catalyzed by galactose-1-phosphate uridyltransferase, Gal7p. This reaction has been studied more intensively because of its relation with the human disease galactosemia. Failure of this reaction accumulates galactose-1-phophate, which is a marker for diagnostic of the disease (Lai, *et al.*, 2009). The toxicity of high concentration of galactose-1-phosphate is not clear, while there have been several proposes about the reasons such as inhibition of enzymes and ATP drain (Lai, *et al.*, 2009). UDP-galactose is converted into UDP-glucose by a reaction of UDP-galactose 4-epimerase, Gal10p, which is also galactose mutarotase. This enzyme has dual activity, which is a unique feature of *S. cerevisiae* and *Kluyveromyces fragilis* (Thoden & Holden, 2005). Prokaryotes and higher eukaryotes have different enzymes to provide these two enzyme activities (Holden, *et al.*, 2003). Since the Leloir pathway is a highly conserved system in most organisms, and yeast supposedly occupies a position between prokaryotes and higher eukaryotes, this distinctive evolutionary history is an open question. In the last step, glucose-1-phosphate is converted into glucose-6-phosphate in a reaction of isomerization by Pgm1p and Pgm2p. Pgm2p is responsible for about 80% of the total activity (Timson, 2007). Further detailed knowledge about the kinetic properties and structures of the enzymes in the Leloir pathway are well explained in biochemistry references (Daugherty, *et al.*, 1975, Schell & Wilson, 1977, Segawa & Fukasawa, 1979, Fukasawa, *et al.*, 1980, Reifenberger, *et al.*, 1997, Holden, *et al.*, 2003).

#### 1.4.1.2. Regulation of GAL genes

Regulation of *GAL* genes is an excellent model for studying a regulated eukaryal gene expression system (Acar, *et al.*, 2005, Ramsey, *et al.*, 2006, Pannala, *et al.*, 2010). The promoter of the *GAL* genes has been used as a strong expression system with galactose induction (Li, *et al.*, 2008). Each of the galactose catabolism enzymes Gal1p, Gal7p and Gal10p exist at about 0.3 to 1.5% of total soluble cytoplasmic protein during growth on galactose (St John & Davis, 1981). There are three regulation mechanisms. First, the presence of glucose represses expression of the *GAL* genes through the transcription factor Mig1p (Timson, 2007) (Fig. 2). The Mig1p interacts with the transcriptional co-repressor complex Cyc8p (Ssn6p)-Tup1p (Treitel & Carlson, 1995). The complex of these three proteins activates the histone deacetylases Hda1p, Hos1p, Hos2p and Rpd3p (Davie, *et al.*, 2003), which ensures keeping the chromatin deacetylated, compact, and hereby in a transcriptionally inactive state. Second, at high glucose concentrations, the Mig1p is

dephosphorylated and stays in the nucleus where it together with the co-repressors block expression of the *GAL* genes, especially *GAL4* that is an activator of the *GAL* genes. Therefore the presence of glucose completely blocks expression of the *GAL* genes.

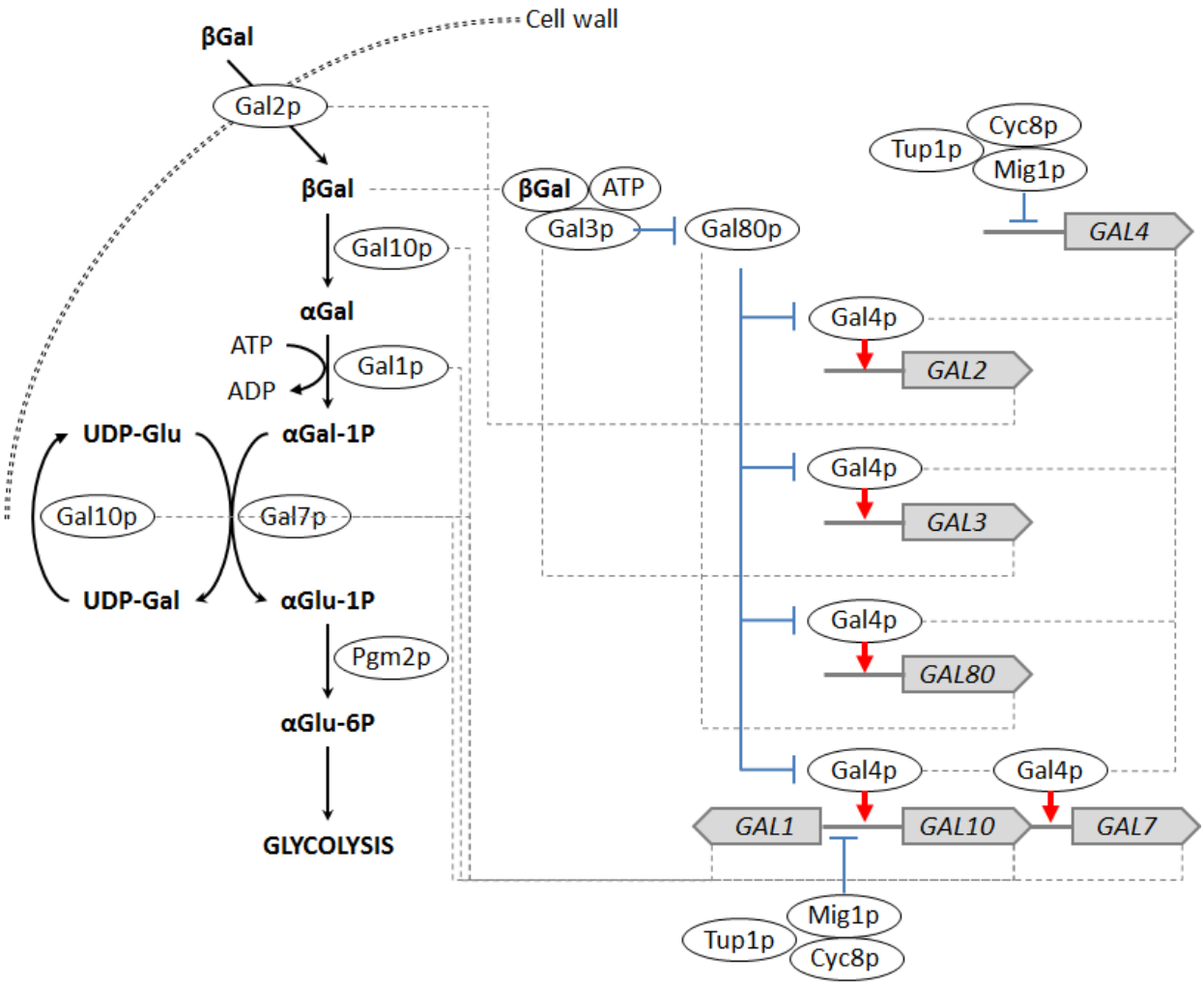


**Fig. 1-2.** Galactose pathway (Leloir pathway) and regulation in *S. cerevisiae*. Pointed arrows mean conversion of intracellular metabolites by enzymatic reactions, red arrows indicate transcriptional activation. Blue blunt arrows mean inhibition. Dotted lines indicate the direct connection between genes and proteins.

The absence of glucose, however, is not adequate to induce the galactose metabolizing enzymes. The existence of galactose is also necessary (Fig. 2). When external galactose is present as a sole carbon source, it can transfer at low rate to the cell through hexose transporters (Hxts), which are not specific to galactose as they have very high Km values for galactose transport. Galactose transporter gene, *GAL2* can be expressed after galactose enters to the cells. Intracellular galactose is combined with Gal3p that is a sensor of galactose. Gal3p binds galactose and ATP, and then traps the repressor Gal80p. Gal80p is present in the cytoplasm and nucleus, which interferes proper binding of Gal4p to the upstream activating sequences ($UAS_{GAL}$) of the GAL genes. Thus, in the absence of galactose, the Gal80p blocks Gal4p and hereby prevents induction of the *GAL* gene expression. When only the complex of Gal3p with galactose and ATP is present, the blocking of the Gal80p binding to Gal4p is released since this complex catches Gal80p (Yano & Fukasawa, 1997). Dual feedback loops have been well elucidated in the galactose control system (Ramsey, *et al.*, 2006). The Gal4p induces not only the Gal2p and Gal3p, but also Gal80p. Induction of Gal2p and Gal3p is positive feed-back loop because the increased expression of Gal2p and Gal3p result in further activation of Gal4p, while induction of Gal80p provides a negative feed-back loop since higher expression of Gal80p blocks Gal4p activation (Fig. 2). Simultaneous operation of these dual opposite controls has provided an excellent model for studies of the dynamics of gene expression regulation in eukaryotes. Third, Lap3p/Gal6p is supposed as a possible regulator, because deletion of this gene increases expression of the *GAL* genes (2.5 fold) (Zheng, *et al.*, 1997). The Lap3p is a cysteine protease and the *S. cerevisiae* homologue of this enzyme is Gal6p. The exact mechanism of how Gal6p carries negative regulation remains unclear (Zheng, *et al.*, 1997).

### 1.4.2. Galactose as a feedstock in industrial biotechnology

#### 1.4.2.1. Galactose content of biomass

In terms of its use as a carbon and energy source for production of fuels and chemicals galactose is mostly found in cheese whey, but with the prospect of using algae as source of biomass it is interesting to note that red seaweed has a high content of galactose (*Gelidium amansii*) (Wi, *et al.*, 2009, Kim, *et al.*, 2012). Cheese whey is an eluent from the dairy industry; and it contains about 85-95% of the milk volume and 55% of milk nutrients. Two types of the cheese whey, sweet (pH 6~7) and acid (pH < 5) are produced dependent on the procedure of

casein precipitation. The main components are lactose, whey protein and minerals (Table 5) (Jelen, 1979, Siso, 1996, Panesar, *et al.*, 2007). Lactose is a disaccharide sugar composed of galactose and glucose through beta-1, 6-linkage. Galactose therefore contains around 22~26 g/l in cheese whey.

**Table 1-5.** Typical composition of sweet and acid whey (Source: Jelen, 1979, Panesar, *et al.*, 2007)

| Components | Sweet whey (g/l) | Acid whey (g/l) |
|---|---|---|
| Total solids | 63-70 | 63-70 |
| Lactose | 45-52 | 44-46 |
| Protein | 6-10 | 6-8 |
| Calcium | 0.4-0.6 | 1.2-1.6 |
| Phosphate | 1-3 | 2-4.5 |
| Lactate | 2 | 6.4 |
| Chloride | 1.1 | 1.1 |

Recently algae have been considered as an attractive biomass source for bio-based products, due to several advantages compared to terrestrial plant biomass such as high production yield, non-food and land usage, little recalcitrant lignin and crystalline cellulose, a higher growth rate and others (Kim, *et al.*, 2011, Wargacki, *et al.*, 2012). One algae, the red seaweed (*Gelidium amansii*) has high galactose content even comparable to the amount of glucose (Wi, *et al.*, 2009, Kim, *et al.*, 2012). Carbohydrate compositions of different biomass sources are given in Fig. 3.

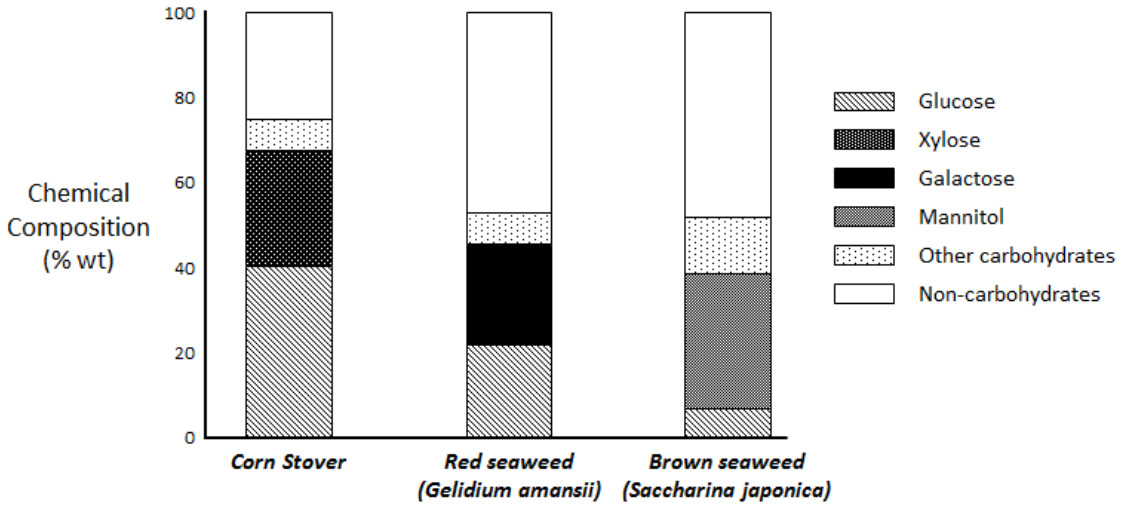


**Fig. 1-3.** Composition of non-food biomass (Source: Kim, *et al.*, 2012)

### 1.4.2.2. Metabolic engineering for improved galactose utilization

Intensive research on galactose metabolism has generated vast amount of information about its metabolic structure and regulation. By exploiting this abundant resource, many elegant metabolic engineering approaches have been executed to improve galactose utilization in terms of a specific galactose uptake rate. Direct genetic modification of *GAL* genes and regulatory genes were done. Over-expression of *GAL* catabolic genes was implemented in a high-copy number plasmid with different combinations (de Jongh, *et al.*, 2008). Over-expression of the *GAL* catabolic genes was expected to increase flux from galactose to glycolysis; however, the result showed reduction of galactose uptake and growth rate. The reason was that changed expression level of the *GAL* genes triggered the fluctuation of concentration of intermediate metabolites in the Leloir pathway. One of them, galactose-1-phosphate, was known as a toxic intracellular metabolite that interfere with galactose metabolism. Therefore genetic modification was focused on regulatory genes. By over-expression of the transcriptional activator, *GAL4* and deletion of negative regulators, *GAL80*, *MIG1* and *GAL6* showed improved galactose uptake rate without growth retardation (Ostergaard, *et al.*, 2000). Especially, the triple knock-out mutant (SO16, which was used as a control strain in this thesis study) showed the highest specific galactose uptake rate. In a follow up study, transcriptome data of these strains was used to find target genes that were related to improvement of galactose availability (Bro, *et al.*, 2005). Commonly changed genes were screened in galactose related pathways and based on this *PGM2*, encoding phosphoglucomutase, was the only gene that showed significant up-regulation. Application of this gene in a high-copy number plasmid clearly showed improvement of galactose utilization (this strain was called PGM2, which was used as another control strain in this study). Since this gene was supposed to be quite highly expressed even at non-galactose growth condition, the result that the rate limiting step enzyme of the Leloir pathway was *PGM2* was surprising. When galactose-1-phosphate was measured, this strain showed no reduction of this metabolite. Higher activity of phosphoglucomutase was checked by checking higher concentration of sugar-6-phosphates that were considered as products of this enzyme such as galactose-6-phosphate, glucose-6-phosphate, mannose-6-phosphate and fructose-6-phosphate. Therefore, even though basal expression level of *PGM2* was relatively higher than other *GAL* genes in the wild-type, over-expression of this gene was still needed to improve galactose utilization. Another study also supported the importance of higher activity of *PGM2* for improved galactose utilization (Lee, *et*

*al.*, 2011). In this study, a genomic library was used to find target genes that were related to galactose utilization. The constructed library was transformed into the wild-type strain, and improved strains were screened. Three beneficial over-expression targets, *SEC3*, *tTUP1*, and *SNR84* were identified. All three targets displayed higher phosphoglucomutase activity. Two of them, Sec3p (phosphomannomutase having activity as phosphoglucomutase) and truncated Tup1p (complex of Mig1p repressor) were confirmative with the previous work due to the function of those genes; while the last target was a new discovery. *SNR84* codes for H/ACA box small nucleolar RNA, and there is no report on the effect of this gene to galactose metabolism. However, higher activity of phosphoglucomutase in the transformant over-expressing *SNR84* proposed a relationship between this gene and galactose metabolism.

# 2. OVERVIEW OF THE THESIS

The motivation of this study is to apply genome-scale analyses for unraveling the molecular and genetic basis of evolutionary strategy of microorganism.

The galactose metabolism of yeast *S. cerevisiae* was chosen as a target for evolution, because of the following reasons.

1) It has relevance for developing yeast strains that can use galactose more efficiently like other hexose carbon sources; since galactose is an abundant sugar in some renewable resources, and *S. cerevisiae* is a vastly useful strain in industrial applications.

2) The galactose metabolism in yeast has been extensively studied, which has led to many trials for the construction of yeast mutant strains by direct genetic engineering. Consequently, several genetic targets related to the improvement of galactose utilization have already been identified, which means it is difficult to find new targets. However, rather less attention was paid on evolutionary engineering, thus if different targets are generated from an evolutionary approach, they are useful for a next round of strain development.

3) Moreover, which was the main purpose of this thesis; can genome-scale analyses be used for the characterization of evolved mutants with the objective to find driving mutations? If this answered positively it could open up for wider use of evolutionary strategies in metabolic engineering.

For these reasons, yeast *S. cerevisiae* CEN.PK113-7D was evolved on galactose minimal media through adaptive evolution for 62 days. Three evolved mutants were generated from independent populations grown in identical serial transfers. Improved galactose utilization ability was confirmed in precisely controlled bioreactors, and genome-scale analyses through transcriptome, metabolome and whole-genome analyses were applied to understand evolutionary strategies of the galactose-evolved yeast mutants. Furthermore, inverse metabolic engineering was applied using identified mutations and new combinations of the genetic changes. The comparison of reconstructed strains with the evolved mutants provided a good example how evolution and engineering work synergistically in strain development. Further characterization of the evolved mutants was done in glucose minimal media to explore the pleiotropy of obtained

traits. Molecular and genetic bases of that pleiotropy were elucidated by genome-scale analyses. The result increased the understanding of evolutionary strategies of the evolved mutants. Consequently, three research studies were designed.

**Paper I:** ***Unravelling evolutionary strategies of yeast for improving galactose utilization through integrated systems level analysis***

Adaptive evolution generated improved galactose availability with different physiology. The molecular and genetic bases that were supposed to be related to improved galactose utilization were analyzed. The significant molecular changes in transcripts and metabolites were detected in around galactose metabolic pathways, but no mutations were found in those regions. Instead the Ras/PKA signaling pathway was detected as a common pathway that had mutations in all the evolved mutants. Introduction of one of those mutations in a reference strain partially provided the genetic bases of the galactose evolved physiology. It was confirmed that adaptive evolution can generate key mutations in unpredictable regions or non-canonical pathways. And the genetic basis (mutations) and resulting molecular basis (transcriptome and metabolome) for evolutionary changes were found to happen in different regions.

**Paper II:** ***Recovery of phenotypes obtained by adaptive evolution through inverse metabolic engineering***

Through adaptive evolution and genome-scale analyses, new genetic targets for improving galactose utilization were identified. As only a few mutations were selected from many mutations, it was necessary to evaluate whether the adaptive phenotype can be recovered by a few mutations. Furthermore, it was speculated how inverse metabolic engineering could give more chances beyond evolutionary engineering itself for strain development. Two groups of engineered mutants were constructed; site-directed mutants that had the identified mutations in the reference strain genetic background, and combined mutants that had new combinations by transforming the *PGM2* over-expression plasmid into the site-directed mutants. Surprisingly, some of the constructed strains showed complete recovery of the galactose adaptive phenotype with just one

or two genetic modifications. Even one of the reconstructed mutants exhibited further improved galactose utilization. These results indicated that far fewer genetic changes were enough to reach the same phenotype as the evolved mutants. Therefore inverse metabolic engineering is an essential step in the application of evolutionary approaches for strain development, i.e. it can enable more strategies for further improvement of desired phenotypes by sieving out beneficial mutations from negative ones and generated new artificial combinations of mutations. Detailed molecular changes by the mutations were also analyzed using transcriptome analysis and the level of a few metabolites. The introduction of key mutations that recovered the adaptive phenotype triggered fewer molecular changes compared to the evolved mutants. This result indicated again that all molecular changes were not necessary for reaching the same phenotype.

**Paper III:** ***Adaptively evolved yeast mutants on galactose show trade-offs in carbon utilization on glucose***

The evolved mutants obtained the trait that was an ability to utilize galactose more efficiently than the ancestor strain. It was, however, interesting to evaluate whether this trait was associated with other effects, i.e. pleiotropy. The galactose-evolved mutants were therefore grown in glucose minimal media. Interestingly, these cultivations showed reduced glucose utilization in the evolved strains. This means that there is trade-off in galactose utilization and glucose utilization. In other words, the evolved mutants likely obtained the increased galactose availability by partly losing their ability to very efficiently utilize glucose. The underlying mechanisms of this trade-off were studied at the molecular and genetic level by integrated genome-scale analyses. Antagonistic pleiotropy was found to be the dominant evolutionary trade-off mechanism. The tight regulation system of glucose catabolic repression was loosened by the mutations in Ras/PKA signaling pathway and unidentified mutations that may be involved in hexokinase regulation and reserve carbohydrates metabolism. Therefore, the glucose utilization ability is likely collateral cost for having improved galactose availability in the evolved mutants. This finding indicates that genetic context such as pleiotropy causing trade-off in traits should be considered, when evolutionary approaches is applied in strain development.

# 3. RESULTS AND DISCUSSION

This section provides a summary of the results, whereas the attached papers in the end of this thesis include detailed materials, methods and experimental design with expanded explanations.

## 3.1. Molecular and genetic basis of evolutionary strategies of the galactose-evolved mutants (Paper I)

Three evolved populations were generated from *S. cerevisiae* CEN.PK113-7D, an ancestor strain, by three independent serial transfers in a galactose (20g/l) minimal medium for 62 days (Fig. 3-1). Single clone isolates were obtained from the last shake flasks, and designated 62A, 62B and 62C. Two engineered strains, SO16 *(gal6*Δ *gal80*Δ *mig1*Δ) and PGM2 (over-expression of *PGM2* gene), showed improved galactose uptake rates in previous studies were used as control strains to elucidate unique strategies of adaptive evolution (Ostergaard, *et al.*, 2000, Bro, *et al.*, 2005).

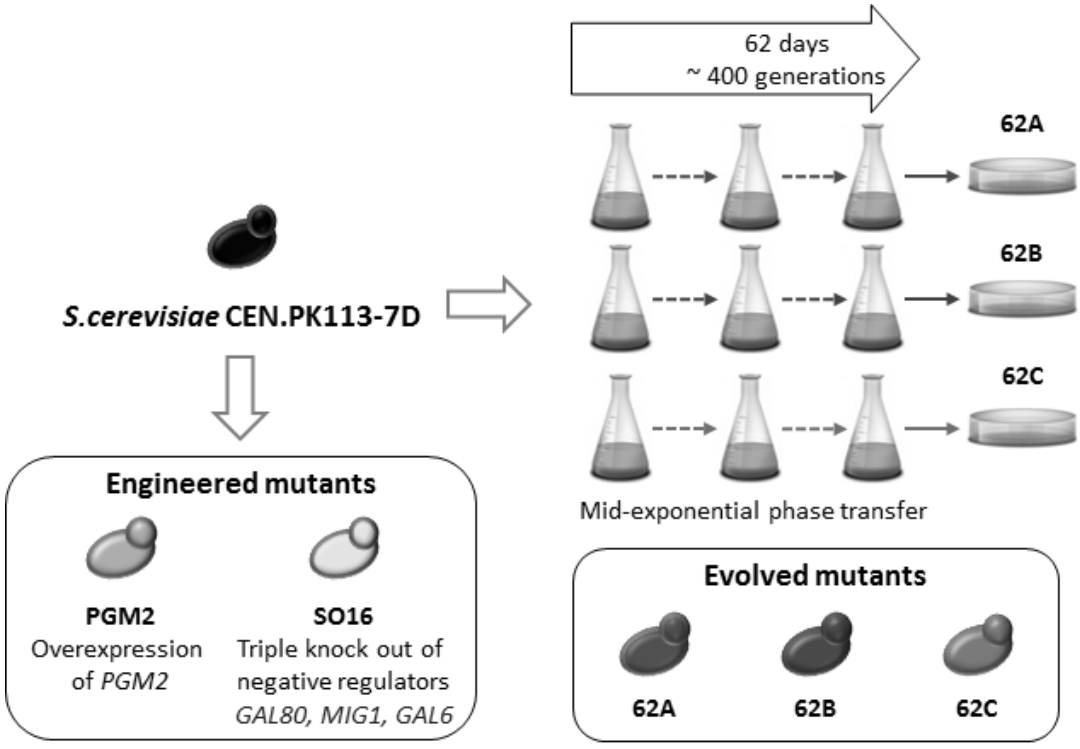


**Fig. 3-1.** Yeast strains used in this study. Engineered mutants were constructed in previous studies (Ostergaard, *et al.*, 2000, Bro, *et al.*, 2005) whereas the three evolved mutants were generated in this study.

Parallel fermentations at aerobic batch mode in precisely controlled bioreactors was performed to estimate physiological parameters and to take samples for omics analyses; and transcriptome and targeted metabolome analysis were applied to all strains including the two engineered strains; and whole-genome analysis were performed on the evolved mutants (Fig. 3-2).

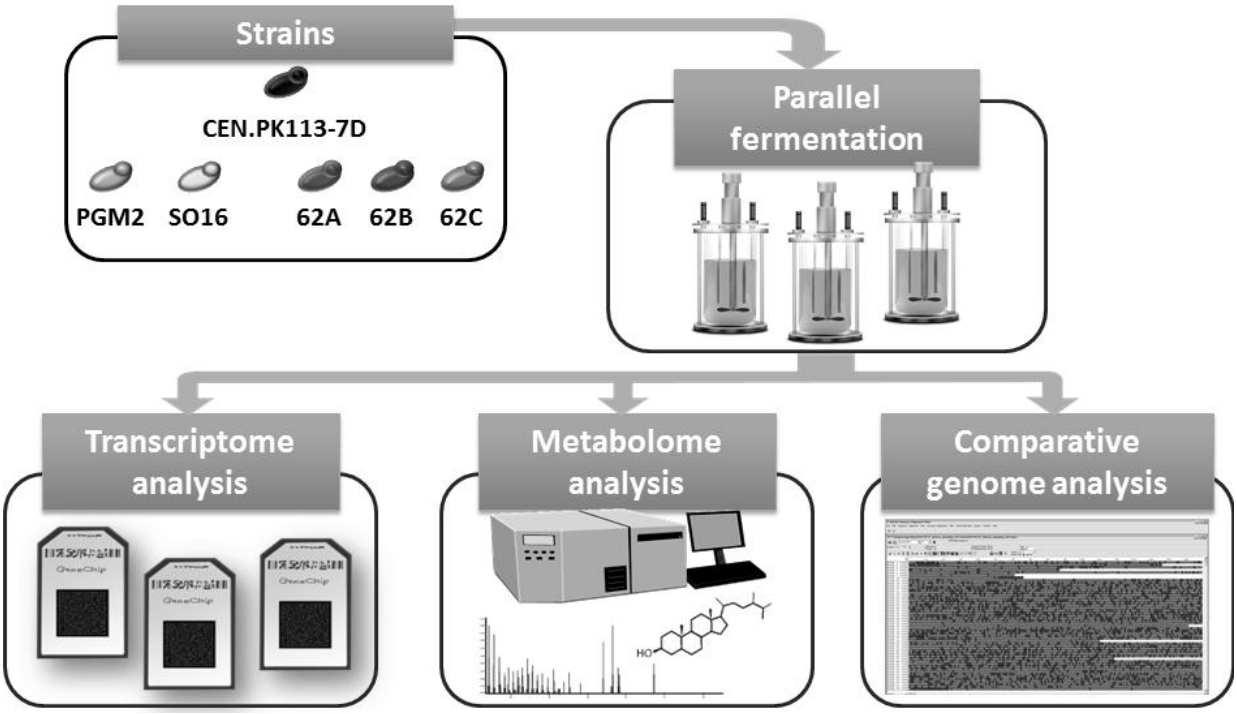


**Fig. 3-2.** Overall flow of the experiments in this study. 6 strains were cultivated in bioreactors, and at mid-exponential phase, samples for omics analyses were collected.

Exact identification of improved galactose utilization in the evolved mutants was compulsory, since only one colony from each of the populations were selected. Population was composed of many diverse individuals, thus it was not sure if the selected colony was really evolved in terms of improved galactose utilization. Of course, based on Darwin's theory, natural preservation, the variants that had higher fitness would take more portions in the population; therefore, there was high chance to select evolved clones with improved fitness. The purpose of this study was to detect evolutionary strategies; hence the confirmation of improved phenotype in the evolved mutants was a prerequisite. The evolved mutants achieved improved galactose utilization in terms

of a maximum specific growth rate and a specific galactose uptake rate, which were a different phenotype compared with the engineered strains (Fig. 3-3). The galactose-evolved mutants showed a 24% increase in the maximum specific growth rate and 18 ~ 36% increase in the specific galactose uptake rate compared to the reference strain. Interestingly, clear grouping was observed between all three evolved mutants and the reference strains in the plot of the specific galactose uptake rate versus a specific ethanol production rate (Fig. 3-3B). These two groups were separated by different regression curves, which observation means that the adaptive evolution has resulted in different phenotypes compared with the engineered strains.

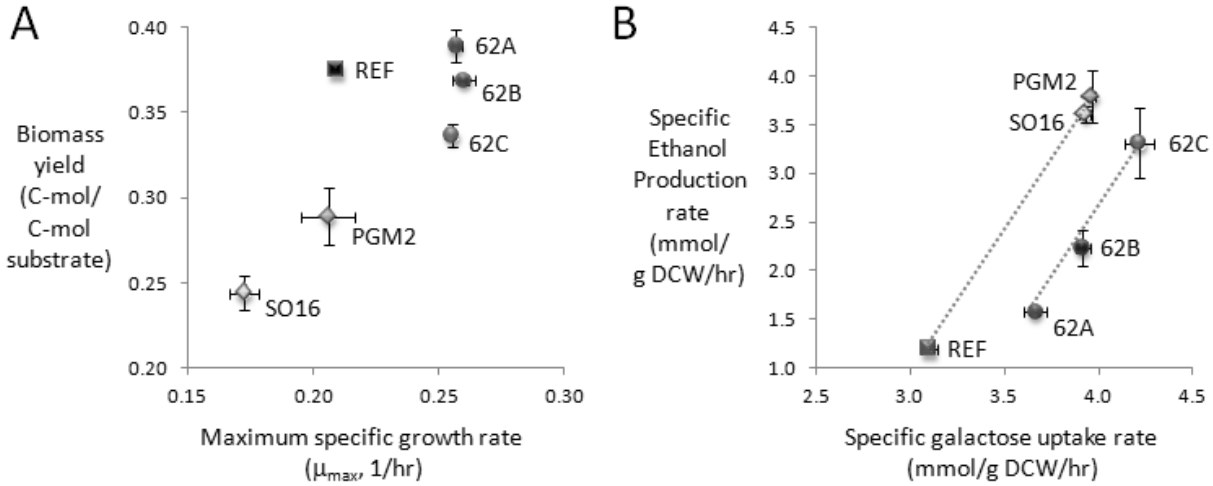


**Fig. 3-3.** Phenotypic changes of evolved mutant strains 62A, 62B and 62C compared with the reference strain CEN.PK113-7D and the two engineered strains SO16 and PGM2. (A) Correlation between a maximum specific a growth rate and biomass yield. (B) Correlation between a specific galactose uptake rate and a specific ethanol production rate. The regression curves of the two lines (from right to left) had a slope of 2.95 and 3.19 and intercept of minus 7.95 (R2 = 0.99) and minus 10.157 (R2 = 0.98), respectively. Both slope values were around 3, which indicated that catabolic repression induced flux re-direction from respiratory metabolism to fermentation one, because if there was not that repression, slope should be around 2.

To investigate the molecular basis, firstly transcriptome analysis was used to check overall changes, and select significantly altered pathways (Paper I, Fig. 2). Secondly, target metabolome was implemented; around 40 metabolites were measured based on the results from the transcriptome data and quantified by diverse analytical instruments (Paper I, Fig. 3). Both data

sets were used to find the molecular basis for the evolutionary strategies by selecting commonly changed metabolisms in all evolved mutants (Fig. 3-4).

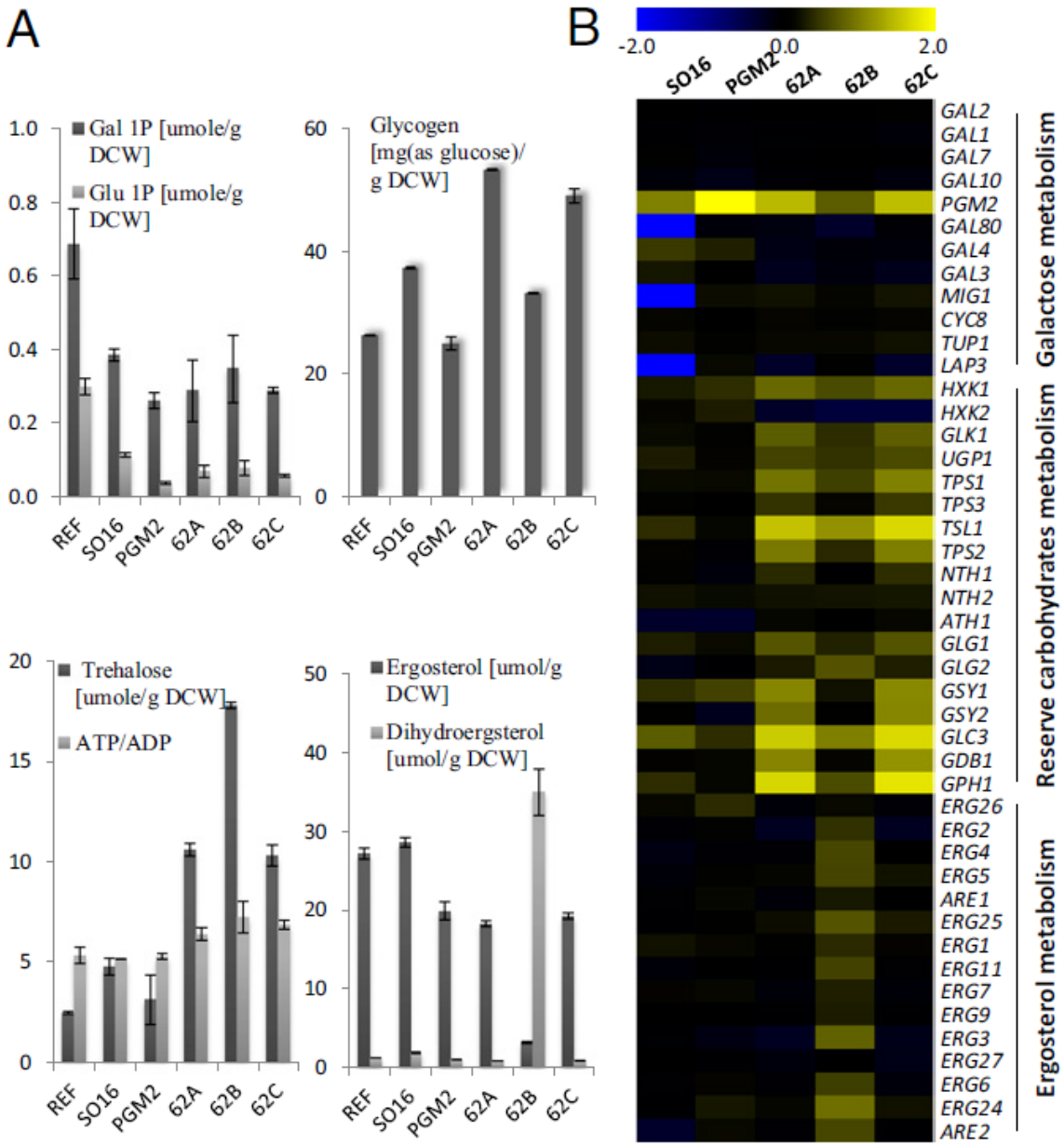


**Fig. 3-4.** Changes in the galactose, reserve carbohydrates, and ergosterol metabolism in the evolved mutants are illustrated by changes in the concentration of metabolites and fold changes of transcriptome compared with the other strains. (A) The concentrations of sugar phosphates, storage carbohydrates, and sterols and the ratio of ATP to ADP. (B) Fold changes of all genes involved in galactose, reserve carbohydrates, and ergosterol metabolism are compared with the reference strain.

Up-regulation of the *PGM2* gene and lower concentration of galactose-1-phosphate and glucose-1-phosphate in the galactose pathway were common in all evolved mutants and engineered mutants compared to the reference strain; while up-regulation of genes in reserve carbohydrates metabolism and down-regulation of *HXK2* that was one of the main glucose catabolic repression controllers (Gancedo, 1998), were unique for the evolved mutants. A unique change among evolved mutants was found in ergosterol metabolism. The 62B strain only showed up-regulation of *ERG* genes with different ratio of the concentration of ergosterol and dihydroergosterol. In terms of fermentation physiology, there was big difference between the 62A and 62C strains; however, in terms of transcriptome and metabolome data, they looked almost identical. 62B was positioned between them in terms of gross physiology, whereas this evolved mutant showed vast differences in the transcriptome and the metabolome. The reason of differences among the evolved mutants was not clear, while the common changes of the evolved mutants from the reference strain likely explained the molecular bases of evolutionary strategies for improving galactose utilization.

To identify the genetic basis of evolutionary strategies, whole-genome sequencing was applied to the evolved mutants (explanation of overall process and raw data are in the supplementary data of Paper I). Surprisingly, there were no mutations or duplications in the *GAL* genes and the regulatory genes involved in galactose metabolism including their promoter and terminator regions. The reaction step by *PGM2* was earlier found as a rate-controlling step in galactose metabolism; hence *PGM2* over-expression was already proven as a beneficial target for metabolic engineering, and several genetic modifications that induced higher expression of this gene were also identified. However, mutations from previous studies were not detected in the evolved mutants. This result indicated that new mutations that induced up-regulation of *PGM2* were generated. Furthermore, genes of the reserve carbohydrates metabolism and hexokinases had no mutations, even though they showed significant alteration in their transcription. Exceptionally, the 62B evolved mutant had mutation in the *EGR5* gene that seemed to be related to changes in ergosterol metabolism (Kelly, *et al.*, 1995) (Table 3-1). No mutations in the metabolisms that showed common molecular changes in all the evolved mutants implied that the key mutations may be involved in regulatory regions. Common genes, pathway or cellular metabolism that had mutation in all three evolved mutants was searched, and it was found that there were common mutations in the regulatory, Ras/PKA signaling pathways (Table 3-1). The

Ras/PKA signaling pathway has been known to take key role in global regulation of glucose sensing and stress response (Estruch, 2000). And *PGM2* and *UGP1* had STER element in their promoter region. Therefore, the mutations in Ras/PKA signaling pathway were suggested as a driven mutation that increased galactose utilization by triggering the activation of galactose and reserve carbohydrates metabolism (Fig. 3-5). Introduction of mutations in *RAS2* genes into a reference strain clearly showed the increase of galactose utilization (supplementary data from Paper I and Paper II). The 62B unique mutation in the *EGR5* gene could explain the changes of transcripts and metabolites in the ergosterol pathway; it may also be explicable why this mutant showed large differences from other evolved mutants by this mutation. However it was not clear how there is a relationship between galactose metabolism and ergosterol pathway.

**Table 3-1.** Genetic changes in the evolved mutants

| Strains | Mutations | Functions | Specific features |
|---|---|---|---|
| 62A | *RAS2* [$Gln^{77} \rightarrow Lys$] | Ras/PKA signaling pathway | Commonly mutated pathway |
| 62B | *RAS2* [$Asp^{112} \rightarrow Tyr$] | | |
| 62C | *CYR1* [$Asp^{822} \rightarrow Asn$] | | |
| 62B | *ERG5* [$Arg^{370} \rightarrow Pro$] | Ergosterol metabolism | Uniquely mutated gene |

In conclusion, key genetic changes were identified in non-canonical metabolism, but in the Ras/PKA signaling pathway; which meant no mutation detected in galactose metabolism not like other direct genetic engineering studies (Ostergaard, *et al.*, 2000, Bro, *et al.*, 2005, Lee, *et al.*, 2011). And, molecular changes were well related to canonical metabolisms; up-regulation of *PGM2* in galactose metabolism, and up-regulation of genes in reserve carbohydrates metabolism that shared the intermediate of galactose metabolic pathway, i.e. glucose-1-phosphate. Hypothetical evolutionary changes were plotted in Fig. 3-5. Therefore, insight about evolutionary strategy that results in non-canonical genetic changes with canonical molecular changes could be applied as an evolutionary approach in strain development.

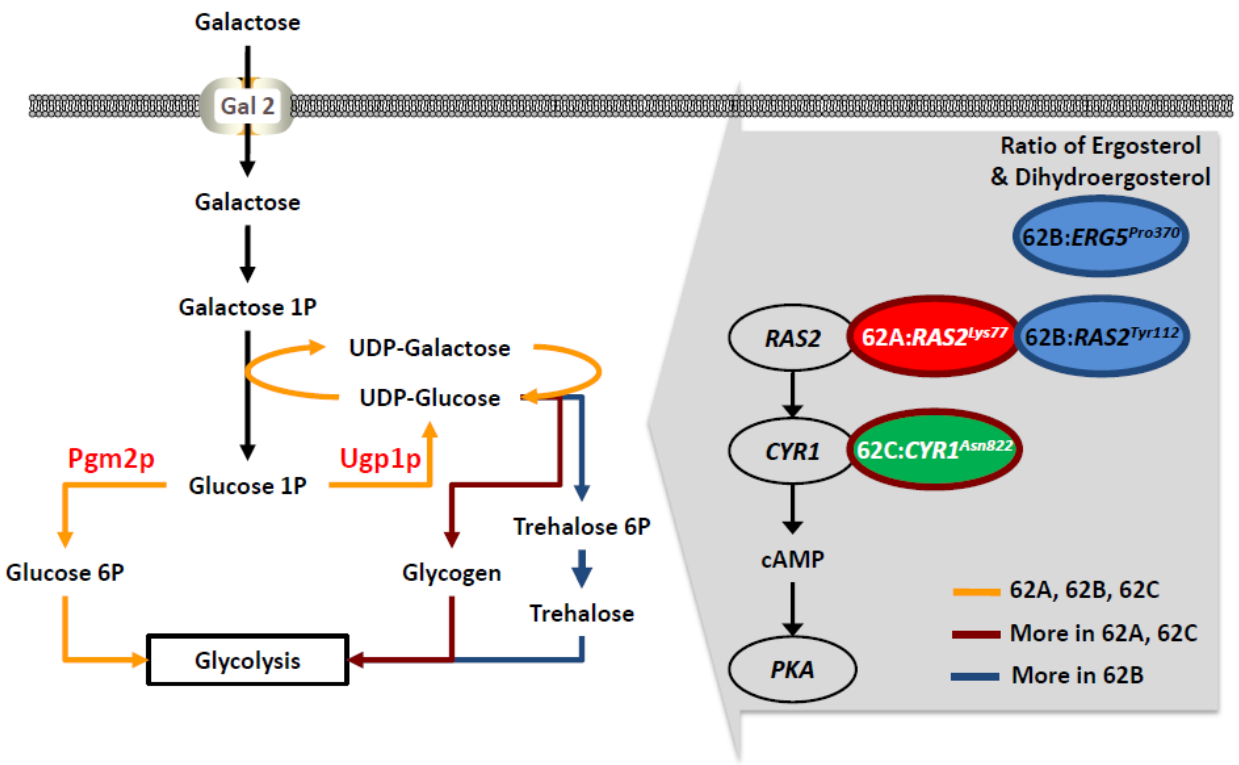


**Fig. 3-5.** Summary of evolution changes in the three evolved mutants; 62A, 62B, and 62C. Color circular boxes indicate genes having genetic mutations. Color lines indicate activated fluxes inferred from transcriptome and metabolome analysis.

### 3.2. Complete recovery of adaptive phenotype through inverse metabolic engineering (Paper II)

The initial purpose of this step was to evaluate the identified mutations. Furthermore, the objective was to explore how inverse metabolic engineering was useful in connection with evolutionary approaches for strain development. Site-directed mutants and combined mutants were constructed for the application of inverse metabolic engineering.

The genome-scale analyses suggested that the genetic basis for the improved galactose utilization could be present in the Ras/PKA signaling pathway, which is not directly involved in galactose metabolism, since all the evolved mutants commonly had mutations in this signaling pathway. One of the evolved mutants, 62B showed significant changes in ergosterol metabolism both at the level of transcripts and metabolites and it carried a mutation in the *ERG5* gene. To clearly confirm the effect of identified mutations on galactose availability, and to examine how much those mutations recovered the adaptive phenotype of the evolved mutants, site-directed mutants carrying each of the mutations independently were constructed. In addition, combined mutants were constructed by introduction of the known beneficial change (*PGM2* overexpression) into the site-directed mutants. These combined mutants were designed to generate new combination of the genetic basis for improving galactose utilization that was not present in the evolved mutants (Table 3-2).

The gross phenotype of the reconstructed strains was compared to the evolved strains (Fig. 3-6). The results of the site-directed mutants clearly confirmed the effects of the identified mutations on galactose utilization. Two site-directed mutants (RAU and RBU) that had mutations in the *RAS2* gene showed a significant increase in the maximum specific growth rate and the specific galactose uptake rate compared with their reference strain (5DU). Especially, the RAU strain that carried the mutation $RAS2^{Lys\ 77}$ exhibited the highest specific galactose uptake rate among all the strains including the evolved mutants. Additionally, when its improvement of maximum specific growth rate was compared to the evolved mutants in terms of increased extent from each of their reference strains, i.e. RAU from 5DU, and the evolved mutants from 7D, even the RAU has a higher relative increase in the specific growth rate than the evolved mutants. Interestingly, even though two mutations were positioned in the same gene, their effect on galactose utilization was quite different. These results highlighted why the concept of inverse

metabolic engineering was important for strain development, because new targets from evolutionary engineering should be re-evaluated and there was space for more improvement of the desired traits by surveying more mutations in that target gene. The results of the combined mutants also showed improvement of the galactose availability. They almost fully recovered the adaptive phenotype of the evolved mutants, since the maximum specific growth rate and the specific galactose uptake rate were in the same level as the evolved mutants. This result again confirmed the importance of inverse metabolic engineering in connection with evolutionary approaches for strain development, because the same phenotype was realized with much fewer traceable genetic modifications providing more space for new engineering strategies.

**Table 3-2.** Reconstructed strains and control strains. *Saccharomyces cerevisiae* CEN.PK113-5D was used to construct site-directed mutants and combined mutants due to its availability of *URA3* marker gene. Prototrophic site-directed mutants (RAU, RBU and EBU) were constructed by transformation with the plasmid, *pSP-GM2* containing the *URA3* gene. The combined mutants RAP, RBP and EBP were constructed by transformation of the plasmid *pPGM2* into the site-directed mutants.

| Strains | Ancestor strains and Genotype | Groups | References |
|---|---|---|---|
| 7D | *MAT**a** SUC2 MAL2-8$^{c}$* (CEN.PK113-7D) | Reference strain | SR&D* |
| 62A | 7D, total no. SNPs: 21 including *RAS2* $^{Lys\ 77}$ | Evolved mutants | This study |
| 62B | 7D, total no. SNPs: 104 including *RAS2* $^{Tyr112}$, *ERG5* $^{Pro\ 370}$ | | |
| 62C | 7D, total no. SNPs: 29 including *CYR1*$^{Asn822}$ | | |
| 5D | *MAT**a** SUC2 MAL2-8$^{c}$ ura3-52* (CEN.PK113-5D) | | SR&D* |
| 5DU | 5D, *pSP-GM2*(*URA3*) | Reference strain | This study |
| **RAU** | 5D, *pSP-GM2*(*URA3*); *RAS2* $^{Lys\ 77}$ (from 62A) | **Site-directed mutants** | This study |
| **RBU** | 5D, *pSP-GM2*(*URA3*); *RAS2* $^{Tyr112}$ (from 62B) | | |
| **EBU** | 5D, *pSP-GM2*(*URA3*); *ERG5* $^{Pro\ 370}$ (from 62B) | | |
| PGM2 | 5D, *pPGM2*(*URA3*, $P_{PMA1}$-*PGM2*) | Engineered mutant | Bro et al 2005 |
| **RAP** | 5D, *pPGM2*(*URA3*, $P_{PMA1}$-*PGM2*); *RAS2* $^{Lys\ 77}$ | **Combined mutants** | This study |
| **RBP** | 5D, *pPGM2*(*URA3*, $P_{PMA1}$-*PGM2*); *RAS2* $^{Tyr112}$ | | |
| **EBP** | 5D, *pPGM2*(*URA3*, $P_{PMA1}$-*PGM2*); *ERG5* $^{Pro\ 370}$ | | |

*Scientific Research & Development GmbH, Oberursel, Germany.

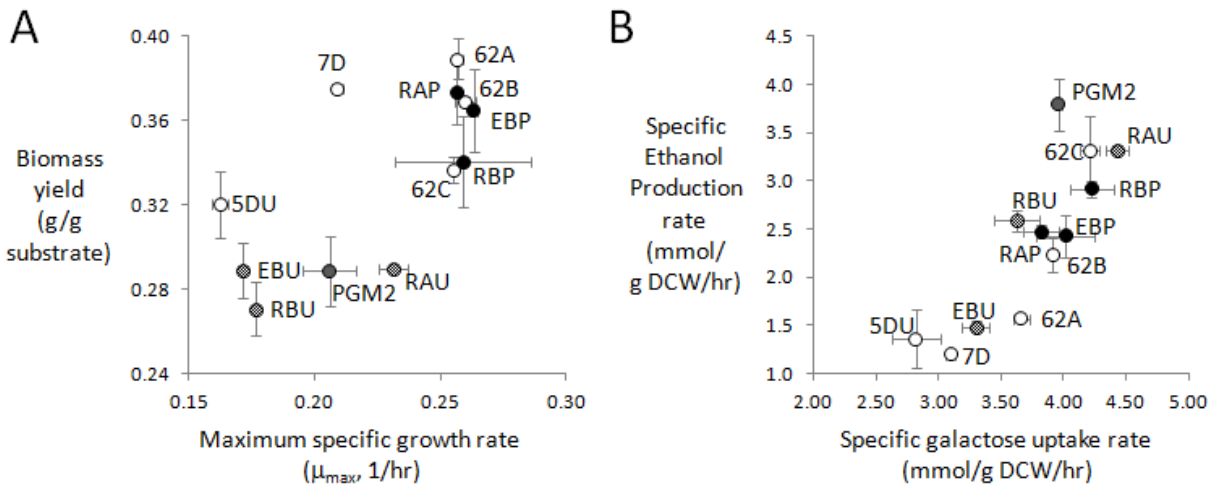


**Fig. 3-6.** Data on overall fermentation physiology of the site-directed mutants (RAU, RBU and EBU) and the combined mutants (RAP, RBP and EBP) are compared to the reference strains 5DU, 7D and the engineered strain PGM2, and the corresponding evolved mutants 62A and 62B. A: Correlation between the maximum specific growth rate and biomass yield B: Correlation between the specific galactose uptake rate and the specific ethanol production rate. Error bars represent standard deviation from biological duplicates.

The critical points were, 1) the evolved mutants accumulated many genetic changes that seemed to be not necessary for improving galactose utilization, because the reconstructed strains showed full recovery of the galactose adaptive phenotype with much fewer genetic changes; 2) the combination effect of genetic changes was different from the sum of each of the changes, for example, the RAP strain that contained a combination of RAU (*RAS2*$^{Lys\ 77}$) and *PGM2* overexpression, showed an increase of the maximum specific growth rate (58% from 5DU); however, the sum of each of genetic changes was bigger (69% = 42% (RAU) + 27% (PGM2 strain)). This phenomenon was much clearer in the specific galactose uptake rate; the combination (RAP) showed even reduced value compared with the RAU and PGM2 strains. Another combination case, the RBP strain with a combination of RBU and PGM2, also showed a negative synergy of beneficial genetic changes even though the extent of the physiological changes was different. On the contrast, the combination of *ERG5* mutation and *PGM2* over-expression looked like synergetic epistasis. The mutation in the *ERG5* gene showed only a small effect on galactose utilization. It looked almost neutral when it was solely present, while the

combination of this mutation with the over-expression of *PGM2* presented the same phenotype like the other combined mutants. In evolutionary biology, the accumulation of negative or neutral mutations and epistasis among mutations is a well-known event during adaptive evolution (Ikeda, *et al.*, 2006, Warner, *et al.*, 2009). That was one of the reasons why cells may not always reach to the optimum point of a specific trait by adaptive evolution, especially in asexual reproduction. Even for a versatile biological system, natural selection or preservation that could enrich only beneficial mutations would possibly require infinite generation time. Thus, to reach the optimum point by laboratorial adaptive evolution could be almost impossible (Sauer, 2001). Therefore, there is likely space for further improvement of desired traits by removing negative mutations and reconstruction of new combinations that may generate synergetic epistasis. Because of these reasons, inverse metabolic engineering is an essential step in evolutionary approaches for strain development.

The molecular basis of the reconstructed strains was investigated to clarify the relationship between the identified mutations and the molecular changes of transcripts and metabolites in specific pathways. First, the overall number of differentially expressed genes was compared (Fig. 3-7). Like the case of genetic changes, the reconstructed mutants showed a much smaller number of differentially expressed genes than the evolved mutants. It confirmed again that many changes in the evolved mutants were not necessary to reach the same phenotype. Second, the detailed molecular changes indicated that the mutations in the *RAS2* gene induced *PGM2*, but not reserve carbohydrates metabolism (Fig. 3-8). This result indicated that there were unidentified mutations triggering up-regulation of reserve carbohydrates metabolism. Maybe the up-regulation of this metabolism was not closely related to improving galactose utilization, or there would be negative epistasis between the mutations in Ras/PKA signaling pathway and the unidentified mutations that activated the reserve carbohydrates metabolism. Both cases could explain the recovery of the galactose adaptive phenotype by the mutations in the *RAS2* gene. Another finding is that the two mutations in the *RAS2* gene showed substantial difference in terms of molecular changes. The RAU strain ($RAS2^{Lys\ 77}$) showed much fewer numbers of transcriptional changes than RBU ($RAS2^{Tyr112}$), while the RAU exhibited higher improvement of galactose utilization than RBU (Fig. 3-7). This finding again emphasized the space for further improvement of galactose utilization by inverse metabolic engineering. The *ERG5* mutation was confirmed as a reason of the changes in the ergosterol pathway (Paper II).

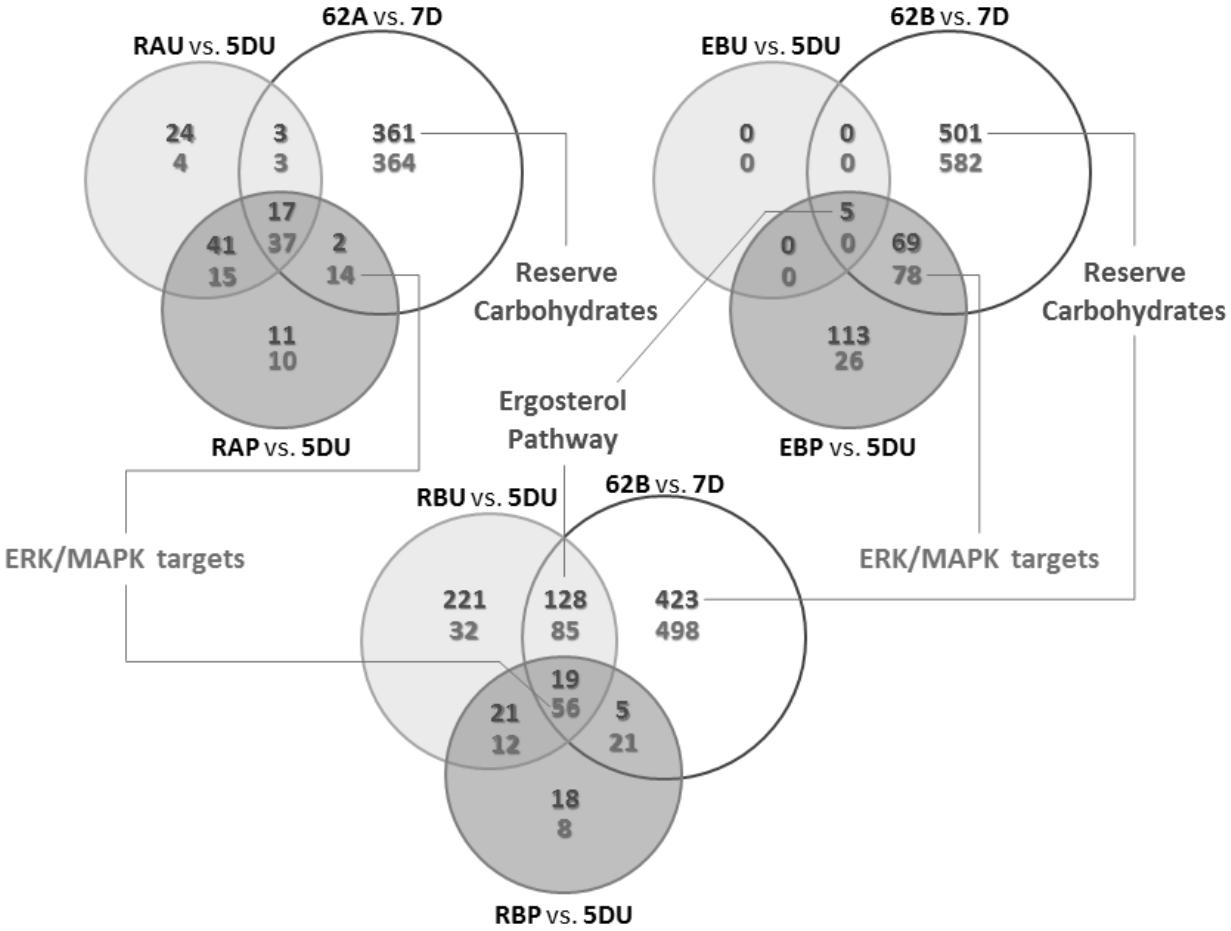


**Fig. 3-7.** Effect of reconstructed strains compared to the evolved strains 62A and 62B by differentially expressed genes. Differentially expressed genes (adjust $p < 0.01$) are categorized as Venn diagrams. The functional enrichment of genes in each part was analyzed by hyper-geometric distribution based on the KEGG, Reactome and GO term databases. Upper numbers in a pair of two numbers mean up-regulation and lower number mean down-regulation.

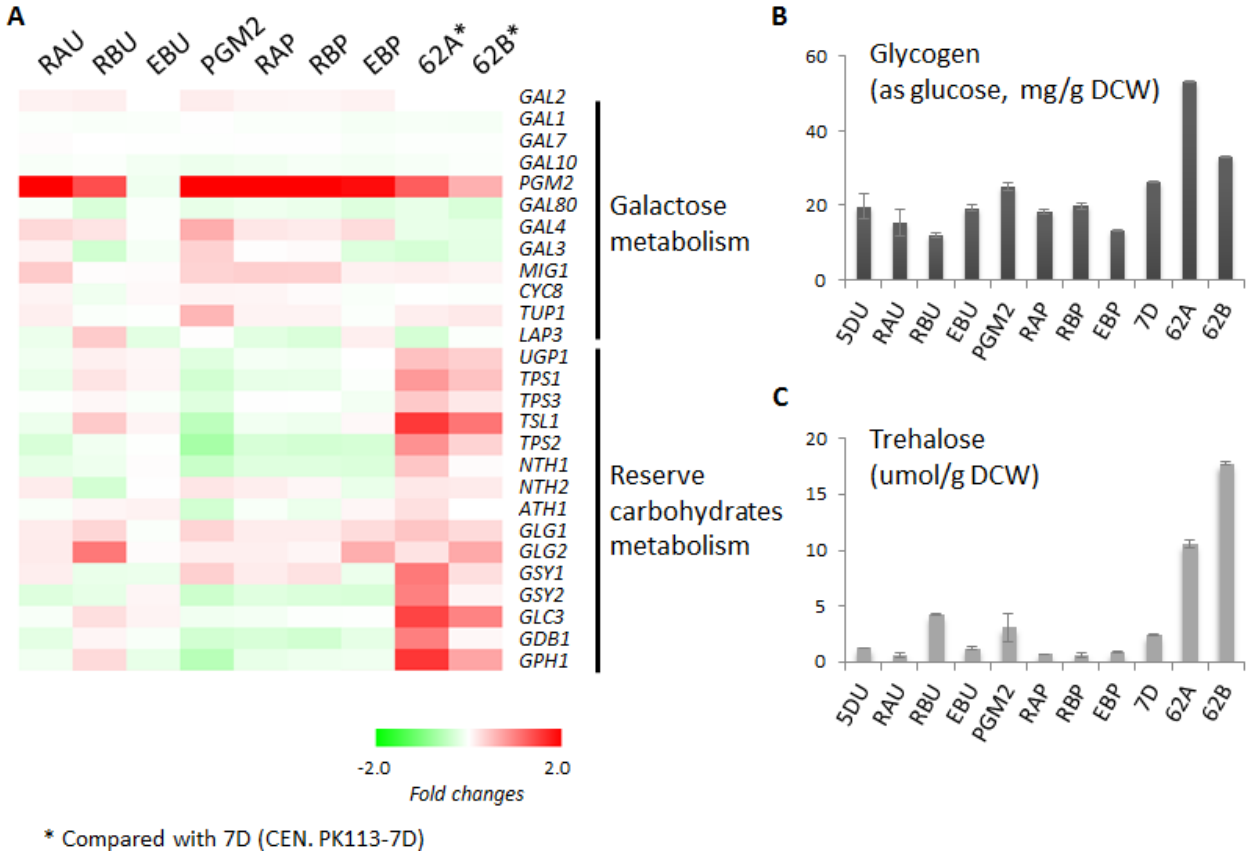


**Fig. 3-8.** Changes in the galactose and reserve carbohydrates metabolisms in the reconstructed strains are shown by changes in fold changes of the transcriptome and the concentration of carbohydrates. A: Fold changes of all genes involved in galactose and reserve carbohydrates metabolisms are compared to the reference strains. B: The concentrations of glycogen. C: The concentration of trehalose. Error bars represent standard deviation from biological duplicates.

### 3.3. Characterization of molecular mechanism of trade-offs in carbon utilization (Paper III)

When galactose is used as a carbon source in industry, glucose would almost always also be present. Therefore, further characterization of the galactose-evolved mutants for growth on glucose was carried out. In addition, it was wondered if there was another effect from the obtained phenotype or traits, or there was collateral cost to get new traits. Considering the short adaptive evolutionary history of the evolved mutants to grow faster on galactose compared to the millions of year of evolution to maximize growth on glucose, a decline in glucose utilization to compensate for the cost of improving galactose utilization was not expected. However, interestingly all galactose-evolved mutants showed reduced glucose utilization (Fig. 3-9). In other word, the trade-off in carbon utilization between galactose and glucose was clearly detected in the evolved mutants. Two engineered mutants, PGM2 and SO16 strains also showed the trade-offs in the specific carbon uptake rate and the specific ethanol production rate (Fig. 3-9B). Since the genetic changes of these engineered strains were known, the genetic bases of this trade-off were easily identified. However, in case of evolved mutants, they showed different pattern of trade-offs, for example the maximum specific growth rate (Fig. 3-9A). Characterization of molecular and genetic bases of this trade-off was the main purpose of this study.

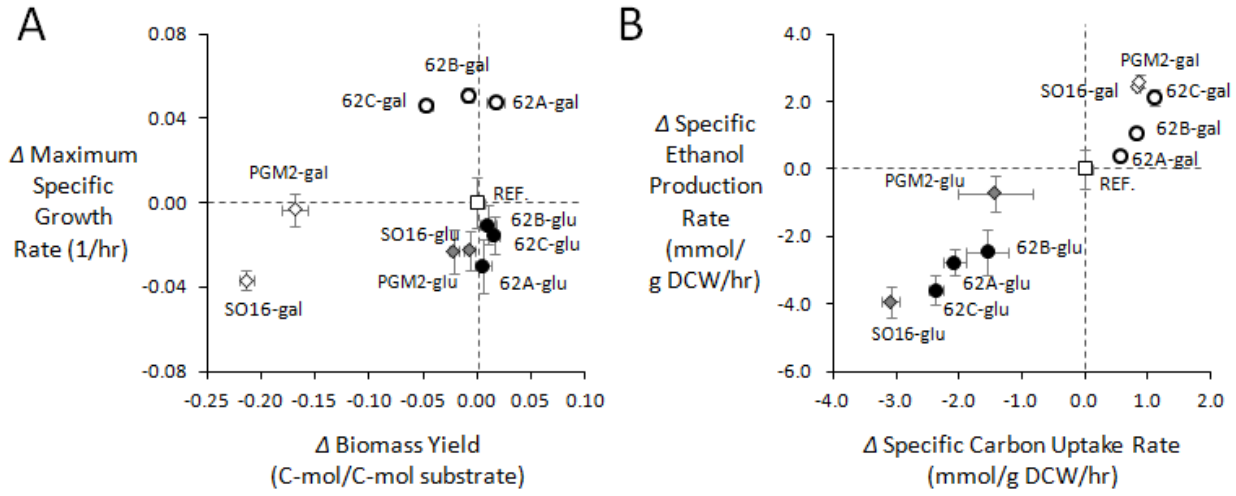


**Fig. 3-9.** Fermentation physiology of the evolved mutants and the engineered mutants compared to the reference strain in galactose (gal) and glucose (glu) through correlation between different values (Δ) of maximum specific growth rate and biomass yield (A), specific carbon uptake rate and specific ethanol production rate (B). Error bar represents standard error from biological duplicates in bioreactors.

Trade-offs among traits is one of the fundamental concepts in evolutionary biology. Two mechanisms for evolutionary trade-off have been suggested (Cooper & Lenski, 2000, Elena & Lenski, 2003, Wenger, *et al.*, 2011); antagonistic pleiotropy (AP) in which the same mutation is related to gain and loss of adaptation in different environment, and mutations accumulation (MA) where different mutations are responsible for the gain and loss of adaptation. Characterization of the trade-off mechanisms is important in the evolutionary approach for strain development, since the strategy for inverse metabolic engineering will dependent on the reason for evolutionary trade-off, AP or MA.

In this study, integrated genome-scale analyses were again applied to elucidate molecular and genetic evolutionary mechanism of the trade-off in the evolved mutants. Firstly, overall transcriptome profile was compared by principal component analysis; the distance between the evolved mutants and the reference strains looked almost identical during growth on both carbon sources (Fig. 3-10). This means that the evolved mutants responded to both carbon sources by similar transcriptional changes. More detailed molecular changes were analyzed by comparison of the differentially expressed genes and functional enrichment of them.

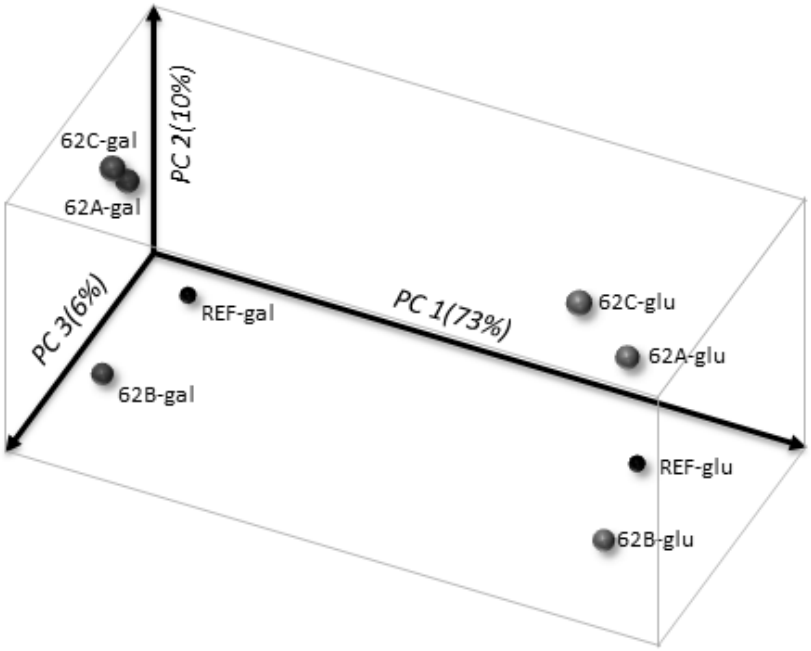


**Fig. 3-10.** Transcriptome analysis of the evolved mutants and the reference strain in galactose (gal) and glucose (glu) through principal component analysis (PCA). The results are projected by the first three PCs, which covered 89% of the variance.

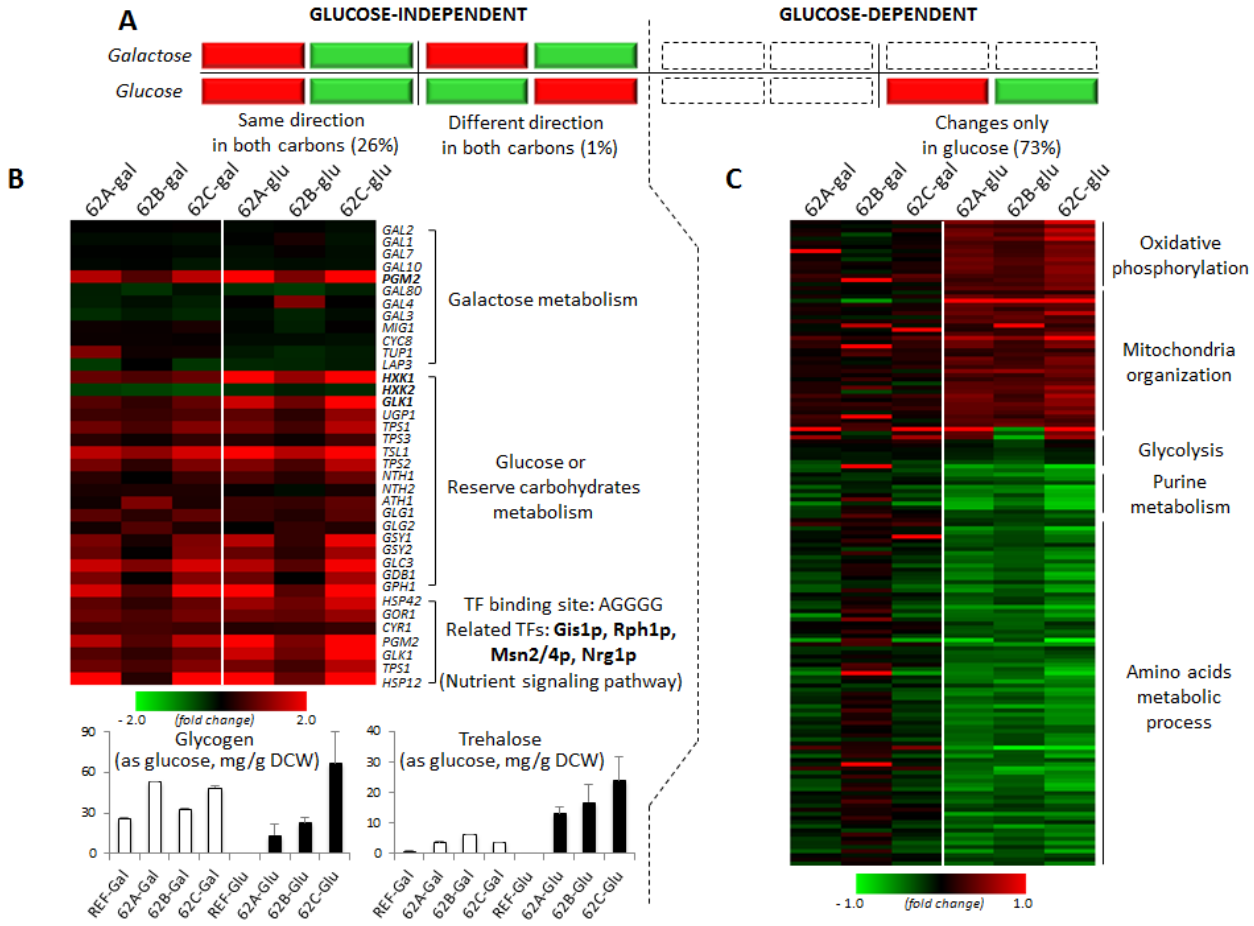


**Fig. 3-11**. Patterns of common molecular changes of the evolved mutants compared to the reference strain based on separation of glucose-independent and glucose-dependent. Glucose-independent, differentially expressed genes in both carbons; glucose-dependent, differentially expressed genes only in glucose (A), Specific pathways and targeted metabolites in glucose-independent (B), Specific pathways in glucose-dependent (C). Error bars in the concentration of glycogen and trehalose represent standard error from biological duplicates in bioreactors.

Conserved pattern of transcripts in specific parts of the metabolisms and reserve carbohydrates were detected (Fig. 3-11); and specific molecules that were likely involved in the trade-off mechanism were identified such as up-regulation of 1) *PGM2*, 2) two non-glucose inducible hexokinase *HXK1*, *GLK1* and 3) genes in reserve carbohydrates metabolism, and down regulation of 4) *HXK2* that is one of the key enzymes of glucose metabolism and is also a regulator of glucose catabolic repression (Gancedo, 1998). Additionally, commonly up-regulated genes on growing both carbon sources had the same transcription factor (TF) binding site

(AGGGG). And TFs related to this site were Gis1p, Rph1p, Msn2/4p and Nrg1p, which are involved in nutrient signaling pathway (Orzechowski Westholm, *et al.*, 2012). These results at the molecular level implied that antagonistic pleiotropy was the dominant mechanism for the trade-off, and the result was loosening the tight glucose control of metabolism.

The genetic bases of the trade-off in carbon utilization were explored. There were three identified mutations; two of them were confirmed as a beneficial mutation for galactose utilization such as mutations in the *RAS2* genes, and one of them was neutral for galactose availability. The site-directed mutants that had each of those mutations supported antagonistic pleiotropy as the mechanism for trade-off (Fig. 3-12).

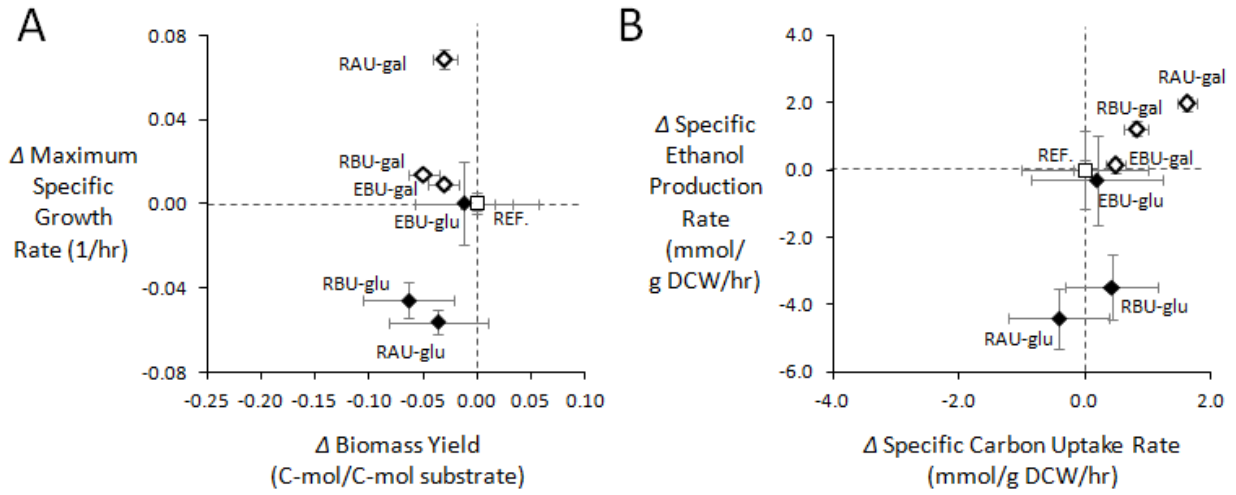


**Fig. 3-12.** Fermentation physiology of the site-directed mutants compared to the reference strain in galactose (gal) and glucose (glu) through correlation between different values (Δ) of maximum specific growth rate and biomass yield (A), specific carbon uptake rate and specific ethanol production rate (B). A reference strain for site-directed mutants is CEN.PK 113-5D having URA3 marker in plasmid. Error bar represents standard error from biological duplicate on galactose in bioreactors and biological triplicate on glucose in baffled flasks. Longer error bars in the reference were from glucose culture, shorter ones came from galactose culture.

The Ras/PKA signaling pathway is involved in the control of transcription factors, Gis1p, Rph1p, Msn2/4p and Nrg1p (Orzechowski Westholm, *et al.*, 2012). The identified mutations in the *RAS2* gene showed up-regulation of *PGM2* in the previous study (Paper II). The mutations

that triggered the molecular changes in reserve carbohydrates metabolism and hexokinases are not still clear. There would be unidentified mutations, which mutations may adjust the change between maximum specific growth rate and specific glucose uptake rate, because that change was the main difference between the evolved mutants and the site-directed mutants containing mutations in the *RAS2* gene. Hypothetical interpretation of the trade-off mechanism in the galactose evolved mutants is illustrated in Fig. 3-13.

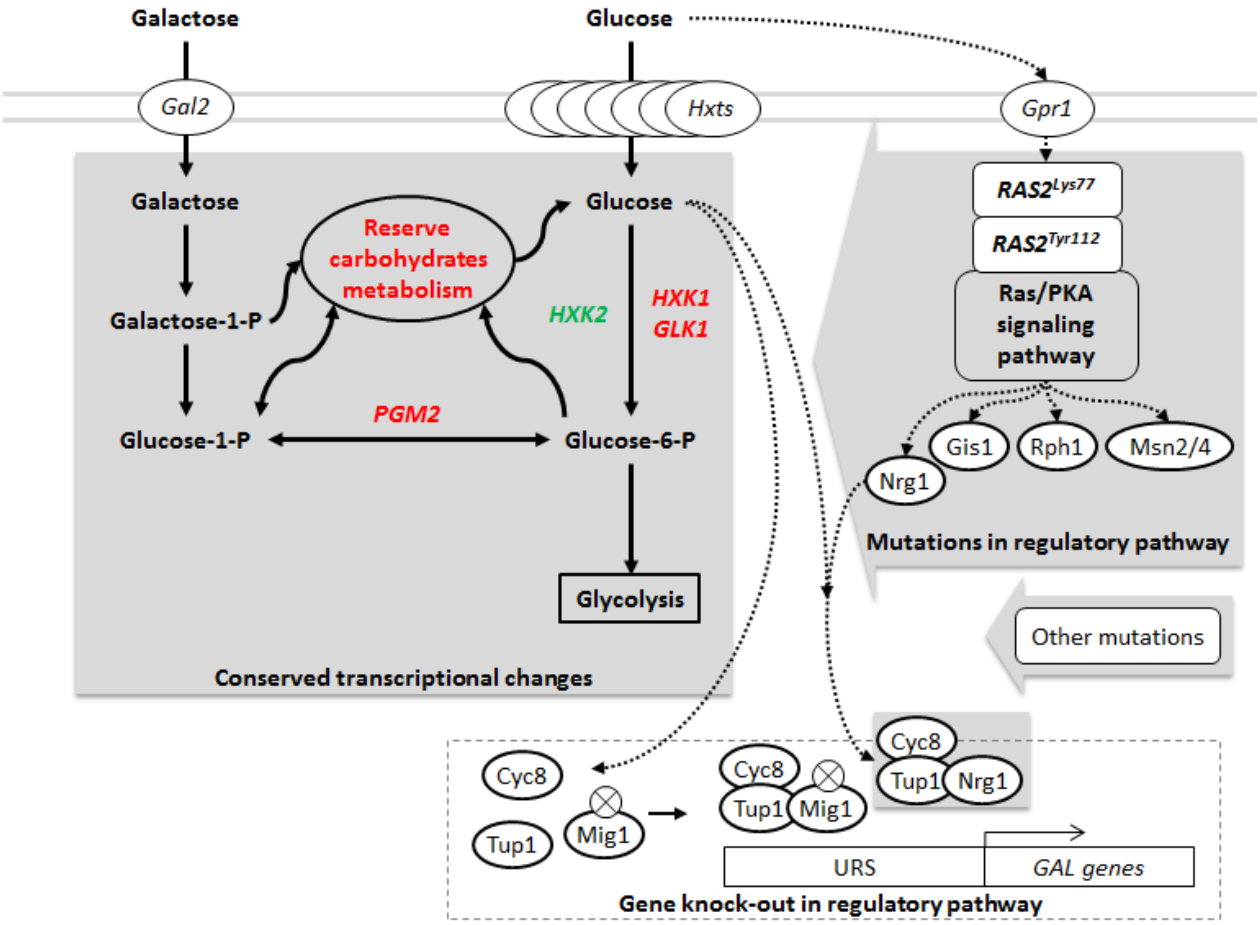


**Fig. 3-13.** Summary of possible molecular mechanism for the trade-off in galactose and glucose utilization. Colored letters (red and green) mean transcriptional change, up-regulation and down-regulation, respectively. Gray boxes (square and arrow shape) exhibit the changes in the evolved mutants. A dotted box means the change only in SO16 strain (knock-out of *MIG1*). Dot arrows represent signaling flow, solid arrows represent metabolic flow.

# 4. CONCLUSIONS

Adaptive evolution generated new strategies for improving galactose utilization in yeast *S.cerevisiae*. Those evolutionary strategies were characterized by integrated genome-scale analyses. Significantly, through this approach, evolutionary strategies of galactose-evolved mutants were elucidated at the molecular and genetic level. This characterization allowed inverse metabolic engineering to be more useful in evolutionary approaches for strain development. In addition, more characterization of the evolved mutants elucidated pleiotropy of the obtained traits. Through examples of this study, one could know what can be expectable or predictable in the application of evolutionary approaches for strain development. The most important findings in this Ph.D. study could be summarized as follow,

**Evolutionary changes of galactose-evolved yeast mutants can be characterized by integrated genome-scale analyses**

- Integration of genome scale analyses such as transcriptome, metabolome and whole-genome sequencing is crucial to identify the molecular and genetic basis of evolutionary changes. Each of these techniques does not allow for drawing comprehensive interpretation, but the combination of them provide a picture that enables understanding of the evolutionary strategies.
- It is important to use several evolved mutants with a reference strain, because this comparison allow identification of conserved mutations that result in the same phenotype. Each of the three evolved mutants has several mutations that probably do not contribute to the evolved phenotype, but by identifying conserved mutations, a clear picture emerged.

**Non-canonical genetic changes results in canonical molecular changes of the evolved mutants**

Transcriptome and metabolome analyses lists up significantly changed metabolisms; and among them, molecular changes likely related to galactose metabolism were found. Whole-genome sequencing identified several mutations, while there were no mutations in the genes or promoter regions that show the molecular changes. Key mutations for improving galactose utilization were found in non-canonical pathways.

- Common molecular changes in all the galactose-evolved mutants are 1) Up-regulation of *PGM2* with reduced concentration of galactose-1-phosphate and glucose-1-phosphate, 2) Up-regulation of reserve carbohydrates metabolism with increased concentration of trehalose and glycogen (Fig. 6-4).
- Adaptive evolution of yeast on galactose generates no mutations in the galactose pathway and its regulatory region, which had been considered as modification targets for metabolic engineering; also no mutations in reserve carbohydrate metabolism.
- A common pathway that contains mutations in all the evolved mutants was the Ras/PKA signaling pathway (Table 6-1).
- Two identified mutations in the *RAS2* gene result in improved galactose utilization with up-regulation of *PGM2* but not reserve carbohydrate metabolism (Fig. 6-6, Fig. 6-8).

**Importance of inverse metabolic engineering in connection with use of evolutionary approaches for strain development**

Few genetic and transcriptional changes are required to reach adaptive phenotypes. Accumulation of deleterious mutations or negative epistasis among beneficial mutations seems to be quite high during adaptive evolution. Therefore, inverse metabolic engineering can give a lot of new strategies for further engineering, such as sieving out beneficial mutations from negative ones and generation of new artificial combination of mutations.

- Site-directed mutants containing only one mutation in the *RAS2* gene, [$Gln^{77}$→Lys] or [$Asp^{112}$→Tyr] show similar improvement in the specific galactose uptake rate with the evolved mutants; also those strains display much smaller transcription changes compared to the evolved mutants (Fig. 6-6, Fig. 6-7).
- The site-directed mutant having *RAS2* [$Gln^{77}$→Lys] mutation even presents the highest specific galactose uptake rate among all the evolved and engineered strains (Fig. 6-6), and also relatively the highest maximum specific growth rate.
- Two mutations in the *RAS* gene have different effects on galactose utilization.
- New combinations of beneficial genetic changes almost completely recovers adaptive phenotypes in terms of galactose utilization, such as constitutive *PGM2* over-expression on a plasmid combined with mutation in *RAS2* [$Gln^{77}$→Lys] or [$Asp^{112}$→Tyr], and in *ERG5*

[Arg$^{370}$→Pro], respectively. These results indicate that new combinations of beneficial targets are one of the strategies for inverse metabolic engineering.

**Molecular and genetic bases of evolutionary pleiotropy: trade-offs in carbon utilization**

Galactose-evolved yeast mutants show trade-offs in carbon utilization between galactose and glucose. Adaptation on galactose seems to be realized by losing capacity for glucose utilization. The characterization results at the molecular and genetic level of this trade-off mechanism reveals that antagonistic pleiotropy is the dominant mechanism in the evolved mutations and this is likely realized by loosening the tight glucose catabolic repression system.

- The cost for improving galactose utilization may come from diminishing glucose utilization.
- Transcriptional changes with key metabolites of the three evolved mutants reveal antagonistic pleiotropy between glucose and galactose.
- Conserved molecular changes on both carbon sources are considered underlying the molecular mechanism by loosening tight glucose catabolic repression such as up-regulation of 1) *PGM2*, 2) non-glucose metabolism related hexokinase *HXK1*, *GLK1* and 3) reserve carbohydrate metabolism; down-regulation of 4) glucose catabolic repression regulator *HXK2*; and 4) involvement of transcription factors in nutrient sensing, *GIS1*, *RPH1*, *MSN2/4*, and *NRG1*.
- The mutations in the *RAS2* gene indicate antagonistic pleiotropy mechanisms for trade-off in carbon utilization by covering the phenotypic changes of the evolved mutants on both carbon sources.
- As mutations in the *RAS2* gene triggered up-regulation of *PGM2* and involved the transcription factors in nutrient sensing (Paper II), there are other unidentified mutations that induce transcriptional changes in the reserve carbohydrate metabolism and hexose kinases.
- Antagonistic pleiotropy between galactose and glucose utilization by attenuation of glucose regulation

# 5. PERSPECTIVE

Engineers have established significant development in the massive production of fuels and chemicals from petroleum and our generation is taking benefits from these technical advances. However, since the petroleum based production is using limited resources and generating serious environmental problems, our generation should prepare new technologies for the next generation, which uses renewable resources and alleviates environmental issues. Microbial fermentation processes could be one of the possible solutions, because this process utilizes biomass that is continuously produced with absorbing carbon dioxide in connection with its growth.

Engineering or reconstructing of microorganisms is the requisite step for the development of fermentation process. The engineering of biological systems is certainly different from mechanical or chemical engineering, since the biology is not only vastly complicated in their reaction networks and regulations, but also has emergent properties. Endy suggested four challenges in the engineering of biological system; 1) biological complexity, 2) the tedious and unreliable construction and characterization of synthetic biological systems, 3) the apparent spontaneous physical variation of biological system behavior, 4) evolution (Endy, 2005). One of the strategies for engineering the microorganism is to learn and apply nature's algorithm (Rothschild, 2010). Nature has produced relevant traits in specific environment; one also has used this valuable mechanism for making domesticated species from wild ones. Currently there are tools available for analysis genome-wide molecular and genetic changes. This means one can trace nature's strategies for obtaining new traits.

In this thesis, mutations in the *RAS2* gene were identified as the genetic bases for improving galactose utilization in yeast *S. cerevisiae*. This result indicates two important finding. Firstly, these mutations were only designable by nature's algorithm, random mutagenesis and natural preservation; because not only the relationship between these mutations and galactose utilization was not predictable, but also even though they were located in the same gene, the effects of each of them were vastly different. Therefore, there are still vast amounts of opportunity to find new strategies for strain development by evolutionary approaches. Recently, artificial mutagenesis methods have been developed, which are called genome engineering techniques that make it possible to generate random mutations on specific regions such as promoters, regulators and limited pathway genes (Santos & Stephanopoulos, 2008, Boyle & Gill, 2012). However, these

techniques still cannot cover the mutations that were found by random mutagenesis. Secondly, these mutations were only detectable by genome-scale analyses, since these analyses can only scan whole genome level changes. Of course there were still unidentified mutations that could be related to the changes in reserve carbohydrates metabolism and hexokinases. Whole-genome sequencing in this study had limitation such as incomplete coverage of whole DNA, insufficient coverage folds, missing copy number changes and rearrangement and so on. In spite of these limitations, whole genome sequencing detected the key mutations. It was also important point to focus on common changes by employing several parallel evolved mutants.

In industry, a lot of mutations are normally accumulated in producing strains, because of long history of evolution and high mutation rate by treating mutagen. The limitation in the number of evolved mutants could make it difficult to find common mutations generated in the same gene or pathways. Therefore, identification of beneficial genetic changes is practically very difficult. As shown in this study, there is a possible solution, namely to do more characterization of the mutants at different conditions with other omics tools such as transcriptome and metabolome analysis. Perturbation of culture conditions could separate conserved changes from others. And those conserved related mutations could be the main reason for the obtained phenotype. For example, the galactose-evolved mutants kept the changes of transcripts and metabolites in specific metabolism when growing on two different carbon sources. So those changes could be interpretive as induced by the same mutations. This process could reduce the number of mutations that is involved in desired traits.

Another point to consider is that engineering of the Ras/PKA signaling pathway might be an efficient way to achieve multiple phenotypes of industrial interest. Two mutations in the same gene showed different phenotypes, and just one mutation was enough to reach the entire adaptive phenotype. These results indicate that the effect of mutations in the *RAS2* gene could be beyond the change of activity of Ras/PKA signaling pathway. In addition, when some mutations in the *RAS2* gene were combined with *PGM2* over-expression, the more diverse phenotype could be expectable. Therefore, constructing a mutation library of the *RAS2* gene or another mutation library of the whole Ras/PKA signaling pathway with adjusting *PGM2* expression could be very useful for the next step in strain development.

It is important to keep in mind that there are several evolutionary mechanisms or genetic context that may be related to strain development, such as negative epistasis and trade-offs in traits. These mechanisms indicate that there are many chances to lose beneficial mutations and their combinations. Therefore, system-level characterization of evolutionary process could detect more number of beneficial mutations, and one can design new combinations of them or generate mutation library of the identified target gene. It is crucial to accumulate the examples of evolutionary mechanisms at detailed molecular levels supported by genome-scale analyses for further advancement of the use of evolutionary engineering in industrial biotechnology.

# 6. ACKNOWLEDGMENTS

It is a lucky chance to spend my doctoral time in Sweden, because of having experience of mature society. It is more luck to live in Göteborg, one of the best cities for taking care of family and lovely atmosphere. It is even greater to study in Systems and Synthetic Biology group at Chalmers. During this time, I have learned and disciplined not only to step in an independent researcher in science and engineering but also to build my view of world and personality.

First of all, I am sincere grateful to my supervisor Prof. Jens Nielsen, for giving an opportunity to be his PhD student, providing profound freedom in my idea development with encouragement, trust and patience. He also has supported a lot of resources in research and education with guiding how my work fit into the big picture of biotechnology. His excellent ability and professional attitude in science are my next goal that I would like to achieve in my life.

I would like to thank Prof. Christer Larsson, Assoc. Prof. Joakim Norbeck, Assist. Prof. Dina Petranovic for many productive discussion and good feedback; also, Assoc. Prof. Carl Johan Franzén, and Assist. Prof. Maurizio Bettiga for good teaching and discussion.

I would like to thank co-authors, Assist. Prof. Goutham Vemuri for his supervision and inspiration in research, Assist Prof. Wanwipa Vongsangnak, Assist. Prof. Jin Hou, Assist. Prof. Sergio Bordel Velasco and Saeed Shoaie for productive collaboration.

In particular, I would like to thank seniors at CJ CheilJedang Dr. Young-Lyeol Yang, Dr. Chang-Gyeom Kim, Sung-Sik Park, Jin-Su Chung for helping to start my PhD; and Seong-Ah Kim, Jaeyong Park, Dr. Jaeyeong Ju, Kyung-Suk Lee, Director of R&D Sangjo Lim, and CEO Chul-Ha Kim for their practical supporting and helping; my leaders, Vice Director of R&D Jinman Cho, Dr. Jong-Kwon Han, Dr. Kwang-Myung Cho, Dr. Kwang-Ho Lee, Assist Prof. Hyun-Soo Kim for their advises and valuable guidance; and Byoung-chun Lee, Jun-Ok Moon for their encouragement.

A special acknowledgement for Dr. Rahul Kumar, he hasn't hesitated to support me with his vast amount of time, effort and attention to my research and writing. He always gives very valuable input and discussion in diverse field, which allow me to have broad knowledge.

I would like to thank my colleagues, Dr. Jie Zhang, Dr. Pramote Chumnanpuen for their helping to initiate experimental systems biology works and cheering; Dr. Luis Caspeta-Guadarrama, Dr. Yun Chen, Klaas Buijs for revising the thesis and good discussion; Kanokarn Kocharin, Zihe Liu, Lifang Liu, Marta Papini, Siavash Partow, Gionata Scalcinati, Tobias Österlund, Juan Valle, Fredrik Karlsson, Christoph Knuf, Bouke de Jong, Dr. Roberto Olivares-Hernández, Dr. Shuobo Shi, Dr. José L. Martínez, Dr. Adil Mardinoglu, Dr. António Roldão Dr. Sakda Khoomrung and other lab mates including industrial biotechnology group members for

making a lab life fun and checking manuscripts. Thank Dr. Verena Siewers, Dr. Intawat Nookaew, Dr. Subir Kumar Nandy, Assist. Prof. Keith Tyo for their valuable comments. Thank Govindprasad Bhutada and Sinisa Bratulic for their questioning and cheers. I would like to deeply thank Erica Dahlin, Martina Butorac, Malin Nordvall, Marie Nordqvist, Ximena Sevilla, Jenny Nilsson, Suwanee Jansa-Ard for their supporting.

There are also very nice Korean friends and seniors who I met in Sweden, Dr. Hee Chul Park, Sang-Mi Na, Hyeongcheol Yoo, Dr. Byung-Soo Kang, Youngwoo Nam, Eunmi Choi, Dr. Keunjae Kim, Myung-Ja Kim, Kwang-Sup Lee, Youngsoon Um and Ingemar Johansson. Without their helps, my family would face many difficulties; thank so much for their support.

I would like to thank Prof. Jin Byung Park for guiding to PhD and close friends Taehee Park, Hyungun Kim for their encouragement and keeping good discussion in many issues.

My sincere gratitude goes to my parents, Sung-Hwan Hong, Kyung-Ja Lee, my younger brother Duk-Ki Hong, parents-in-law Jae-Seok Kang, Sung-Won Hong, and sister-in-law Min-Sun Kang. Their devotion, pray for me and enthusiasm in their life are great resource for sustaining and keeping values of my life.

And most of all, I would like to thank my wife Min-Jin Kang, for exploring unpredictable life together, with being beautiful, lively, wise and encouraging me to overcome any barriers in the doctoral works. It is great to see and learn how to adapt new environment from my daughter Jin-Seo Hong; she gets well along with her international friends, and I am busy to drop her on and off at her friends' home. Jin-Ha Hong, my son doesn't care about his dad's situation, just always try to hug me and sit on my knees. Thank my angels for your presence.

Kuk-Ki Hong

Aug 2012

Göteborg, Sweden

# 7. REFERENCES


Acar M, Becskei A & van Oudenaarden A (2005) Enhancement of cellular memory by reducing stochastic transitions. *Nature* 435: 228-232.

Albertsen L, Chen Y, Bach LS*, et al.* (2011) Diversion of flux toward sesquiterpene production in *Saccharomyces cerevisiae* by fusion of host and heterologous enzymes. *Appl Environ Microbiol* 77: 1033-1040.

Asadollahi MA, Maury J, Schalk M, Clark A & Nielsen J (2010) Enhancement of farnesyl diphosphate pool as direct precursor of sesquiterpenes through metabolic engineering of the mevalonate pathway in *Saccharomyces cerevisiae*. *Biotechnol Bioeng* 106: 86-96.

Bailey JE, Sburlati A, Hatzimanikatis V, Lee K, Renner WA & Tsai PS (2002) Inverse metabolic engineering: a strategy for directed genetic engineering of useful phenotypes. *Biotechnol Bioeng* 79: 568-579.

Becker JV, Armstrong GO, van der Merwe MJ, Lambrechts MG, Vivier MA & Pretorius IS (2003) Metabolic engineering of *Saccharomyces cerevisiae* for the synthesis of the wine-related antioxidant resveratrol. *FEMS Yeast Res* 4: 79-85.

Bengtsson O, Jeppsson M, Sonderegger M, Parachin NS, Sauer U, Hahn-Hagerdal B & Gorwa-Grauslund MF (2008) Identification of common traits in improved xylose-growing *Saccharomyces cerevisiae* for inverse metabolic engineering. *Yeast* 25: 835-847.

Boles E & Hollenberg CP (1997) The molecular genetics of hexose transport in yeasts. *FEMS Microbiol Rev* 21: 85-111.

Boyle NR & Gill RT (2012) Tools for genome-wide strain design and construction. *Curr Opin Biotechnol*.

Bro C & Nielsen J (2004) Impact of 'ome' analyses on inverse metabolic engineering. *Metab Eng* 6: 204-211.

Bro C, Knudsen S, Regenberg B, Olsson L & Nielsen J (2005) Improvement of galactose uptake in *Saccharomyces cerevisiae* through overexpression of phosphoglucomutase: example of transcript analysis as a tool in inverse metabolic engineering. *Appl Environ Microbiol* 71: 6465-6472.

Brochado AR, Matos C, Moller BL, Hansen J, Mortensen UH & Patil KR (2010) Improved vanillin production in baker's yeast through in silico design. *Microb Cell Fact* 9: 84.

Cabib E (1970) Research on sugar nucleotides brings honor to Argentinian biochemist (Luis Leloir). *Science* 170: 608-609.

Chen X, Nielsen KF, Borodina I, Kielland-Brandt MC & Karhumaa K (2011) Increased isobutanol production in *Saccharomyces cerevisiae* by overexpression of genes in valine metabolism. *Biotechnol Biofuels* 4: 21.

Chigira Y, Oka T, Okajima T & Jigami Y (2008) Engineering of a mammalian O-glycosylation pathway in the yeast *Saccharomyces cerevisiae*: production of O-fucosylated epidermal growth factor domains. *Glycobiology* 18: 303-314.

Conrad TM, Lewis NE & Palsson BO (2011) Microbial laboratory evolution in the era of genome-scale science. *Mol Syst Biol* 7: 509.

Cooper VS & Lenski RE (2000) The population genetics of ecological specialization in evolving *Escherichia coli* populations. *Nature* 407: 736-739.

Daugherty JP, Kraemer WF & Joshi JG (1975) Purification and properties of phosphoglucomutase from Fleischmann's yeast. *Eur J Biochem* 57: 115-126.

Davie JK, Edmondson DG, Coco CB & Dent SY (2003) Tup1-Ssn6 interacts with multiple class I histone deacetylases in vivo. *J Biol Chem* 278: 50158-50162.

de Jongh WA, Bro C, Ostergaard S, Regenberg B, Olsson L & Nielsen J (2008) The roles of galactitol, galactose-1-phosphate, and phosphoglucomutase in galactose-induced toxicity in *Saccharomyces cerevisiae*. *Biotechnol Bioeng* 101: 317-326.

Demain AL (2000) Microbial biotechnology. *Trends Biotechnol* 18: 26-31.

Dettman JR, Rodrigue N, Melnyk AH, Wong A, Bailey SF & Kassen R (2012) Evolutionary insight

from whole-genome sequencing of experimentally evolved microbes. *Mol Ecol* 21: 2058-2077.

Egel-Mitani M, Andersen AS, Diers II, Hach M, Thim L, Hastrup S & Vad K (2000) Yield improvement of heterologous peptides expressed in yps1-disrupted *Saccharomyces cerevisiae* strains. *Enzyme Microb Technol* 26: 671-677.

Elena SF & Lenski RE (2003) Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat Rev Genet* 4: 457-469.

Elkins JG, Raman B & Keller M (2010) Engineered microbial systems for enhanced conversion of lignocellulosic biomass. *Curr Opin Biotechnol* 21: 657-662.

Endy D (2005) Foundations for engineering biology. *Nature* 438: 449-453.

Estruch F (2000) Stress-controlled transcription factors, stress-induced genes and stress tolerance in budding yeast. *FEMS Microbiol Rev* 24: 469-486.

Eudes A, Baidoo E, Yang F*, et al.* (2011) Production of tranilast [*N*-(3',4'-dimethoxycinnamoyl)-anthranilic acid] and its analogs in yeast *Saccharomyces cerevisiae*. *Appl Microbiol Biot* 89: 989-1000.

Fukasawa T, Obonai K, Segawa T & Nogi Y (1980) The enzymes of the galactose cluster in *Saccharomyces cerevisiae*. II. Purification and characterization of uridine diphosphoglucose 4-epimerase. *J Biol Chem* 255: 2705-2707.

Gancedo JM (1998) Yeast carbon catabolite repression. *Microbiology and Molecular Biology Reviews* 62: 334-+.

Garcia Sanchez R, Hahn-Hagerdal B & Gorwa-Grauslund MF (2010) PGM2 overexpression improves anaerobic galactose fermentation in *Saccharomyces cerevisiae*. *Microb Cell Fact* 9: 40.

Goeddel DV, Heyneker HL, Hozumi T*, et al.* (1979) Direct expression in *Escherichia coli* of a DNA sequence coding for human growth hormone. *Nature* 281: 544-548.

Guadalupe Medina V, Almering MJ, van Maris AJ & Pronk JT (2010) Elimination of glycerol production in anaerobic cultures of a *Saccharomyces cerevisiae* strain engineered to use acetic acid as an electron acceptor. *Appl Environ Microbiol* 76: 190-195.

Hackel BJ, Huang D, Bubolz JC, Wang XX & Shusta EV (2006) Production of soluble and active transferrin receptor-targeting single-chain antibody using *Saccharomyces cerevisiae*. *Pharm Res* 23: 790-797.

Hahn-Hagerdal B, Karhumaa K, Fonseca C, Spencer-Martins I & Gorwa-Grauslund MF (2007) Towards industrial pentose-fermenting yeast strains. *Appl Microbiol Biotechnol* 74: 937-953.

Hazelwood LA, Walsh MC, Luttik MA, Daran-Lapujade P, Pronk JT & Daran JM (2009) Identity of the growth-limiting nutrient strongly affects storage carbohydrate accumulation in anaerobic chemostat cultures of *Saccharomyces cerevisiae*. *Appl Environ Microbiol* 75: 6876-6885.

Heinemann M & Sauer U (2010) Systems biology of microbial metabolism. *Curr Opin Microbiol* 13: 337-343.

Herrgard MJ, Swainston N, Dobson P*, et al.* (2008) A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nat Biotechnol* 26: 1155-1160.

Herring CD, Raghunathan A, Honisch C*, et al.* (2006) Comparative genome sequencing of Escherichia coli allows observation of bacterial evolution on a laboratory timescale. *Nat Genet* 38: 1406-1412.

Holden HM, Rayment I & Thoden JB (2003) Structure and function of enzymes of the Leloir pathway for galactose metabolism. *Journal of Biological Chemistry* 278: 43885-43888.

Ideker T, Thorsson V, Ranish JA*, et al.* (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* 292: 929-934.

Ikeda M, Ohnishi J, Hayashi M & Mitsuhashi S (2006) A genome-based approach to create a minimally mutated *Corynebacterium glutamicum* strain for efficient L-lysine production. *J Ind Microbiol Biotechnol* 33: 610-615.

Ikeda M, Mitsuhashi S, Tanaka K & Hayashi M (2009) Reengineering of a *Corynebacterium glutamicum* L-arginine and L-citrulline producer. *Appl Environ Microbiol* 75: 1635-1641.

Jelen P (1979) Industrial Whey Processing Technology - Overview. *Journal of Agricultural and Food Chemistry* 27: 658-661.

Johnson IS (1983) Human insulin from recombinant DNA technology. *Science* 219: 632-637.

Kelly SL, Lamb DC, Corran AJ, Baldwin BC, Parks LW & Kelly DE (1995) Purification and reconstitution of activity of *Saccharomyces cerevisiae* P450 61, a sterol Delta(22)-desaturase. *FEBS Lett* 377: 217-220.

Kim E-J, Park Y-K, Lim H-K, Park Y-C & Seo J-H (2009) Expression of hepatitis B surface antigen S domain in recombinant *Saccharomyces cerevisiae* using GAL1 promoter. *J Biotech* 141: 155-159.

Kim HJ, Lee SJ & Kim H-J (2010) Optimizing the secondary structure of human papillomavirus type 16 L1 mRNA enhances L1 protein expression in *Saccharomyces cerevisiae*. *J Biotech* 150: 31-36.

Kim NJ, Li H, Jung K, Chang HN & Lee PC (2011) Ethanol production from marine algal hydrolysates using *Escherichia coli* KO11. *Bioresour Technol* 102: 7466-7469.

Kim SR, Ha SJ, Wei N, Oh EJ & Jin YS (2012) Simultaneous co-fermentation of mixed sugars: a promising strategy for producing cellulosic ethanol. *Trends Biotechnol* 30: 274-282.

Kinoshita S, Udaka S & Shimono M (1957) Studies on the amino acid fermentation. Part I. Production of L-glutamic acid by various microorganisms. . *J. Gen. Appl. Microbiol.* 3: 193-205.

Krivoruchko A, Siewers V & Nielsen J (2011) Opportunities for yeast metabolic engineering: Lessons from synthetic biology. *Biotechnol J* 6: 262-276.

Kummel A, Ewald JC, Fendt SM*, et al.* (2010) Differential glucose repression in common yeast strains in response to HXK2 deletion. *FEMS Yeast Res* 10: 322-332.

Lai K, Elsas LJ & Wierenga KJ (2009) Galactose toxicity in animals. *IUBMB Life* 61: 1063-1074.

Le Crom S, Schackwitz W, Pennacchio L*, et al.* (2009) Tracking the roots of cellulase hyperproduction by the fungus Trichoderma reesei using massively parallel DNA sequencing. *Proc Natl Acad Sci U S A* 106: 16151-16156.

Lee KS, Hong ME, Jung SC*, et al.* (2011) Improved galactose fermentation of *Saccharomyces cerevisiae* through inverse metabolic engineering. *Biotechnol Bioeng* 108: 621-631.

Lee W & Dasilva NA (2006) Application of sequential integration for metabolic engineering of 1,2-propanediol production in yeast. *Metab Eng* 8: 58-65.

Li A, Liu Z, Li Q, Yu L, Wang D & Deng X (2008) Construction and characterization of bidirectional expression vectors in *Saccharomyces cerevisiae*. *FEMS Yeast Res* 8: 6-9.

Lipinsky ES (1981) Chemicals from biomass: petrochemical substitution options. *Science* 212: 1465-1471.

Lowin T, Raab U, Schroeder J, Franssila R & Modrow S (2005) Parvovirus B19 VP2-proteins produced in *Saccharomyces cerevisiae*: comparison with VP2-particles produced by baculovirus-derived vectors. *J Vet Med B Infect Dis Vet Public Health* 52: 348-352.

MacLean D, Jones JD & Studholme DJ (2009) Application of 'next-generation' sequencing technologies to microbial genetics. *Nat Rev Microbiol* 7: 287-296.

Madsen KM, Udatha GD, Semba S*, et al.* (2011) Linking genotype and phenotype of *Saccharomyces cerevisiae* strains reveals metabolic engineering targets and leads to triterpene hyper-producers. *PLoS ONE* 6: e14763.

Mapelli V, Hillestrøm PR, Kápolna E, Larsen EH & Olsson L (2011) Metabolic and bioprocess engineering for production of selenized yeast with increased content of seleno-methylselenocysteine. *Metab Eng* 13: 282-293.

Mardis ER (2008) Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet* 9: 387-402.

Marie G-G, Bärbel H-H & Maurizio B (2009) Metabolic Engineering in Yeast. *The Metabolic Pathway Engineering Handbook,*ed.^eds.), p.^pp. 22-21-22-48. CRC Press.

Metzker ML (2010) Sequencing technologies - the next generation. *Nat Rev Genet* 11: 31-46.

Mutka SC, Bondi SM, Carney JR, Da Silva NA & Kealey JT (2006) Metabolic pathway engineering for complex polyketide biosynthesis in *Saccharomyces cerevisiae*. *FEMS Yeast Res* 6: 40-47.

Nakayama K, Kitada S & Kinoshita S (1961) Studies on lysine fermentation I. The control mechanism on lysine accumulation by homoserine and threonine. *J. Gen. Appl. Microbiol.* 7: 145-154.

Nevoigt E (2008) Progress in metabolic engineering of *Saccharomyces cerevisiae*. *Microbiol Mol Biol Rev* 72: 379-412.

Nielsen J & Jewett MC (2008) Impact of systems biology on metabolic engineering of *Saccharomyces cerevisiae*. *FEMS Yeast Res* 8: 122-131.

Ohnishi J, Mitsuhashi S, Hayashi M, Ando S, Yokoi H, Ochiai K & Ikeda M (2002) A novel methodology employing *Corynebacterium glutamicum* genome information to generate a new L-lysine-producing mutant. *Appl Microbiol Biotechnol* 58: 217-223.

Orzechowski Westholm J, Tronnersjo S, Nordberg N, Olsson I, Komorowski J & Ronne H (2012) Gis1 and Rph1 regulate glycerol and acetate metabolism in glucose depleted yeast cells. *PLoS One* 7: e31577.

Ostergaard S, Olsson L, Johnston M & Nielsen J (2000) Increasing galactose consumption by *Saccharomyces cerevisiae* through metabolic engineering of the GAL gene regulatory network. *Nat Biotechnol* 18: 1283-1286.

Oud B, van Maris AJ, Daran JM & Pronk JT (2012) Genome-wide analytical approaches for reverse metabolic engineering of industrially relevant phenotypes in yeast. *FEMS Yeast Res* 12: 183-196.

Ozcan S & Johnston M (1999) Function and regulation of yeast hexose transporters. *Microbiol Mol Biol Rev* 63: 554-569.

Panesar PS, Kennedy JF, Gandhi DN & Bunko K (2007) Bioutilisation of whey for lactic acid production. *Food Chemistry* 105: 1-14.

Pannala VR, Bhat PJ, Bhartiya S & Venkatesh KV (2010) Systems biology of GAL regulon in *Saccharomyces cerevisiae*. *Wiley Interdiscip Rev Syst Biol Med* 2: 98-106.

Patil KR & Nielsen J (2005) Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc Natl Acad Sci U S A* 102: 2685-2689.

Patnaik R (2008) Engineering complex phenotypes in industrial strains. *Biotechnol Prog* 24: 38-47.

Peralta-Yahya PP, Ouellet M, Chan R, Mukhopadhyay A, Keasling JD & Lee TS (2011) Identification and microbial production of a terpene-based advanced biofuel. *Nat Commun* 2: 483.

Petranovic D & Vemuri GN (2009) Impact of yeast systems biology on industrial biotechnology. *J Biotechnol* 144: 204-211.

Portnoy VA, Bezdan D & Zengler K (2011) Adaptive laboratory evolution--harnessing the power of biology for metabolic engineering. *Curr Opin Biotechnol* 22: 590-594.

Pronk JT (2002) Auxotrophic yeast strains in fundamental and applied research. *Appl Environ Microbiol* 68: 2095-2100.

Raab AM, Gebhardt G, Bolotina N, Weuster-Botz D & Lang C (2010) Metabolic engineering of *Saccharomyces cerevisiae* for the biotechnological production of succinic acid. *Metab Eng* 12: 518-525.

Rakestraw JA, Sazinsky SL, Piatesi A, Antipov E & Wittrup KD (2009) Directed evolution of a secretory leader for the improved expression of heterologous proteins and full-length antibodies in *Saccharomyces cerevisiae*. *Biotechnol Bioeng* 103: 1192-1201.

Ramsey SA, Smith JJ, Orrell D*, et al.* (2006) Dual feedback loops in the GAL regulon suppress cellular heterogeneity in yeast. *Nat Genet* 38: 1082-1087.

Reifenberger E, Boles E & Ciriacy M (1997) Kinetic characterization of individual hexose

transporters of *Saccharomyces cerevisiae* and their relation to the triggering mechanisms of glucose repression. *Eur J Biochem* 245: 324-333.

Reimand J, Arak T & Vilo J (2011) g:Profiler--a web server for functional interpretation of gene lists (2011 update). *Nucleic Acids Res* 39: W307-315.

Rico J, Pardo E & Orejas M (2010) Enhanced production of a plant monoterpene by overexpression of the 3-hydroxy-3-methylglutaryl coenzyme A reductase catalytic domain in *Saccharomyces cerevisiae*. *Appl Environ Microbiol* 76: 6449-6454.

Rothschild LJ (2010) A powerful toolkit for synthetic biology: Over 3.8 billion years of evolution. *Bioessays* 32: 304-313.

Santos CN & Stephanopoulos G (2008) Combinatorial engineering of microbes for optimizing cellular phenotype. *Curr Opin Chem Biol* 12: 168-176.

Sauer U (2001) Evolutionary engineering of industrially important microbial phenotypes. *Adv Biochem Eng Biotechnol* 73: 129-169.

Schell MA & Wilson DB (1977) Purification and properties of galactokinase from *Saccharomyces cerevisiae*. *J Biol Chem* 252: 1162-1166.

Segawa T & Fukasawa T (1979) The enzymes of the galactose cluster in *Saccharomyces cerevisiae*. Purification and characterization of galactose-1-phosphate uridylyltransferase. *J Biol Chem* 254: 10707-10709.

Shendure J & Ji H (2008) Next-generation DNA sequencing. *Nat Biotechnol* 26: 1135-1145.

Siewers V, San-Bento R & Nielsen J (2010) Implementation of communication-mediating domains for non-ribosomal peptide production in *Saccharomyces cerevisiae*. *Biotechnol Bioeng* 106: 841-844.

Siso MIG (1996) The biotechnological utilization of cheese whey: A review. *Bioresour Technol* 57: 1-11.

Snyder M & Gallagher JE (2009) Systems biology from a yeast omics perspective. *FEBS Lett* 583: 3895-3899.

Sonderegger M & Sauer U (2003) Evolutionary engineering of *Saccharomyces cerevisiae* for anaerobic growth on xylose. *Appl Environ Microbiol* 69: 1990-1998.

St John TP & Davis RW (1981) The organization and transcription of the galactose gene cluster of Saccharomyces. *J Mol Biol* 152: 285-315.

Tavares S, Grotkjaer T, Obsen T, Haslam RP, Napier JA & Gunnarsson N (2011) Metabolic engineering of *Saccharomyces cerevisiae* for production of Eicosapentaenoic Acid, using a novel {Delta}5-Desaturase from *Paramecium tetraurelia*. *Appl Environ Microbiol* 77: 1854-1861.

Thoden JB & Holden HM (2005) The molecular architecture of galactose mutarotase/UDP-galactose 4-epimerase from *Saccharomyces cerevisiae*. *J Biol Chem* 280: 21900-21907.

Timson DJ (2007) Galatose Metabolism in *Saccharomyces cerevisiae*. *Dynamic Biochemistry, Process Biotechnology and Molecular Biology* 1: 63-73.

Treitel MA & Carlson M (1995) Repression by SSN6-TUP1 is directed by MIG1, a repressor/activator protein. *Proc Natl Acad Sci USA* 92: 3132-3136.

Tyo KE, Alper HS & Stephanopoulos GN (2007) Expanding the metabolic engineering toolbox: more options to engineer cells. *Trends Biotechnol* 25: 132-137.

Vai M, Brambilla L, Orlandi I, Rota N, Ranzi BM, Alberghina L & Porro D (2000) Improved secretion of native human insulin-like growth factor 1 from gas1 mutant *Saccharomyces cerevisiae* cells. *Appl Environ Microbiol* 66: 5477-5479.

van Maris AJ, Geertman JM, Vermeulen A*, et al.* (2004) Directed evolution of pyruvate decarboxylase-negative *Saccharomyces cerevisiae*, yielding a C2-independent, glucose-tolerant, and pyruvate-hyperproducing yeast. *Appl Environ Microbiol* 70: 159-166.

van Maris AJ, Abbott DA, Bellissimi E*, et al.* (2006) Alcoholic fermentation of carbon sources in biomass hydrolysates by *Saccharomyces cerevisiae*: current status. *Antonie Van Leeuwenhoek* 90: 391-418.

Vellanki RN, Komaravelli N, Tatineni R & Mangamoori LN (2007) Expression of hepatitis B surface antigen in *Saccharomyces cerevisiae* utilizing glyceraldeyhyde-3-phosphate dehydrogenase promoter of Pichia pastoris. *Biotechnol Lett* 29: 313-318.

Vennestrom PN, Osmundsen CM, Christensen CH & Taarning E (2011) Beyond petrochemicals: the renewable chemicals industry. *Angew Chem Int Ed Engl* 50: 10502-10509.

Verwaal R, Wang J, Meijnen JP, Visser H, Sandmann G, van den Berg JA & van Ooyen AJ (2007) High-level production of beta-carotene in *Saccharomyces cerevisiae* by successive transformation with carotenogenic genes from *Xanthophyllomyces dendrorhous*. *Appl Environ Microbiol* 73: 4342-4350.

Vijayendran C, Barsch A, Friehs K, Niehaus K, Becker A & Flaschel E (2008) Perceiving molecular evolution processes in *Escherichia coli* by comprehensive metabolite and gene expression profiling. *Genome Biol* 9: R72.

Wargacki AJ, Leonard E, Win MN*, et al.* (2012) An engineered microbial platform for direct biofuel production from brown macroalgae. *Science* 335: 308-313.

Warner JR, Patnaik R & Gill RT (2009) Genomics enabled approaches in strain engineering. *Curr Opin Microbiol* 12: 223-230.

Wenger JW, Piotrowski J, Nagarajan S, Chiotti K, Sherlock G & Rosenzweig F (2011) Hunger Artists: Yeast Adapted to Carbon Limitation Show Trade-Offs under Carbon Sufficiency. *Plos Genetics* 7.

Werpy T & Petersen G (2004) Top value added chemicals from biomass. Volume 1, Results of screening for potential candidates from sugars and synthesis gas. U.S. Dept. of Energy, [Washington, D.C.].

Westfall PJ, Pitera DJ, Lenihan JR*, et al.* (2012) Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. *Proc Natl Acad Sci USA* 109: E111-118.

Wi SG, Kim HJ, Mahadevan SA, Yang DJ & Bae HJ (2009) The potential value of the seaweed Ceylon moss (*Gelidium amansii*) as an alternative bioenergy resource. *Bioresour Technol* 100: 6658-6660.

Wieczorke R, Krampe S, Weierstall T, Freidel K, Hollenberg CP & Boles E (1999) Concurrent knock-out of at least 20 transporter genes is required to block uptake of hexoses in *Saccharomyces cerevisiae*. *FEBS Lett* 464: 123-128.

Yano K & Fukasawa T (1997) Galactose-dependent reversible interaction of Gal3p with Gal80p in the induction pathway of Gal4p-activated genes of *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* 94: 1721-1726.

Yu KO, Jung J, Kim SW, Park CH & Han SO (2012) Synthesis of FAEEs from glycerol in engineered *Saccharomyces cerevisiae* using endogenously produced ethanol by heterologous expression of an unspecific bacterial acyltransferase. *Biotechnol Bioeng* 109: 110-115.

Zaldivar J, Borges A, Johansson B, Smits HP, Villas-Boas SG, Nielsen J & Olsson L (2002) Fermentation performance and intracellular metabolite patterns in laboratory and industrial xylose-fermenting *Saccharomyces cerevisiae*. *Appl Microbiol Biotechnol* 59: 436-442.

Zhang B, Carlson R & Srienc F (2006) Engineering the monomer composition of polyhydroxyalkanoates synthesized in *Saccharomyces cerevisiae*. *Appl Environ Microbiol* 72: 536-543.

Zhang YX, Perry K, Vinci VA, Powell K, Stemmer WP & del Cardayre SB (2002) Genome shuffling leads to rapid phenotypic improvement in bacteria. *Nature* 415: 644-646.

Zhao L, Wang J, Zhou J, Liu L, Du G & Chen J (2011) Modification of carbon flux in *Sacchromyces cerevisiae* to improve L-lactic acid production. *Wei Sheng Wu Xue Bao* 51: 50-58.

Zheng WJ, Xu HE & Johnston SA (1997) The cysteine-peptidase bleomycin hydrolase is a member of the galactose regulon in yeast. *Journal of Biological Chemistry* 272: 30350-30355.