ห้องสมุดงานวิจัย สำนักงานคณะกรรมการวิจัยแห่งชาติ
E46912

# APPLICATION OF *LACTOBACILLUS* GENOME-SCALE MODEL TO AUXOTROPH DEVELOPMENT

MISS PORNTIP CHIEWCHANKASET

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE
(BIOINFORMATICS AND SYSTEMS BIOLOGY)
SCHOOL OF BIORESOURCES AND TECHNOLOGY AND
SCHOOL OF INFORMATION TECHNOLOGY
KING MONGKUT'S UNIVERSITY OF TECHNOLOGY THONBURI
2010

hec 246525

ห้องสมุดงานวิจัย สำนักงานคณะกรรมการวิจัยแห่งชาติ
E46912

# Application of *Lactobacillus* Genome-Scale Model to Auxotroph Development

Miss Porntip Chiewchankaset B.Sc. (Microbiology)

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science (Bioinformatics and Systems Biology)
School of Bioresources and Technology and School of Information Technology
King Mongkut's University of Technology Thonburi
2010

THE NATIONAL RESEARCH COUNCIL OF THAILAND RESEARCH LIBRARY

Thesis Committee

| | |
|---|---|
| .................................................... (Researcher, Vethachai Plengvidhya, Ph.D.) | Chairman of Thesis Committee |
| .................................................... (Asst. Prof. Asawin Meechai, Ph. D.) | Member and Thesis Advisor |
| .................................................... (Lecturer, Taweerat Vichitsoonthonkul, Ph.D.) | Member |
| .................................................... (Researcher, Peter Kurdi, Ph.D.) | Member |

Copyright reserved

# PREFACE

This thesis was written for accomplishment of my master degree of Bioinformatics and Systems biology. The topic of the study is "Application of *Lactobacillus* Genome-Scale Model to Auxotroph Development" or Thai title is "การประยุกต์ใช้แบบจำลองระดับจีโนมเพื่อพัฒนาแลคโตแบซิลลัสที่มีคุณสมบัติเป็นออกโซโทรพ". This work was done at King Mongkut's University of Technology Thonburi (KMUTT), Thailand. This thesis is consists of five main chapters, which are introduction, backgrounds and literature reviews, methodology, results and discussions, and conclusions.

| | |
|---|---|
| Thesis Title | Application of *Lactobacillus* Genome-Scale Model to Auxotroph Development |
| Thesis Credits | 12 |
| Candidate | Miss Porntip Chiewchankaset |
| Thesis Advisor | Asst. Prof. Dr. Asawin Meechai |
| Program | Master of Science |
| Field of Study | Bioinformatics and Systems Biology |
| Department | Bioinformatics and Systems Biology |
| Faculty | School of Bioresources and Technology and School of Information Technology |
| B.E. | 2553 |

Abstract

E46912

At present, Nham, a Thai fermented sausage product, is manufactured using commercial starter bacteria. Yet, these bacterial starters can readily be sub-cultured by Nham producers themselves, making it economically infeasible for starter providers to run this business. To address this issue, auxotrophic starters should be developed to be used in Nham production process. Although fast and simple, the traditional approach for auxotroph development using replica plating techniques is an indirected approach in which researchers have no clue as to what genes and what mechanisms that lead to cell inability to synthesize certain substances required for its growth. Another possible way to develop auxotroph strains is to use molecular biology techniques. This approach is quite powerful as it may enable us to develop genetically designed Nham starters which are difficult for Nham producers to reproduce, and thus they must continuously acquire the starters from Nham venders only. An important step of strain development using the directed approach is the step of designing of what gene(s) to be engineered. In this study, we apply a published genome-scale model of *Lactobacillus plantarum* WCFS 1 to help predict a set of genes/metabolites possible for auxotroph development as Nham starters. The published model consisting of 724 genes, 766 reactions, and 660 metabolites is implemented on MATLAB using the COBRA Toolbox version 1.3.3. This genome scale model is simulated by setting the objective function as maximize biomass and the nutrient uptake and secretion rates as the constraints. It is employed using single reaction and single gene knockout techniques to identify sets of metabolite-

E46912

gene pairs essential for growth of cells grown on 4 media, i.e., Chemically defined medium (CDM), de Man, Rogosa, and Sharpe (MRS), Simulated Nham Broth (SNB), and Actual Nham (AN). It is found that there are 4 pairs of genes and metabolite candidates, i.e. *glmM* with D-glucosamine 1-phosphate, *murI* with D-glutamate, *alr* with D-alanine, and *dfrA* with 5,6,7,8-tetrahydrofolate for all media, and only one pair of gene and metabolite candidate (*gadB* with 4-aminobutanoate) on SNB and AN. These selected gene and metabolites are potential candidates for *L. plantarum* WCFS 1 auxotroph development and improvement of Nham starter culture in the future.

Keywords: Genome-scale metabolic model / *Lactobacillus plantarum* WCFS 1 / Nham / Auxotroph / Gene and metabolite candidate

| | |
|---|---|
| หัวข้อวิทยานิพนธ์ | การประยุกต์ใช้แบบจำลองระดับจีโนมเพื่อพัฒนาแลคโตแบซิลลัสที่มีคุณสมบัติเป็นออกโซโทรพ |
| หน่วยกิต | 12 |
| ผู้เขียน | นางสาวพรทิพย์ เชี่ยวชาญเกษตร |
| อาจารย์ที่ปรึกษา | ผศ.ดร.อัศวิน มีชัย |
| หลักสูตร | วิทยาศาสตรมหาบัณฑิต |
| สาขาวิชา | ชีวสารสนเทศและชีววิทยาระบบ |
| ภาควิชา | ชีวสารสนเทศและชีววิทยาระบบ |
| คณะ | ทรัพยากรชีวภาพและเทคโนโลยี และเทคโนโลยีสารสนเทศ |
| พ.ศ. | 2553 |

E46912

## บทคัดย่อ

ในปัจจุบันนี้แหนมหรือผลิตภัณฑ์ไส้กรอกหมักของไทยถูกผลิตขึ้นโดยการใช้หัวเชื้อแบคทีเรียแบบเชิงพาณิชย์ แต่หัวเชื้อแบคทีเรียเหล่านี้สามารถที่จะขยายเชื้อได้ทันทีโดยกลุ่มผู้ผลิตแหนมที่ได้ซื้อหัวเชื้อแหนมเหล่านี้ไปใช้ ส่งผลให้หัวเชื้อแหนมนี้ไม่สามารถนำไปใช้ขายหรือต่อยอดทางธุรกิจได้ เพื่อที่จะแก้ไขปัญหาดังกล่าวข้างต้น หัวเชื้อแบบออกโซโทรพ (auxotroph) จึงถูกพัฒนาขึ้นเพื่อนำมาใช้ในกระบวนการผลิตแหนม โดยวิธีการแบบดั้งเดิมที่ง่าย และรวดเร็วสำหรับการพัฒนาออกโซโทรพ คือการใช้วิธีเรพลิคาเพลติง (replica plating) ซึ่งเป็นวิธีการทางอ้อม โดยวิธีการดังกล่าวนี้นักวิจัยจะไม่ทราบว่ายีนหรือกลไกอะไรที่จะนำไปสู่การทำให้เซลล์สูญเสียความสามารถในการสังเคราะห์สารอาหารที่จำเป็นต่อการเจริญ นอกจากวิธีการดังกล่าวแล้วยังมีวิธีการอื่นอีกที่จะนำมาใช้ในการพัฒนาออกโซโทรพ ซึ่งวิธีการหนึ่งที่นำมาใช้คือ การใช้เทคนิคของอณูชีววิทยา (molecular biology) ซึ่งวิธีการนี้จะมีประสิทธิภาพสูง เนื่องจากว่าวิธีการนี้จะช่วยในการพัฒนาการออกแบบทางด้านพันธุกรรมของหัวเชื้อแหนมที่ยากสำหรับกลุ่มผู้ผลิตแหนมที่จะนำหัวเชื้อเหล่านี้ไปขยายเชื้อต่อ ส่งผลให้กลุ่มผู้ผลิตแหนมดังกล่าวต้องหันกลับมาซื้อหัวเชื้อแหนมอย่างต่อเนื่องจากผู้จำหน่ายหัวเชื้อแหนมเท่านั้น โดยขั้นตอนที่สำคัญสำหรับการพัฒนาออกโซโทรพโดยใช้วิธีการทางตรงนี้จะเกี่ยวข้องกับขั้นตอนของการออกแบบว่าจะตัดต่อที่ยีนอะไร ซึ่งการศึกษาครั้งนี้ได้นำแบบจำลองวิถีเมตาบอลิซึมในระดับจีโนม (genome-scale model) ของ *Lactobacillus plantarum* สายพันธุ์ดับเบิลยูซีเอฟเอสหนึ่ง (*Lactobacillus plantarum* WCFS 1) ที่ถูกตีพิมพ์แล้วนั้นมาใช้ในการทำนายหากลุ่มของยีนและเมตาโบไลท์ที่เป็นไปได้สำหรับการพัฒนา

E46912

ออกโซโทรพของหัวเชื้อแหนม โดยแบบจำลองนี้จะประกอบไปด้วย 724 ยีน, 766 ปฏิกิริยา และ 660 เมตาบอไลท์ ที่จะถูกนำมาจำลองโดยใช้โปรแกรมย่อยที่มีชื่อว่าโคบรา (COBRA Toolbox) เวอร์ชั่น 1.3.3 โดยทำงานบนซอฟท์แวร์แมทแลป (MATLAB) ซึ่งแบบจำลองวิถีเมตาบอลิซึมในระดับจีโนมนี้จะถูกจำลองภายใต้เงื่อนไขการกำหนดค่าฟังชั่น (objective function) เป็นค่าการสร้างชีวมวลที่มากที่สุด (maximum biomass) และเงื่อนไขการกำหนดปริมาณของอัตราการบริโภคและการผลิตสารอาหาร หลังจากนั้นแบบจำลองนี้จะถูกนำมาใช้ในการระบุหาคู่ของเมตาโบไลท์และยีน (metabolite-gene pairs) ที่จำเป็นสำหรับการเจริญของเซลล์โดยเปรียบเทียบในอาหารเลี้ยงเชื้อ 4 ชนิด ได้แก่ อาหาร Chemically defined medium (CDM), de Man Rogosa and Sharpe (MRS), Simulated Nham Broth (SNB) และ Actual Nham (AN) ภายใต้การจำลองการตัดปฏิกิริยา (single reaction knockout) และการจำลองการตัดยีน (single gene knockout) จากผลการทดลองพบว่ามี 4 ยีนและเมตาบอไลท์ ที่ประกอบไปด้วย *glmM* กับ D-glucosamine 1-phosphate, *murI* กับ D-glutamate, *alr* กับ D-alanine และ *dfrA* กับ 5,6,7,8-tetrahydrofolate ที่ถูกคัดเลือกให้มีความสำคัญสำหรับการเจริญของเซลล์ในอาหารทุกชนิด และมีเพียง 1 ยีนและเมตาบอไลท์ คือ *gadB* กับ 4-aminobutanoate เท่านั้นที่ถูกคัดเลือกให้มีความสำคัญในการเติบโตบนอาหาร SNB และ AN โดยยีนและเมตาบอไลท์ที่ถูกคัดเลือกเหล่านี้สามารถนำไปใช้ในการพัฒนาออกโซโทรพสำหรับ *Lactobacillus plantarum* สายพันธุ์ดับเบิลยูซีเอฟเอสหนึ่ง และการปรับปรุงสายพันธุ์ของหัวเชื้อแหนมบริสุทธิ์ต่อไปในอนาคต

คำสำคัญ: แบบจำลองระดับจีโนม / *Lactobacillus plantarum* สายพันธุ์ดับเบิลยูซีเอฟเอสหนึ่ง / แหนม / ออกโซโทรพ / ยีนและเมตาบอไลท์ที่ถูกคัดเลือก

# ACKNOWLEDGEMENTS

I am deeply indebted to my advisor, Asst. Prof. Dr. Asawin Meechai, for his insights and suggestions that helped to shape my research skills. His valuable feedback contributed greatly to this thesis. I also thank the committee members, Dr. Taweerat Vichitsoonthonkul, Dr. Peter Kurdi, and Dr. Vethachai Plengvidhya for all the invaluable advices.

I would like to thank Mrs. Amornpan Boontongchuay for her help, support, interest and valuable encouragement. I also thank my friends and my family for their infinite love, care, support, understanding, and encouragement.

Last, but not least, I would also like to thank to all of my teachers in the Bioinformatics program at King Mongkut's University of Technology Thonburi (KMUTT) for all given knowledge. Furthermore, I would like to convey my deepest gratitude to KMUTT for providing the facilities. Finally, my sincere appreciation is expressed to both National Center for Genetic Engineering and Biotechnology (BIOTEC), Thailand and KMUTT for the full scholarship, providing me an opportunity to complete this master degree.

# CONTENTS

| | PAGE |
|---|---|
| PREFACE | i |
| ENGLISH ABTRACT | ii |
| THAI ABTRACT | iv |
| ACKNOWLEDGEMENTS | vi |
| CONTENTS | vii |
| LIST OF TABLES | x |
| LIST OF FIGURES | xii |
| LIST OF TECHNICAL VOCABULARY AND ABBREVIATIONS | xiv |

**CHAPTER**

| | | |
|---|---|---|
| **1 INTRODUCTION** | | **1** |
| 1.1 | Background and rationale | 1 |
| 1.2 | Objectives | 3 |
| 1.3 | Scopes of works | 4 |
| 1.4 | Expected outputs | 4 |
| **2 THEORY AND LITERATURE REVIEWS** | | **5** |
| 2.1 | Lactic Acid Bacteria | 5 |
| 2.1.1 | Metabolism of Lactic Acid Bacteria | 5 |
| 2.1.2 | Lactic Acid Bacteria in Fermented Foods | 6 |
| 2.2 | Nham | 10 |
| 2.3 | Auxotroph | 12 |
| 2.3.1 | The Traditional Cause of Auxotroph | 13 |
| 2.3.2 | Isolation and Identification of Auxotroph | 14 |
| 2.4 | Genome-scale Metabolic Model | 16 |
| 2.4.1 | Tools for Metabolic Reconstruction | 17 |
| 2.4.2 | Software Packages for Genome-Scale Model | 17 |
| 2.4.3 | Flux Balance Analysis | 18 |
| 2.5 | The Constraint-based Reconstruction and Analysis (COBRA) Toolbox | 21 |

**3 MATERIALS AND METHODOLOGY** **25**

3.1 Materials 25

3.2 Methodology 25

3.2.1 Model Implementation 25

3.2.1.1 Data retrieval 27

3.2.1.2 Simulation with COBRA Toolbox 27

3.2.1.3 Validation with experiment data 30

3.2.2 Simulation and identification of essential genes 31

3.2.2.1 Simulation of *L. plantarum* WCFS 1 in various media 31

3.2.2.2 Essential genes identification using the single reaction and single gene knockout 32

3.2.3 Comparison and selection of essential genes based on constraint and no constraint simulations 33

3.2.3.1 Compared list of essential genes in each medium 33

3.2.3.2 Selected essential gene and metabolites 33

**4 RESULTS AND DISCUSSION** **35**

4.1 Model implementation 35

4.1.1 Data Retrieval 35

4.1.2 Simulation with COBRA Toolbox 42

4.1.3 Validation with experimental data 43

4.2 Simulation and identification of essential genes 43

4.2.1 Simulation model of *L. plantarum* WCFS 1 in 4 media 43

4.2.2 Identification of essential genes for 4 media using the single reaction and single gene knockout 46

4.3 Comparison and selection of essential genes based on constraint and no constraint simulations 50

4.3.1 Compared list of essential genes in each medium 50

4.3.2 Selected essential gene and metabolites 51

**5 CONCLUSIONS AND SUGGESTION** **60**

5.1 Conclusions 60

5.2 Suggestion 61

**REFERENCES** **62**

**APPENDIX**


| | | |
|---|---|---|
| A | The *Lactobacillus plantarum* WCFS 1 metabolic network data | 68 |
| B | The official name of metabolites' abbreviation in *Lactobacillus plantarum* WCFS1 metabolic network data | 123 |
| C | Flux constraints of many media used in simulation model | 144 |
| D | Media Compositions | 157 |

**CURRICULUM VITAE** **163**

# LIST OF TABLES

**TABLE** **PAGE**


| TABLE | | PAGE |
|---|---|---|
| 2.1 | Distribution of lactic acid bacteria and allied bacteria in fermented foods in Thailand (Tanasupawat and Komagata, 1995) | 9 |
| 4.1 | Basic characteristics of the *L. plantarum* WCFS 1 genome-scale metabolic model developed in 2006 and 2009 (Teusink et al., 2006 and 2009) | 37 |
| 4.2 | Comparison of different reaction between the *L. plantarum* WCFS 1 genome-scale metabolic models in 2006 and 2009 (Teusink et al., 2006 and 2009) | 38 |
| 4.3 | List of reactions whose corresponding EC number were added in this study | 39 |
| 4.4 | Metabolite uptake and secretion rates used in model constraints from the experiment of 25 and 100 mM glucose concentrations in CDM (mmol $gDW^{-1}$ $h^{-1}$) | 42 |
| 4.5 | Comparing biomass production rate of experiments and *In-silico* simulations at various growth dilution rates (Teusink et al., 2006) | 43 |
| 4.6 | Comparison of the metabolite in each medium | 44 |
| 4.7 | Simulated biomass values ($h^{-1}$) under 4 media according to constraint and no constraint simulation | 46 |
| 4.8 | A list of metabolite and gene candidates for auxotroph development | 54 |
| A.1 | Metabolic network data of carbohydrate metabolisms | 69 |
| A.2 | Metabolic network data of energy metabolisms | 76 |
| A.3 | Metabolic network data of lipid metabolisms | 77 |
| A.4 | Metabolic network data of nucleotide metabolisms | 82 |
| A.5 | Metabolic network data of amino acid metabolisms | 87 |
| A.6 | Metabolic network data of other amino acids metabolism | 94 |
| A.7 | Metabolic network data of glycan biosynthesis and metabolism | 95 |
| A.8 | Metabolic network data of biosynthesis of polyketides and nonribosomal peptides | 96 |
| A.9 | Metabolic network data of metabolism of cofactors and vitamins | 97 |
| A.10 | Metabolic network data of biosynthesis of secondary metabolites | 100 |
| A.11 | Metabolic network data of translation | 101 |
| A.12 | Metabolic network data of unassigned pathway | 103 |
| A.13 | Metabolic network data of transport metabolites | 110 |
| A.14 | Metabolic network data of exchange metabolites | 116 |
| A.15 | Metabolic network data of biomass reaction and macromolecule | 122 |
| B.1 | Abbreviations of metabolites in *L. plantarum* WCFS 1 metabolic network data | 124 |

C.1 Flux constraints (mmol gDW$^{-1}$ h$^{-1}$) of metabolite in Chemically Defined Medium (CDM) at dilution rates 0.11 and 0.47 h$^{-1}$ with 25 mM glucose concentration 145

C.2 Flux constraints (mmol gDW$^{-1}$ h$^{-1}$) of metabolite in Chemically Defined Medium (CDM) at dilution rates 0.08 and 0.11 h$^{-1}$ with 100 mM glucose concentration 149

C.3 Flux constraints (mmol gDW$^{-1}$ h$^{-1}$) of metabolite in de Man, Rogosa and Sharpe (MRS) medium, Simulated Nham Broth (SNB), and Actual Nham (AN) based on dilution rates 0.11 h$^{-1}$ with 25 mM glucose concentration of Chemically Defined Medium (CDM) 153

D.1 Composition of nutrient in Chemically Defined Medium (CDM) 158

D.2 Composition of nutrient in de Man, Rogosa, and Sharpe (MRS) medium (difco, Becton Dickinson, Sparks, MD, USA) 159

D.3 Composition of nutrient in Simulated Nham Broth (SNB) 159

D.4 Composition of nutrient in Actual Nham (AN) medium 159

D.5 Composition of metabolite in many medium 160

# LIST OF FIGURES



| FIGURE | | PAGE |
|---|---|---|
| 2.1 | The pathway of Homofermentative in Lactic Acid Bacteria (McDonald et al., 1991) | 7 |
| 2.2 | The pathway of Heterofermentative in Lactic Acid Bacteria (McDonald et al., 1991) | 8 |
| 2.3 | Nham or the traditional Thai fermented pork sausage | 10 |
| 2.4 | A pyrimidine dimer | 14 |
| 2.5 | Replica plating techniques to isolate auxotrophic mutants | 16 |
| 2.6 | The strategy of reconstruction metabolic network for a constraint-based approach | 22 |
| 2.7 | The application of using the COBRA Toolbox | 23 |
| 3.1 | Overview of methodology used to identify the essential components for auxothroph development | 26 |
| 3.2 | Input file preparation for simulation in COBRA Toolbox | 28 |
| 3.3 | The format of metabolic data stored in Excel spreadsheet for converting to the text file (.txt) before using with PHP script | 29 |
| 3.4 | The reaction list from the Excel spreadsheet was pasted to the workspace of EditPlus and the first and last lines of the list should be deleted every time before saving as the text file | 29 |
| 3.5 | Running of PHP script in command window | 30 |
| 4.1 | The format of metabolic reactions in Excel spreadsheet | 36 |
| 4.2 | Functional groups of metabolic reactions in *L. plantarum* WCFS 1 | 40 |
| 4.3 | The overview of *Lactobacillus plantarum* WCFS 1 metabolic network | 41 |
| 4.4 | Percentage of the intermediate, lethal, and silent genes from single reaction (blue circle) and single gene (pink circle) knockout simulations based on constraint condition | 48 |
| 4.5 | Percentage of the intermediate, lethal, and silent genes from single reaction (blue circle) and single gene (pink circle) knockout simulations based on no constraint condition | 49 |
| 4.6 | Comparing the number of lethal genes from the single gene knockout based on the constraint (the number of lethal genes in both circle combined together) and the no constraint (small circle) simulation | 50 |
| 4.7 | A number of essential genes in each medium under the constraint simulation | 51 |
| 4.8 | The number of essential genes after applying with the both criteria | 52 |
| 4.9 | The number of identified essential metabolites according to selected essential genes under criteria I and II | 53 |
| 4.10 | The number of essential metabolites after screening with criteria III | 53 |

4.11 Metabolic network pathways of potential gene and metabolite candidates for auxotroph development for all media 57

4.12 Metabolic network pathways of *dfrA* (dihydrofolate reductase) gene and 5, 6, 7, 8-tetrahydrofolate (thf) metabolite candidate in all media 58

4.13 Metabolic network pathways of *gadB* (glutamatedecarboxylase) gene and 4-aminobutanoate (4abut) metabolite candidate in all media 59

# LIST OF TECHNICAL VOCABULARY AND ABBREVIATIONS

| | | |
|---|---|---|
| % error | = | percentage of error |
| AN medium | = | actual nham medium |
| ATP | = | adenosine triphosphate |
| CDM | = | chemically defined medium |
| COBRA Toolbox | = | constraint-based reconstruction and analysis toolbox |
| $CO_2$ | = | carbon dioxide |
| EC number | = | enzyme Commission number |
| EMP | = | embden-meyerhof-parnas pathway |
| FBA | = | flux balance analysis |
| GAP | = | glyceraldehyde phosphate |
| $gDW^{-1}$ | = | per gram dry weight |
| $h^{-1}$ | = | per hour |
| KEGG | = | kyoto encyclopedia of genes and genomes |
| LAB | = | lactic acid bacteria |
| LP | = | linear programming |
| MRS medium | = | de man, rogosa, and sharpe medium |
| mM | = | millimolar |
| NADH | = | nicotinamide adenine dinucleotide |
| ODE | = | ordinary differential equation |
| $O_2$ | = | oxygen |
| PEP | = | phosphoenolypyruvate |
| PTS | = | phosphoenolypyruvate sugar phosphotransferase system |
| SBML | = | systems biology markup language |
| SNB | = | simulated nham broth |
| UV | = | ultraviolet radiation |