



โครงการ

การพัฒนาวิธีการสร้างเครือข่ายควบคุมการแสดงออกของยีนแบบมีเงื่อนไขด้วยเทคนิคบูลีน

Constraint-based Boolean network for reconstructing gene regulatory network

คณะผู้วิจัย

หัวหน้าโครงการ: นางสาวสมคิด บุมิ
ผู้ร่วมโครงการ: ผศ. ดร. อัครวิน มีชัย

สถาบันพัฒนาและฝึกอบรมโรงงานต้นแบบ
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี

งานวิจัยนี้ได้รับทุนอุดหนุนงบประมาณแผ่นดิน มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี
ปีงบประมาณ 2554



250400

โครงการ

การพัฒนาวิธีการสร้างเครือข่ายควบคุมการแสดงออกของยีนแบบมีเงื่อนไขด้วยเทคนิคบูลีน

Constraint-based Boolean network for reconstructing gene regulatory network



คณะผู้วิจัย

หัวหน้าโครงการ: นางสาวสมคิด บุมิ

ผู้ร่วมโครงการ: ผศ. ดร. อัครวิน มีชัย

สถาบันพัฒนาและฝึกอบรมโรงงานต้นแบบ
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรีงานวิจัยนี้ได้รับทุนอุดหนุนงบประมาณแผ่นดิน มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี
ปีงบประมาณ 2554

รายงานการวิจัยฉบับสมบูรณ์

ประจำปีงบประมาณ 2554

1. ชื่อโครงการวิจัย

ชื่อเรื่อง (ภาษาไทย) การพัฒนาวิธีการสร้างเครือข่ายควบคุมการแสดงออกของยีนแบบมีเงื่อนไขด้วยเทคนิคบูลีน
(ภาษาอังกฤษ) Constraint-based Boolean network for reconstructing gene regulatory network

2. หน่วยงานหลักที่รับผิดชอบงานวิจัย

ชื่อหน่วยงาน สถาบันพัฒนาและฝึกอบรมโรงงานต้นแบบ
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี
เลขที่ 49 ซอยเทียนทะเล 25 ถนนบางขุนเทียน - ชายทะเล
แขวงท่าข้าม เขตบางขุนเทียน, กรุงเทพฯ 10150

3. คณะผู้วิจัย

3.1 หัวหน้าโครงการ

นางสาวสมคิด บุนี
ตำแหน่ง ผู้ช่วยนักวิจัย
สถานที่ทำงาน สถาบันพัฒนาและฝึกอบรมโรงงานต้นแบบ
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี (บางขุนเทียน)
83 หมู่ 8 แขวงท่าข้าม เขตบางขุนเทียน กรุงเทพฯ 10150
โทรศัพท์ 02-4707531
โทรสาร 02-4523455
ประสบการณ์ในงานวิจัย Systems biology and Bioinformatics
ความรับผิดชอบต่อโครงการที่เสนอ คิดเป็น 90% ของงานทั้งหมด

3.2 ผู้ร่วมโครงการ

ผศ.ดร. อัครวิน มีชัย
ตำแหน่ง ผู้ช่วยศาสตราจารย์
สถานที่ทำงาน: ภาควิชาวิศวกรรมเคมี คณะวิศวกรรมศาสตร์
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี
126 ถนนประชาอุทิศ ทุ่งครุ กรุงเทพฯ 10140
โทรศัพท์: 02-4709616
โทรสาร: 02-4279623
ประสบการณ์ในงานวิจัย Systems biology, Metabolic engineering & Bioinformatics
ความรับผิดชอบต่อโครงการที่เสนอ คิดเป็น 10% ของงานทั้งหมด

ปัจจุบันเทคนิคไมโครอะเรย์ทำให้เราสามารถศึกษาเครือข่ายการแสดงออกของยีนในสิ่งมีชีวิตได้ และวิธีการทางคอมพิวเตอร์ที่มีการพัฒนาขึ้นมานั้นโดยส่วนมากใช้ศึกษาเครือข่ายควบคุมการแสดงออกของยีนโดยปราศจากการคำนึงถึงเงื่อนไขของกลไกการทำงานที่มีในสิ่งมีชีวิต ในการศึกษารั้งนี้ ผู้วิจัยได้นำเสนอวิธีการสร้างเครือข่ายควบคุมการแสดงออกของยีนแบบมีเงื่อนไขโดยใช้เทคนิคบูลีน เพื่อใช้ศึกษาเครือข่ายควบคุมการแสดงออกของยีนในกระบวนการสลายน้ำตาลกาแลคโตส ภายใต้เงื่อนไขการเปลี่ยนแปลงสภาวะจากการหายใจแบบไม่ใช้ออกซิเจนโดยใช้กลูโคสเป็นแหล่งพลังงานเป็นการหายใจแบบใช้ออกซิเจน จากข้อมูลการแสดงออกของยีนที่ได้มาจากเทคนิคไมโครอะเรย์ วิธีการดังกล่าวประกอบด้วย 3 ขั้นตอนหลัก ขั้นตอนแรก คือ การแปลงข้อมูลจริงของการแสดงออกของยีนมาเป็นข้อมูลเพียงสองระดับคือ 0 และ 1 โดยเปรียบเทียบ 3 วิธีด้วยกัน คือ Max-x%Max Mean และ Sign of log ratio ซึ่งจะใช้ค่าความเหมือนกันระหว่างข้อมูลที่มีการแปลงเป็นสองระดับและข้อมูลจริงเป็นตัวตัดสินว่าควรใช้วิธีการใดในการแปลงข้อมูลก่อนที่จะสร้างเครือข่ายควบคุมการแสดงออกของยีน ขั้นตอนที่สองคือ การสร้างเครือข่ายควบคุมการแสดงออกของยีนจากข้อมูลที่แปลงเป็น 0 และ 1 ด้วยวิธีการสร้างเครือข่ายควบคุมการแสดงออกของยีนแบบมีเงื่อนไขด้วยเทคนิคบูลีนที่ได้พัฒนาขึ้น จากภาษา C++ เรียกโปรแกรมนี้ว่า *เทคนิคบูลีนแบบมีเงื่อนไข* ขั้นตอนสุดท้ายคือการสร้างเครือข่ายควบคุมการแสดงออกของยีนในรูปแบบกราฟที่แสดงถึงการควบคุมแบบกระตุ้นและการควบคุมแบบยับยั้ง ซึ่งพบว่าเทคนิคบูลีนแบบมีเงื่อนไขดังกล่าวสามารถใช้สร้างเครือข่ายควบคุมการแสดงออกของยีนได้รวมทั้งมีความแม่นยำมากขึ้นกว่าเครือข่ายควบคุมการแสดงออกของยีนที่สร้างด้วยวิธีเทคนิคบูลีนแบบไม่พิจารณาเงื่อนไขสามารถลดความซับซ้อนของเครือข่ายควบคุมการแสดงออกของยีน โดยการลดความผิดพลาดในการทำนายความสัมพันธ์ โดยสามารถเพิ่มความถูกต้องขึ้น 25%

นอกจากนี้ เทคนิคบูลีนแบบมีเงื่อนไขที่พัฒนาขึ้นนั้นยังสามารถนำไปประยุกต์ใช้สร้างเครือข่ายควบคุมการแสดงออกของยีนในกระบวนการสร้างแป้งในพืช *A. thaliana* ภายใต้สภาวะกลางวันและกลางคืนได้ โดยสามารถทำนายโปรตีนควบคุมที่เกี่ยวข้องกับเอนไซม์ในกระบวนการสร้างแป้งทั้งกระบวนการสังเคราะห์แป้ง และกระบวนการสลาย พบว่า โปรตีนควบคุม 2 ตัว คือ At4g39780 and At2g43010 เกี่ยวข้องกับ เอนไซม์ในกระบวนการสลายแป้ง นั่นคือ a cytosolic disproportionating enzyme like protein (At2g40840; DPE2) alpha-amylase like 3 (At1g69830; AMY3) และ a plastidic alpha-glucan phosphorylase (At3g29320; PHS1) และมีโปรตีนควบคุม 1 ตัว คือ At2g28200 ที่เกี่ยวข้องกับเอนไซม์ 5 ตัว คือ คือ phosphoglucomutase (At5g51820; PGM1) Starch synthase II (At3g01180; STS2) Starch branching enzyme III (At2g36390; SBE3) Glucan water dikinase 1 (At1g10760; GWD1/SEX1) และ Glucanotransferase (At5g64860; DPE1) จะเห็นได้ว่า เทคนิคบูลีนแบบมีเงื่อนไขนั้นสามารถใช้สร้างเครือข่ายควบคุมการแสดงออกของยีนจากข้อมูลไมโครอะเรย์ และทำให้ได้เครือข่ายควบคุมการแสดงออกของยีนที่สอดคล้องกับข้อมูลทางชีววิทยามากขึ้น

คำสำคัญ : ไมโครอะเรย์ / เครือข่ายควบคุมการแสดงออกของยีน / เทคนิคบูลีนแบบมีเงื่อนไข

Abstract

250400

The availability of large-scale expression microarray data allows the inference of underlying gene regulatory network of organisms. Various computational techniques developed for this task often infer the circuits directly from the expression data without taking into consideration of biological information *in priori*. Here, we present a constraint-based Boolean approach that integrates biological constraints into a Boolean algorithm to formulate a genetic circuit of *Saccharomyces cerevisiae* during the diauxic shift in a glucose-limited culture from time-series expression data. This approach includes three main steps. First, the gene expression data was discretized to either '0' or '1' based on three discretization methods, Max-x%Max, Mean, and Sign of log ratio. A similarity score was used as a criterion for selecting a suitable discretization method. Second, the discretized data was employed to determine gene relationships using a developed constraint-based Boolean algorithm. This algorithm implemented by C⁺⁺ programming considers all possibility of gene circuits and then specifically selects the circuits that correspond with both the expression data and a set of pre-defined biological constraints, so called "*constraint-based Boolean network*". Last, the gene regulatory network was depicted as a directed graph model describing activation and inhibition among genes. The evaluating result from inferred network between constraint and non-constraint-based algorithm shows that our developed algorithm can reduce false prediction and gain 25% increase of accuracy over non-constraint-based algorithm.

Moreover, we demonstrated the validity of the technique by employing it to infer the gene regulatory network of starch metabolism from time series expression data of *Arabidopsis thaliana* taken diurnally. Inferred network showed that two transcription factors, *At4g39780* and *At2g43010*, co-regulate three enzymes in starch degradation, a cytosolic disproportionating enzyme like protein (*At2g40840*; *DPE2*), alpha-amylase like 3 (*At1g69830*; *AMY3*), and a plastidic alpha-glucan phosphorylase (*At3g29320*; *PHS1*). There is only one transcription factor, *At2g28200*, regulates transcription of five enzymes in starch metabolism, phosphoglucomutase (*At5g51820*; *PGM1*) Starch synthase II (*At3g01180*; *STS2*) Starch branching enzyme III (*At2g36390*; *SBE3*) Glucan water dikinase 1 (*At1g10760*; *GWD1/SEX1*) and Glucanotransferase (*At5g64860*; *DPE1*) The results show constraint-based Boolean network shall enable the better gene regulatory network from large-scale gene expression data which is corresponding to biological data.

Keywords: Microarray / gene regulatory network / constraint-based Boolean network

สารบัญเรื่อง

	หน้า
บทคัดย่อภาษาไทย	3
บทคัดย่อภาษาอังกฤษ	4
สารบัญเรื่อง	5
สารบัญตาราง	7
สารบัญรูป	8
บทที่ 1 บทนำ	9
1.1 ความสำคัญ และที่มาของปัญหาที่ทำการวิจัย	9
1.2 วัตถุประสงค์ของโครงการ	10
1.3 ขอบเขตของโครงการวิจัย	10
1.4 ประโยชน์ที่คาดว่าจะได้รับ	11
1.5 แผนการถ่ายทอดเทคโนโลยีหรือผลการวิจัยสู่กลุ่มเป้าหมาย	11
1.6 แผนการดำเนินงานตลอดโครงการ	11
บทที่ 2 ระเบียบวิธีวิจัย	12
2.1 การเก็บและรวบรวมข้อมูล	12
2.2 การเตรียมข้อมูล	12
2.3 ประเมินวิธีการที่ใช้ในการแปลงข้อมูลการแสดงผลการออกของยีน	13
2.4 พัฒนาโปรแกรมที่ใช้ในการสร้างเครือข่ายการแสดงผลการออกของยีนแบบมีเงื่อนไข	13
2.5 การทดสอบและประเมินโปรแกรม	13
2.6 การประยุกต์ใช้โปรแกรมเพื่อสร้างและศึกษาเครือข่ายควบคุมการแสดงผลการออกของยีนในกระบวนการสังเคราะห์แบ่ง	13
บทที่ 3 ผลการดำเนินงานวิจัย	15
3.1 การเก็บและรวบรวมข้อมูล	15
3.2 การเตรียมข้อมูล	15
3.3 แปลงค่าการแสดงผลการออกของยีนและการประเมินวิธีการที่ใช้ในการแปลงข้อมูลการแสดงผลการออกของยีน	19
3.4 พัฒนาโปรแกรมที่ใช้ในการสร้างเครือข่ายการแสดงผลการออกของยีนแบบมีเงื่อนไข	24
3.5 การทดสอบและประเมินโปรแกรม	28
3.6 การประยุกต์ใช้โปรแกรมเพื่อสร้างและศึกษาเครือข่ายควบคุมการแสดงผลการออกของยีนในกระบวนการสังเคราะห์แบ่ง	34
บทที่ 4 สรุปและเสนอแนะ	37
เอกสารอ้างอิง	38
ผลงานตีพิมพ์	40

สารบัญเรื่อง (ต่อ)

	หน้า
ภาคผนวก	41
ก โค้ดสคริป MATLAB สำหรับคำนวณหา similarity score	42
ข โค้ดสคริป C++ สำหรับสร้าง gene regulatory network ของกลุ่มยีน ที่สนใจด้วยวิธี Constraint-based Boolean network	44
ค ยีนที่ใช้ในการสร้างเครือข่ายควบคุมการแสดงออกของยีน (gene regulatory network) ของกระบวนการสังเคราะห์แบ่ง	57

สารบัญตาราง

ตารางที่	หน้า
1.1	11
3.1	15
3.2	20
3.3	21
3.4	21
3.5	22
3.6	22
3.7	23
3.8	28
3.9	28
3.10	30
3.11	31
3.12	31
3.13	31
3.14	33
3.15	33
3.16	34

สารบัญรูป

รูปที่		หน้า
2.1	วิธีการดำเนินวิจัยโดยสังเขป	11
2.2	ขั้นตอนการวิเคราะห์เครือข่ายควบคุมการแสดงออกของยีนในกระบวนการสังเคราะห์แป้ง	13
3.1	การกระจายตัวของข้อมูลการแสดงออกของยีนของทุกช่วงเวลา (T_1-T_7) ในรูปแบบ \log_2 ratio	16
3.2	การกระจายตัวของข้อมูลการแสดงออกของยีนของทุกช่วงเวลา (T_1-T_7) ที่ผ่านการ normalization ด้วย background correction แล้ว	16
3.3	เปรียบเทียบ IR plot ของข้อมูลการแสดงออกของยีนแต่ละช่วงเวลา (T_1-T_4)	17
3.4	เปรียบเทียบ IR plot ของข้อมูลการแสดงออกของยีนแต่ละช่วงเวลา (T_5-T_7)	18
3.5	การกระจายตัวของข้อมูลการแสดงออกของยีนของทุกช่วงเวลา (T_1-T_7) ที่ผ่านการ normalization ด้วยวิธี <i>lowess</i>	19
3.6	การแสดงออกของกลุ่มยีนที่เกี่ยวข้องกับ galactose pathway ในทุกช่วงเวลา (T_1-T_7) ที่ผ่านการ normalization ด้วยวิธี <i>lowess</i>	20
3.7	การเปรียบเทียบความเหมือนกันระหว่างเดนโดแกรมของข้อมูลที่ใช้วิธี Mean	23
3.8	Similarity score ของการเปรียบเทียบเดนโดแกรมที่ได้จากวิธีการ Max-x%Max	24
3.9	ขั้นตอนการพัฒนาโปรแกรมการสร้างเครือข่ายการแสดงออกของยีนแบบมีเงื่อนไข	25
3.10	เปรียบเทียบ Classical Boolean network) และ Constraint-based Boolean network	26
3.11	เครือข่ายควบคุมการแสดงออกของยีนด้วยโปรแกรม Constraint-based Boolean network	32
3.12	ผลของการประเมินประสิทธิภาพของโปรแกรม Constraint-based Boolean network	33
3.13	การเปรียบเทียบประสิทธิภาพของการเครือข่ายควบคุมการแสดงออกของยีนที่สร้างจากโปรแกรม Constraint-based Boolean network ที่พัฒนาขึ้นเปรียบเทียบกับ Non-Constraint-based Boolean network (classical Boolean algorithm)	34
3.14	เครือข่ายควบคุมการแสดงออกของยีนในกระบวนการสังเคราะห์แป้ง	36