การค้นหาหน้าที่และความสัมพันธ์ของกลุ่มยีนด้วย Gene Ontology

จุฑารัตน์ มณีวัฒนาพฤกษ์ และ นพดล คีรีเพ็ชร

ฝ่ายวิจัยและพัฒนาเทคโนโลยีคอมพิวเตอร์เพื่อการคำนวณ
ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ
112 อุทยานวิทยาศาสตร์แห่งประเทศไทย ถนนพหลโยธิน ต. คลองหนึ่ง อ. คลองหลวง จ. ปทุมธานี 12120
โทรศัพท์ 025646900 โทรสาร 025646772

E-mail: jutarat.maneewattanapluk@nectec.or.th, noppadon.khiripet@nectec.or.th

บทคัดย่อ - บทความนี้กล่าวเกี่ยวกับรูปแบบด้วยความก้าวหน้าทางเทคโนโลยีการถอดรหัสพันธุกรรมของแบคทีเรีย ทำให้พบยีนอีกเป็นจำนวนมากที่ขาดข้อมูลที่อธิบายการทำงานและรายละเอียดของยีน ปัจจุบันมีแหล่งข้อมูลอ้างอิง เช่น Gene Ontology (GO) ที่ทำการรวบรวมความหมายของยีน และข้อมูลที่เกี่ยวข้องกับยีนที่รวบรวมได้จากสิ่งมีชีวิตชนิดต่างๆ แต่จะพบข้อจำกัดที่ข้อมูลไม่ครอบคลุมถึงทุกยีนของแบคทีเรียทุกสายพันธุ์ ดังนั้นเพื่อเป็นการเพิ่มเติมข้อมูลหน้าที่ของยีนส่วนที่ขาดไป งานวิจัยนี้ใช้ข้อมูลกลุ่มยีนที่มีความคล้ายของแบคทีเรีย ด้วยสมมุติฐานว่ายีนในกลุ่มเดียวกันน่าจะมีหน้าที่การทำงานที่คล้ายกัน เป็นข้อมูลตั้งต้นเพื่อค้นหาหน้าที่ของยีนส่วนใหญ่ในกลุ่ม โดยใช้ข้อมูลอ้างอิงจาก GO ที่ทำการสรุปความสัมพันธ์ระหว่าง gene products ในยีนที่กำหนดคำนิยามตามมาตรฐานในระบบ GO มาทำการสืบค้นกับข้อมูลกลุ่มยีน ซึ่งได้สรุปรายการคำนิยามที่พบ จากนั้นใช้วิธีการวิเคราะห์ตัววัดหลายๆ ตัว (multiple test) เพื่อใช้วัดความเหมาะสมของนิยามที่แทนยีนส่วนใหญ่ในกลุ่ม ผลที่ได้จากการจัดกลุ่มและการหาหน้าที่ของยีนสามารถนำไปศึกษารายละเอียดเพิ่มเติมเพื่อทำความเข้าใจเกี่ยวกับหน้าที่การทำงานของยีนในสภาวะต่างๆ เช่น การเกิดโรค หรือศึกษาพิษวิทยาของยาหรือสารต่างๆ ที่มีผลต่อสิ่งมีชีวิต เป็นต้น

คำสำคัญ - Gene Ontology, multiple test, การเปรียบเทียบพันธุกรรม, ลักษณะทางพันธุกรรมของจุลินทรีย์

1. บทนำ

นับตั้งแต่มีการถอดรหัสพันธุกรรมของสิ่งมีชีวิตสำเร็จ ความต้องการศึกษาคุณสมบัติของยีนได้รับความสนใจเพิ่มขึ้น ทั้งการเปรียบเทียบความคล้ายระหว่างยีนในสิ่งมีชีวิตสายพันธุ์เดียวกันและข้ามสายพันธุ์ รวมถึงส่วนประกอบในระดับชีวโมเลกุลของยีน โดยมี Gene Ontology consortium [1] ที่ทำการรวบรวมความหมายของยีน และข้อมูลที่เกี่ยวข้องกับยีนที่รวบรวมได้จากสิ่งมีชีวิตชนิดต่างๆ

ผู้ใช้งานสามารถค้นหาหน้าที่ของยีนที่สนใจในระบบของ GO ด้วยการใช้รหัสของ GO (GO ID) หรือคำนิยาม (GO term) ซึ่งจะอธิบายการความหมายยีนที่พบเป็นโครงสร้างแบบลำดับชั้น (hierarchical) ประกอบด้วยระดับ molecular functions, biological processes และ cellular locations พร้อมแหล่งข้อมูลสิ่งมีชีวิตที่พบยีน โดยรวบรวมสิ่งมีชีวิตได้เพียงบางชนิด ได้แก่ *Saccaromyces cerevisiae* (budding yeast)[2], drosophila [3] และ mouse [4] เป็นต้น

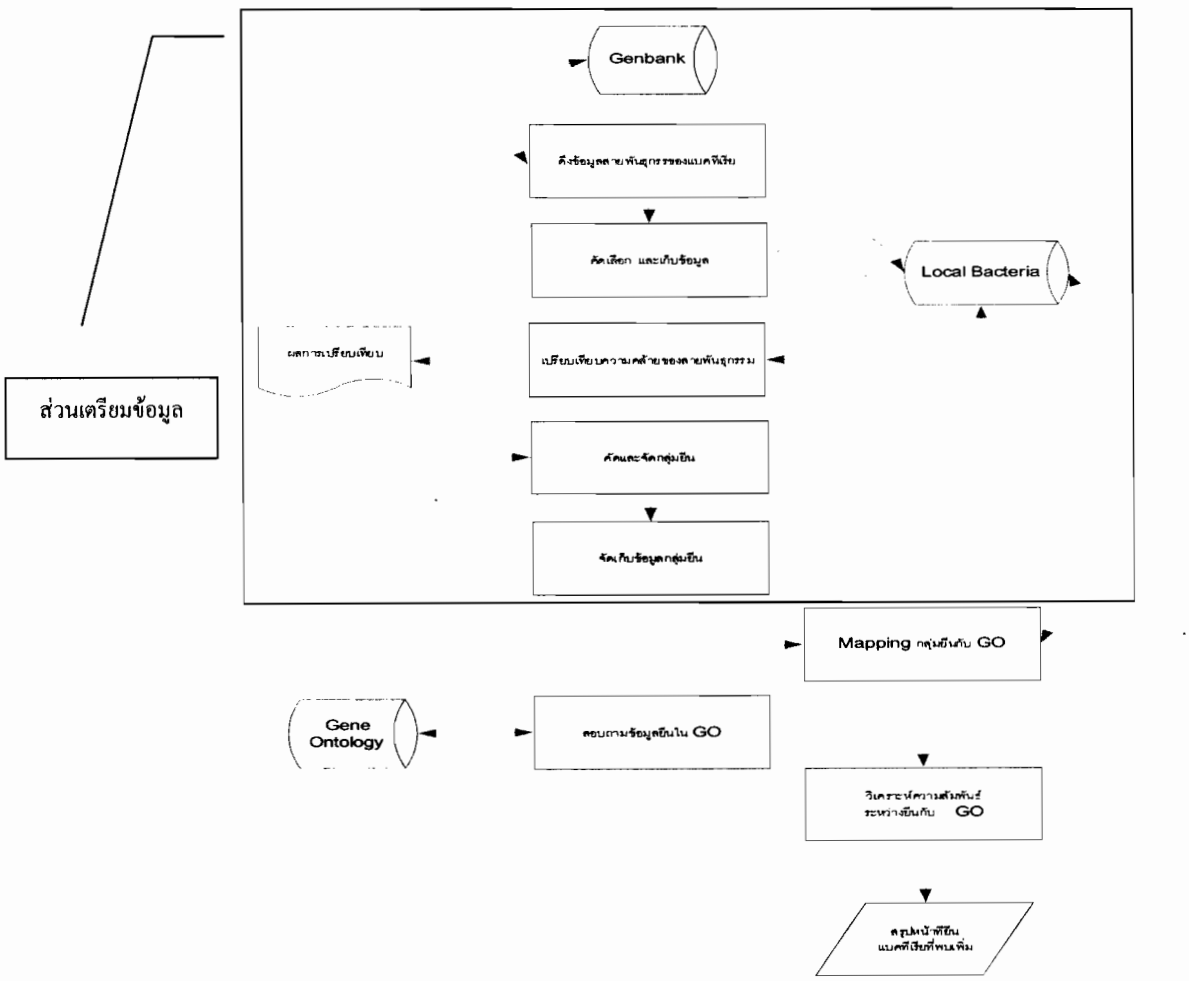
นอกจากนี้พบว่าโปรแกรมช่วยในการวิเคราะห์ความสัมพันธ์ระหว่างรายชื่อยีนกับ GO term หลายโปรแกรมใช้ข้อมูล gene expression มาผ่านการจัดกลุ่มยีน จากนั้นนำผลมาหาความสัมพันธ์ระหว่างรายชื่อยีนที่ได้กับ GO โดยมีการวิเคราะห์ทางสถิติช่วยให้ผู้ใช้ตัดสินใจยอมรับผลที่ได้ เป็นต้นว่า GoMiner [5] และ GOTree Machine [6] ใช้ค่า P-Value พิจารณายอมรับคำนิยามยีนที่สนใจ หรือ MAPPFinder [7] ใช้การคำนวณเปอร์เซ็นต์เพื่อคิดจำนวนยีนที่สัมพันธ์กับคำนิยาม และ Z score คิดค่าการยอมรับคำนิยามที่ได้ ส่วน FatiGO [8] นั้นใช้การ

วิเคราะห์ Fisher's Exact Test [9] เพื่อหา Go term ที่สัมพันธ์กับรายชื่อยีนในรูปตารางความสัมพันธ์แบบ 2x2 และวิธีการวิเคราะห์ตัววัดหลายๆ ตัว (multiple test) ด้วยการใช้ P-Value เพื่อพิสูจน์ว่า Go term ที่ได้สัมพันธ์กับยีน

ถึงแม้ว่า Gene Ontology consortium ได้รวบรวมข้อมูลที่สามารถใช้เพื่ออ้างอิงยีนจากสิ่งมีชีวิตหลายสายพันธุ์ แต่สำหรับข้อมูลของแบคทีเรียที่มีการถอดรหัสพันธุกรรมสายพันธุ์ใหม่ๆ นั้น ข้อมูลหน้าที่ยีนมีไม่ครอบคลุมแบคทีเรียทุกสายพันธุ์ ดังนั้นการค้นหาหน้าที่ของยีนส่วนที่ขาดไปจึงเป็นปัญหาสำหรับงานวิจัยนี้ที่ต้องการวิธีที่สามารถแสดงข้อมูลนิยามของยีน ข้อมูลเริ่มต้นใช้รายชื่อยีนของแต่ละกลุ่มยีนซึ่งเป็นผลจากการจัดกลุ่มยีนที่ได้จากการเปรียบเทียบความคล้ายด้วยโปรแกรม Blast [10] ซึ่งเป็นตามสมมุติฐานว่ายีนในกลุ่มเดียวกันน่าจะมีหน้าที่การทำงานที่คล้ายกัน ได้ถูกนำมาใช้ร่วมกับการค้นหาคำนิยามที่เหมาะสมตามมาตรฐานของ Gene Ontology ด้วยการ mapping นิยามของยีนใน GO กับรายชื่อยีน จากนั้นคำนวณทางสถิติเพื่อการยอมรับคำนิยามที่สัมพันธ์กับยีนส่วนใหญ่ในกลุ่มด้วยวิธีการวิเคราะห์ตัววัดหลายๆ ตัว (multiple test) ซึ่งเป็นการวิเคราะห์ผลว่า Go term ที่ได้เกี่ยวข้องกับจำนวนยีนส่วนใหญ่

2. วิธีการ

งานวิจัยนี้เป็นการใช้ข้อมูลยีนในกลุ่มยีนที่มีความคล้าย ซึ่งเป็นผลจากการจัดกลุ่มยีนที่ได้ผลจากการใช้โปรแกรม Blast ทำการเปรียบเทียบลำดับสายพันธุกรรมของยีนในแบคทีเรียจำนวน 143 สายพันธุ์ [11] โดยกำหนดใช้ข้อมูลตัวอย่างจากกลุ่มยีนที่มีความคล้ายกลุ่มหนึ่งที่มีจำนวนยีนรวม 14 ยีน สรุปเป็นผังของงานดังแสดงในรูปที่ 1



รูปที่ 1 แสดงผังลำดับขั้นตอนของงานการค้นหาหน้าที่และความสัมพันธ์ของกลุ่มยีนด้วย Gene Ontology

ขั้นตอนงานเริ่มจากใช้ระบบฐานข้อมูลจัดเก็บข้อมูลที่มีความสัมพันธ์กับการค้นหาหน้าที่ของยีน ตามรูปประกอบส่วนการเตรียมข้อมูลซึ่งประกอบไปด้วยข้อมูลดังต่อไปนี้

1. ข้อมูลยีน ซึ่งใช้ระบุว่ายีนที่สนใจคือยีนอะไร โดยทำการค้นหาและจัดเก็บข้อมูลจาก Genbank ข้อมูลที่จัดเก็บประกอบไปด้วย gene index, gene product และ organism
2. ข้อมูลยีนในกลุ่มยีนที่มีความคล้าย ประกอบด้วย gene index, group number และค่า expect value

ด้วยการออกแบบระบบฐานข้อมูลที่ใช้ gene index เป็น key เชื่อมความสัมพันธ์ระหว่างข้อมูล ทำให้สามารถทำการสืบค้น gene product ของยีนในแต่กลุ่มยีนได้พร้อมกัน รวมทั้งช่วยลดขั้นตอนการค้นหาและสืบค้นโดยตรงไปที่ Genebank แบบเดิม ที่ทำการค้นหายีนได้ครั้งละยีน สามารถดูสรุปข้อมูลที่เป็นผลจากการทำงานในขั้นตอนนี้ได้ส่วนผลการทดลอง

ลำดับงานต่อไปเป็นการค้นหา Gene term ที่สัมพันธ์กับ gene product ซึ่งเป็น key ที่ได้จากการเตรียมข้อมูลส่วนแรก โดยส่งค่า gene product ของแต่ละยีน ไปทำการสืบค้นในระบบของ GO จากนั้นนำผลการสืบค้นทั้งหมดซึ่งเป็นยีนที่สัมพันธ์กับ GO term มาทำการวิเคราะห์ตัดสินใจเลือกว่า GO term ใดที่ควรเป็นคำนิยามที่อธิบายกลุ่มยีนได้ วิธีการป้องกันปัญหาไม่ให้โอกาสสรุปผิดในกรณีวิเคราะห์แต่ละ GO term ที่พบยีนนั้น วิธีหนึ่งทำได้คือการใช้ Fisher's Exact Test ซึ่งเป็นวิธีการวิเคราะห์ข้อมูลที่สองกลุ่มไม่มีความสัมพันธ์กัน โดยแสดงในรูปตาราง 2x2 ดังแสดงในรูปที่ 2

A เป็นผลรวมจำนวนยีนที่พบใน Go term แรก
B เป็นผลรวมจำนวนยีนที่ไม่พบใน Go term แรก
C เป็นผลรวมจำนวนยีนที่พบใน Go term สอง
D เป็นผลรวมจำนวนยีนที่ไม่พบใน Go term สอง
N คือจำนวนตัวอย่างยีน

A	**B**	**A+B**
C	**D**	**C+D**
A+C	**B+D**	**N**

รูปที่ 2 แสดงตาราง 2x2 เพื่อคำนวณ Fisher's Exact Test

โดยวิธีการคำนวณเป็นการคิดค่าความน่าจะเป็นจากผลความถี่ของยีนแต่ละ GO term ตามสูตรการคำนวณ คือ

$$p = \frac{\binom{A+C}{A}\binom{B+D}{B}}{\binom{N}{A+B}} = \frac{(A+B)!(C+D)!(A+C)!(B+D)!}{A!\,B!\,C!\,D!\,N!}$$

โดยการสรุปว่า GO term ที่พบมีความแตกต่างกันสามารถเป็นตัวแทนคำนิยามให้กับกลุ่มยีนที่พบได้ ก็ต่อเมื่อค่า P-Value ที่วิเคราะห์ได้นี้มีค่าน้อยกว่าหรือเท่ากับค่า cut off P-Value ที่คำนวณมาจากผลรวมของ P-Value ของแต่ละ GO term หารด้วยจำนวน GO term ซึ่งวิธีตัดสินใจนี้ใช้ผลการวิเคราะห์สถิติแบบหลายตัว (multiple test) โดยใช้ค่า P-Value เป็นค่าเบื้องต้นเพื่อพิสูจน์สมมุติฐานการยอมรับ GO term สามารถดูสรุปผลได้จากส่วนของการทดลอง

3. ผลการทดลอง

การทดลองใช้ผลของการจัดกลุ่มยีน จากตัวอย่างแบคทีเรีย 5 สายพันธุ์ รวมจำนวนยีนทั้งหมด 16552 ยีน โดยข้อมูลของแบคทีเรียทั้ง 5 สายพันธุ์ มีดังนี้

ชื่อแบคทีเรีย	จำนวนยีน	จำนวนข้อมูลที่เกิดใน Blast
1. Aeropyrum_pernix_K1	1841	487681
2. Agrobacterium_tumefaciens_C58_Cereon	5299	393132
3. Agrobacterium_tumefaciens_C58_UWash	5402	1365803
4. Aquifex_aeolicus	1560	326049
5. Archaeoglobus_fulgidus	2420	1359554

เมื่อนำข้อมูลซึ่งเป็นผลการจับคู่ยีนด้วย Blast มาจัดกลุ่มยีนที่คล้าย พบว่าจะได้ข้อมูลจำนวน 5248 กลุ่ม และมีจำนวนยีนที่ถูกจัดกลุ่มทั้งสิ้น 13069 ยีน สามารถสรุปเป็นข้อมูลจำนวนกลุ่มที่มีจำนวนยีนที่คล้ายได้ตามตารางที่ 1

ตารางที่ 1 แสดงสรุปจำนวนกลุ่มแยกตามจำนวนยีนที่พบในกลุ่ม

จำนวนยีน	รวมจำนวนกลุ่ม (กลุ่ม)	รวมจำนวนยีนทั้งหมด (ยีน)
14	1	14
13	1	13
12	3	36
11	2	22
10	9	90
9	15	135
8	24	192
7	51	357
6	59	354
5	222	1110
4	214	856
3	596	1788
2	4051	8102

จากนั้นนำค่า gene index จากในกลุ่มยีนตัวอย่างที่มียีนจำนวน 14 ยีน มาทดลองหาคำนิยามที่สัมพันธ์กับยีนในกลุ่มนี้ด้วยการสร้างฐานข้อมูลที่เชื่อมความสัมพันธ์ระหว่าง gi, accession, organism, gene, synonym และ product ของยีนตัวอย่างนี้ ซึ่งสรุปเป็นข้อมูลดังในตารางที่ 2

ตารางที่ 2 แสดงสรุปข้อมูลเริ่มต้นที่เตรียมใช้ต่อในการค้นหาคำนิยามของยีน

Gi	Accession	Organism	gene	synonym	product
11498123	NC_000917	Archaeoglobus_fulgidus		AF0512	chloroplast inner envelope membrane protein
11499903	NC_000917	Archaeoglobus_fulgidus		AF2322	L-isoaspartyl protein carboxyl methyltransferase (pcm-2)
14601138	NC_000854	Aeropyrum_pernix		APE1011	hypothetical protein-L-isoaspartate O-methyltransferase
14601233	NC_000854	Aeropyrum_pernix		APE1183	hypothetical protein
15606625	NC_000918	Aquifex_aeolicus		aq_1457	hypothetical protein
15607083	NC_000918	Aquifex_aeolicus		aq_2139	putative protein
15888383	NC_003062	Agrobacterium_tumefaciens_C58_Cereon	AGR_C_1920	AGR_C_1920p	
15888938	NC_003062	Agrobacterium_tumefaciens_C58_Cereon	AGR_C_2998	AGR_C_2998p	
15889008	NC_003062	Agrobacterium_tumefaciens_C58_Cereon	AGR_C_3127	AGR_C_3127p	
15889027	NC_003062	Agrobacterium_tumefaciens_C58_Cereon	AGR_C_3159	AGR_C_3159p	
17934948	NC_003304	Agrobacterium_tumefaciens_C58_UWash		Atu1041	methyltransferase
17935521	NC_003304	Agrobacterium_tumefaciens_C58_UWash		Atu1625	methyltransferase
17935595	NC_003304	Agrobacterium_tumefaciens_C58_UWash		Atu1701	L-isoaspartyl protein carboxyl methyltransferase
17935615	NC_003304	Agrobacterium_tumefaciens_C58_UWash		Atu1721	protein-L-isoaspartate O-methyltransferase

จากยีนตัวอย่าง 14 ยีน พบว่ายีนพบข้อมูล product จำนวน10 ยีน ซึ่งต้องนำคำที่เป็น product นี้ไปใช้ค้นหาในระบบของ GO และทำการคำนวณ P-Value ตามสูตร Fisher's Exact Test ซึ่งได้ผลสรุปในตารางที่ 3

ตารางที่ 3 แสดงสรุปผล P-Value ของแต่ละ GO term

GO ID	GO term	P-Value
GO:0008150	biological_process unknown	0.6758
GO:0005525	GTP binding	0.6758
GO:0005554	molecular_function	0.6758
GO:0006355	regulation of transcription, DNA-dependent	0.6758
GO:0008168	methyltransferase activity	0.6758
GO:0008372	cellular_component unknown	0.1111
GO:0016020	Membrane	0.1111
GO:0004719	protein-L-isoaspartate (D-aspartate) O-methyltransferase activity	0.5
GO:0030091	protein repair	0.6612

การเลือกว่า GO term ใดที่จะแทนหน้าที่ของกลุ่มยีนนี้ ต้องพิจารณาค่า P-Value ของ GO term ที่มีค่าน้อยกว่า cut off P-Value ซึ่งมีค่า 0.529 ซึ่งได้แก่ GO:0008372 GO:0016020 และ GO:0004719

4. ปัญหาและแนวทางการพัฒนาต่อ

งานวิจัยนี้ใช้การวิเคราะห์ทางสถิติแบบ multiple test ช่วยในการแสดงความน่าเชื่อถือจากผลการค้นหาหน้าที่และความสัมพันธ์ของกลุ่มยีนกับ GO โดยใช้ข้อมูลกลุ่มยีนที่มีความคล้ายของแบคทีเรียมาทำการหาคำนิยามของยีนตามมาตรฐาน GO ซึ่งขั้นตอนในการนำยีนไป mapping กับ GO อาศัยข้อมูลของ gene product เป็น key ในการค้นหา และใช้การกำหนดค่าสถิติ P-Value ที่ได้จากวิธี Fisher's Exact Test เพื่อพิสูจน์การยอมรับผลที่ได้และตั้งอยู่บนสมมุติฐานที่ว่ากลุ่มทั้งสองไม่เกี่ยวข้องกันเลย

ด้วยการใช้วิธีการวิเคราะห์ดังที่ได้กล่าวมานั้นพบปัญหาที่ผลการคำนวณค่า P-Value ที่ใช้ตัดสินเลือก GO term ที่ระบุแทนกลุ่มยีนนั้นมีค่ามากกว่าหนึ่งค่า แม้จะปรับปรุงค่าด้วยการใช้ cut off P-Value ทั้งนี้เป็นเพราะข้อมูลที่เป็น key ในส่วนของการใช้ค้นหา GO term นั้น มีค่าว่างประกอบอยู่ เพื่อเป็นการแก้ปัญหาส่วนนี้ จึงต้องเพิ่มการปรับปรุงวิธีการค้นหาข้อมูลที่ขาดหายไปในการค้นหา gene product เพื่อเพิ่มความถูกต้องในการค้นหา GO term

นอกจากนี้แล้วกรณีข้อมูลมีจำนวนมากขึ้น ปัญหาที่ต้องพิจารณาเพื่อให้ผลการวิเคราะห์สมบูรณ์ ต้องเลือกการวิเคราะห์ทางสถิติแบบอื่นเพิ่มเติม เนื่องจากการใช้วิธี Fisher's Exact Test นั้นเหมาะกับปริมาณข้อมูลน้อยๆ ใช้การเปรียบเทียบครั้งละสองกลุ่ม ถ้าข้อมูลมีจำนวนมากต้องใช้เวลาการวิเคราะห์ระหว่างกลุ่มหลายรอบ และทำให้ผลมีค่าผิดพลาดได้ ซึ่งวิธีการวิเคราะห์ด้วย Chi square เป็นอีกวิธีที่เหมาะกับข้อมูลปริมาณมาก

ส่วนที่ต้องดำเนินการต่อจากการทดลองข้อมูลข้อมูลตัวอย่าง คือการสร้าง engine ของระบบที่สามารถทำการคัดเลือก GO term โดยไม่ต้องพึ่งระบบของ GO และส่วนของการวิเคราะห์ผลให้ทำงานอัตโนมัติ รวมทั้งกรณีที่พบว่า GO term ไปปรากฏในกลุ่มยีนข้ามกลุ่มกัน จำเป็นต้องพิสูจน์สมมุติฐานเพิ่มจากเดิมที่การจัดกลุ่มยีนกลุ่มเดียวกันมีหน้าที่คล้ายกัน มาเป็นว่ายีนคนละกลุ่มอาจมียีนร่วมที่มีความคล้าย ซึ่งเป็นงานที่จะทำให้โปรแกรมทำงานทำงานได้สมบูรณ์ขึ้น

เอกสารอ้างอิง

[1] The Gene Ontology Consortium. (2000). Gene Ontology: tool for the unification of biology. Nature Genet., 25, 25–29.

[2] SGD (Saccaromyces cerevisiae orbudding yeast): http://www.yeastgenome.org/help/yeastGeneNomenclature.shtml

[3] FlyBase (Drosophila): http://flybase.bio.indiana.edu/

[4] MGI (Mouse): http://www.informatics.jax.org/

[5] Ashburner M., Ball C., Blake J., Botstein D., Butler H., Cherry J., Davis A., Dolinski K., Dwight S., Eppig J., et al. (2000). Gene Ontology: tool for the unification of biology. *Nat Genet*. 25:25-29.

[6] Zhang B., Schmoyer D., Kirov S., Snoddy J. (2004). GOTree Machine (GOTM): a web-based platform for interpreting sets of interesting genes using Gene Ontology hierarchies. *Bioinformatics*, 5(1):16.

[7] Doniger S.W., Salomonis N., Dahlquist K.D., Vranizan K., Lawlor S.C., Conklin B.R. (2002). MAPPFinder: using Gene Ontology and GenMAPP to create a global gene-expression profile from microarray data. *Genome Biol*, 4:R7.

[8] Al-Shahrour F., Diaz-Uriarte R.and Dopazo J. (2004). FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes. *Bioinformatics*, 20, 578–580.

[9] Zhong S., Tian L., Li C., Storch F.K., and Wong W.H. (2004). Comparative Analysis of Gene Sets in the Gene Ontology Space under the Multiple Hypothesis Testing Framework. *Proc IEEE Computational Systems Bioinformatics*, 425-435.

[10] Blast: http://www.ncbi.nlm.nih.gov/Education/BLASTinfo/information3.html

[11] Khiripet N., Thammarongtham C. (2004). Discovery Common Genes Using Approximate Maximal Clique Detection. International Conference in Bioinformatics 2004, September 5-8, Auckland, New Zeland.

จุฑารัตน์ มณีวัฒนาพจนกษ์ จบการศึกษาระดับปริญญาโทสาขาเทคโนโลยีการจัดการระบบสารสนเทศ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยมหิดล ปัจจุบันทำงานในตำแหน่งผู้ช่วยนักวิจัยงานวิจัยเทคโนโลยีคลังข้อมูล ฝ่ายวิจัยและพัฒนาเทคโนโลยีคอมพิวเตอร์เพื่อการคำนวณ ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ

นพดล คีรีเพ็ชร สำเร็จการศึกษาระดับปริญญาตรีสาขาฟิสิกส์จากมหาวิทยาลัยสงขลานครินทร์ จังหวัดสงขลาในปี 2536 และระดับปริญญาเอกในสาขา Electrical and Computer Engineering จาก Georgia Institute of Technology ประเทศสหรัฐอเมริกาในปี 2544 ปัจจุบัน ดร. นพดล เป็นนักวิจัยสาขาชีวสารสนเทศศาสตร์ ศูนย์เทคโนโลยีอิเลคทรอนิกส์และคอมพิวเตอร์แห่งชาติ และมีความสนใจด้านการค้นหาความหมายของข้อมูลจีโนมเพื่องานประยุกต์ด้านการแพทย์และสาธารณสุข ด้วยเทคนิควิธีการทาง datamining, machine learning และ graph algorithms