



มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี
รายงานวิจัยฉบับสมบูรณ์
ทุนวิจัยหมวดเงินอุดหนุน (ว.1)
ประจำปีงบประมาณ 2557

การศึกษาหาความหลากหลายในระดับรหัสพันธุกรรมใน *Arthrospira* โดยวิธีการเปรียบเทียบจีโนม:
ก้าวแรกสู่การหาตำแหน่งจีโนมเครื่องหมายสำหรับการจัดจำแนกสายพันธุ์แบบบูรณาการ

คณะผู้วิจัย

นายวิรัช กิตติโชติรัตน์ หัวหน้าโครงการ
สังกัด สถาบันพัฒนาและฝึกอบรมโรงงานต้นแบบ
นางสาวพริดา พรหมมีเนตร ผู้ร่วมโครงการ
สังกัด หน่วยปฏิบัติการวิจัยและพัฒนาวิศวกรรมชีวเคมีและโรงงานต้นแบบ
นางสาววารุณี แก้วงาม ผู้ร่วมโครงการ
สังกัด สถาบันพัฒนาและฝึกอบรมโรงงานต้นแบบ
นายจิตติศักดิ์ เสนาจักร ผู้ร่วมโครงการ
สังกัด หน่วยปฏิบัติการวิจัยและพัฒนาวิศวกรรมชีวเคมีและโรงงานต้นแบบ
นางสุภาภรณ์ ชีวะธนรักษ์ ที่ปรึกษาโครงการ
สังกัด คณะทรัพยากรชีวภาพและเทคโนโลยี มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี

เดือน ธันวาคม พ.ศ. 2557

กิตติกรรมประกาศ

งานวิจัยนี้ได้รับการสนับสนุนจากทุนอุดหนุนโครงการวิจัยเงินงบประมาณประจำปี 2557
จากมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี ตั้งแต่เดือนตุลาคม 2556 ถึงเดือนกันยายน 2557

บทคัดย่อ

Arthrospira หรือสไปรูลินาเป็นสิ่งมีชีวิตในกลุ่มไซยาโนแบคทีเรียที่มีองค์ประกอบหลายอย่างที่เป็นประโยชน์ ไม่ว่าจะเป็นการมีปริมาณโปรตีนสูงมากถึงร้อยละ 70 ของน้ำหนักแห้งและเป็นแหล่งของสารมูลค่าสูงหลากหลายชนิด ด้วยเหตุนี้ *Arthrospira* จึงถูกนำไปใช้ประโยชน์อย่างกว้างขวางในหลากหลายอุตสาหกรรม โดยยังไม่มีการพบรายงานเกี่ยวกับผลกระทบในเชิงลบต่อผู้ใช้แต่อย่างใด ด้วยศักยภาพอันมากมายของ *Arthrospira* จึงทำให้เกิดความสนใจที่จะใช้ *Arthrospira* เป็นเซลล์เจ้าบ้านในการผลิตสารมูลค่าสูงต่างๆ และทำให้การหาสายพันธุ์ของ *Arthrospira* จากธรรมชาติที่มีความสามารถพิเศษในการสร้างหรือสะสมสารมูลค่าสูงนั้นเป็นที่ต้องการทั้งในหน่วยงานวิจัยและในภาคอุตสาหกรรม อย่างไรก็ตามการคัดเลือกสายพันธุ์ของ *Arthrospira* สายพันธุ์ต่างๆในระดับพันธุกรรมยังมีค่อนข้างน้อย จึงทำให้การแยกแยะสายพันธุ์ที่ได้จากธรรมชาติอย่างแม่นยำยังไม่สามารถทำได้อย่างมีประสิทธิภาพ ในงานวิจัยนี้ทางคณะวิจัยจึงใช้ประโยชน์ของข้อมูลจีโนมของ *Arthrospira* ที่มีอยู่ฐานข้อมูลจีโนมสาธารณะทั้งหมด ซึ่งประกอบด้วย *Arthrospira maxima* CS-328, *Arthrospira platensis* C1, *Arthrospira platensis* NIES-39, *Arthrospira platensis* str. Paraca และ *Arthrospira* sp. PCC 8005 โดยนำมาเปรียบเทียบแบบทั้งหมดเทียบกับทั้งหมดอย่างละเอียด เพื่อทำความเข้าใจถึงความหลากหลายในระดับพันธุกรรมที่พบใน *Arthrospira* สายพันธุ์ต่างๆ ผลของการเปรียบเทียบแสดงให้เห็นว่า ในจำนวนยีนทั้งหมดที่พบในจีโนมของ *Arthrospira* นั้นจะมีประมาณ 85-90% ที่เป็นยีนที่พบได้ในทุกสายพันธุ์ และที่เหลืออีก 10-15% จะเป็นยีนที่มีการพบเพียงบางสายพันธุ์ นอกจากนั้นยังมียีนจำนวน 0-3% ที่เป็นยีนที่พบเฉพาะในจีโนมอื่นๆ ซึ่งยีนเหล่านี้จะเป็นยีนจำเพาะที่เกี่ยวข้องโดยตรงกับคุณลักษณะพิเศษของสายพันธุ์ต่างๆ ทั้งนี้ทั้งนั้นผลของการวิเคราะห์ยังแสดงให้เห็นว่าไม่มีการพบยีนที่จำเพาะกับ *Arthrospira platensis* str. Paraca แต่อย่างไรก็ดีผลดังกล่าวไม่ได้หมายความว่า *Arthrospira platensis* str. Paraca ไม่มีลักษณะพิเศษ เพียงแต่ข้อมูลจีโนมของ *Arthrospira* สายพันธุ์นี้อาจจะมาพร้อมกับข้อมูลตำแหน่งยีนที่ไม่สมบูรณ์ ซึ่งมีความเป็นไปได้สูงเนื่องจากจำนวนยีนที่ถูกกำหนดมานั้นมีค่อนข้างน้อยเมื่อเปรียบเทียบกับขนาดของจีโนม นอกเหนือจากยีนจำเพาะแล้วนั้นข้อมูลลำดับเบสของยีนที่พบในทุกๆจีโนมยังถูกนำมาเปรียบเทียบผ่านการทำ Multiple sequence alignment เพื่อแสดงถึงตำแหน่งเบสภายในยีนดังกล่าวที่มีความแตกต่างระหว่างสายพันธุ์ที่ต่างกัน โดยข้อมูลความหลากหลายในระดับรหัสพันธุกรรมทั้งในระดับยีนและในระดับเบสที่ได้จากงานวิจัยนี้จะเป็นข้อมูลที่มีประโยชน์สำหรับการค้นหาตำแหน่งจีโนมเครื่องหมายสำหรับการจัดจำแนกสายพันธุ์ต่างๆของ *Arthrospira* ได้ โดยผู้ที่สนใจสามารถเข้าถึงข้อมูลดังกล่าวได้ผ่านทาง http://www1.sbi.kmutt.ac.th/~weerayuth/w1_57/

Abstract

Arthrospira or *Spirulina* is a group of living organisms, which are classified as cyanobacteria, that have great potential in many applications. They are known to have high protein content (up to 70% of their dry weight) and can accumulate a variety of high value products. For these reasons, many industries and research laboratories have used *Arthrospira* as raw materials in many products without any report on their adverse side effects. In addition, with *Arthrospira*'s great potentials and safeness, there is an increasing interest in using *Arthrospira* as host cells for producing many valuable molecules. This has resulted in an increasing demand for identification of *Arthrospira* from nature that has specific ability to accumulate certain products of interest. However, there is a limited knowledge on genomics variability among different species/strains of various *Arthrospira*, which hinders an accurate characterization or classification of new species/strains. In this research project, we take advantage of the *Arthrospira* whole genome sequence data that are available on the public database, which consists of *Arthrospira maxima* CS-328, *Arthrospira platensis* C1, *Arthrospira platensis* NIES-39, *Arthrospira platensis* str. Paraca and *Arthrospira* sp. PCC 8005, and try to characterize that variations at genetic level. Result shows that among all genes that are found *Arthrospira* genomes, about 85-90% are considered "core gene" as they are found in all 5 genomes and about 10-15% are missing in one of the genome. We also found that about 0-3% of the genes are genome specific. These genome specific genes may be directly linked with the special characteristics of each species/strain. Our result also shows that *Arthrospira platensis* str. Paraca does not possess any genome specific genes. However, we think that this is an artifact that has resulted from the fact that the gene prediction data that comes with this genome sequence data is not complete. Specifically, we found that the number of genes that were assigned by the group who release the genome sequence data of *Arthrospira platensis* str. Paraca is much lower than it should based on its genome size. Beside genome-specific genes, DNA sequences for genes that are found in multiple genomes were analyzed using multiple sequence alignment to identify variation at single nucleotide level. Genomics variations among 5 *Arthrospira* species/strains that were characterized in this research project will be useful for finding genomic markers for distinguishing various closely related species/strains of *Arthrospira*. All data can be access via our online tool at http://www1.sbi.kmutt.ac.th/~weerayuth/w1_57/

สารบัญเรื่อง

	หน้า
บทคัดย่อ	3
สารบัญเรื่อง	5
สารบัญตาราง	6
สารบัญภาพ	7
บทที่ 1 บทนำ	8
ความสำคัญและที่มาของปัญหาที่ทำวิจัย	8
วัตถุประสงค์ของโครงการวิจัย	9
ขอบเขตของโครงการวิจัย	9
ทฤษฎีและกรอบแนวคิดของโครงการวิจัย	9
การทบทวนวรรณกรรม/สารสนเทศที่เกี่ยวข้อง	10
ประโยชน์ที่คาดว่าจะได้รับ	12
บทที่ 2 วิธีการดำเนินการวิจัย	13
บทที่ 3 ผลการวิจัยและวิจารณ์	17
บทที่ 4 สรุปผลการดำเนินงานวิจัยและข้อเสนอแนะ	24
เอกสารอ้างอิงของโครงการวิจัย	25

สารบัญตาราง

	หน้า
ตารางที่ 1 ข้อมูลจีโนมของ <i>Arthrospira</i> สายพันธุ์ต่างๆ	11
ตารางที่ 2 ตัวอย่างฐานข้อมูลจีโนมสาธารณะที่สามารถเข้าไปดาวโหลดข้อมูลจีโนมได้	13
ตารางที่ 3 ตัวอย่างรูปแบบไฟล์จีโนมที่ได้จากการดาวโหลดข้อมูลจีโนมจากฐานข้อมูลสาธารณะ	14
ตารางที่ 4 ตัวอย่างโปรแกรมสำเร็จรูปสำหรับเปรียบเทียบข้อมูลจีโนม	15
ตารางที่ 5 ข้อมูลจีโนมของ <i>Arthrospira</i> ที่ได้จากฐานข้อมูล NCBI Genome Database	17
ตารางที่ 6 สรุปข้อมูลยีนส์ที่พบในจีโนมของ <i>Arthrospira</i> ที่รวบรวมได้	18
ตารางที่ 7 จำนวนยีนส์ที่จำเพาะกับแต่ละจีโนม	19

สารบัญภาพ

	หน้า
รูปที่ 1 A) <i>Arthrospira</i> ที่พบในธรรมชาติ B) เซลของ <i>Arthrospira</i> ในระดับ Microscopic C) ตัวอย่างอาหารเสริม ที่ผลิตจาก <i>Arthrospira</i>	10
รูปที่ 2 การเติบโตของโปรเจ็คอ่านลำดับนิวคลีโอไทด์ในจีโนมสิ่งมีชีวิต ใน GOLD Genome database	11
รูปที่ 3 วิธีการดำเนินการวิจัยเพื่อหาความแตกต่างในระดับจีโนมระหว่าง <i>Arthrospira</i> สายพันธุ์ต่างๆ	13
รูปที่ 4 ตัวอย่าง Entity-relationship model ของระบบฐานข้อมูลเพื่อจัดเก็บข้อมูลจีโนม	15
รูปที่ 5 ผลของการจัดกลุ่มยีนส์ทั้งหมด 30,038 ยีนส์ที่พบในจีโนมของ <i>Arthrospira</i> ทั้ง 5 สายพันธุ์ แสดงให้เห็นว่ายีนส์ทั้งหมดสามารถแบ่งออกได้เป็น 6,782 กลุ่มยีนส์ โดยมี 5,359 (79%) กลุ่มเป็นยีนส์ที่พบในทุกๆ จีโนม และ 1,423 กลุ่มเป็นยีนส์ที่พบเฉพาะในบางจีโนม	19
รูปที่ 6 การแบ่งกลุ่มของยีนส์ที่พบเฉพาะในบางจีโนม ตามจำนวนจีโนมที่มีการพบยีนส์ดังกล่าว	19
รูปที่ 7 การแบ่งแยกยีนส์ที่พบในแต่ละจีโนมของ <i>Arthrospira</i> ตามจำนวนจีโนมที่มีการพบยีนส์ดังกล่าว	20
รูปที่ 8 หน้าเริ่มต้นของระบบเว็บแอพลิเคชันสำหรับค้นหายีนส์ที่จำเพาะกับ <i>Arthrospira</i> จำนวน 5 สายพันธุ์ http://www1.sbi.kmutt.ac.th/~weerayuth/w1_57/	21
รูปที่ 9 หน้าตารางแสดงยีนส์และการพบหรือไม่พบในจีโนมต่างๆ	22
รูปที่ 10 กล่องแสดงข้อมูลยีนส์ในจีโนมที่ผู้ใช้สามารถเรียกดูได้โดยการกดที่ช่องมีปรากฏการมีของยีนส์	22
รูปที่ 11 ตารางแสดงรายชื่อยีนส์ที่พบในทุกๆจีโนม	23
รูปที่ 12 กล่องแสดงข้อมูลเพิ่มเติมของยีนส์ที่พบในทุกๆจีโนม	23
รูปที่ 13 หน้าแสดง Multiple sequence alignment ของยีนส์เดียวกันที่พบในจีโนมของ <i>Arthrospira</i> ทั้ง 5 สายพันธุ์	23

บทที่ 1 บทนำ

ความสำคัญและที่มาของปัญหาที่ทำวิจัย

Arthrospira หรือสไปรูลิน่าเป็นสาหร่ายสีเขียวแกมน้ำเงิน (จัดอยู่ในกลุ่มของไซยาโนแบคทีเรีย) ที่มีความสามารถในการสังเคราะห์แสงเช่นเดียวกับพืช [1] และสามารถพบได้ทั่วโลกทั้งในน้ำจืดและน้ำเค็ม *Arthrospira* มีองค์ประกอบหลายอย่างที่เป็นประโยชน์คือ มีปริมาณโปรตีนสูงมากถึงร้อยละ 70 ของน้ำหนักแห้ง [2] ทั้งยังเป็นแหล่งของสารมูลค่าสูงหลากหลายชนิด [3] จึงทำให้ *Arthrospira* ถูกนำไปใช้ประโยชน์อย่างกว้างขวางในหลากหลายอุตสาหกรรม ไม่ว่าจะเป็นในอุตสาหกรรมอาหารเสริมเพื่อสุขภาพ [4] ตลอดจนอาหารสัตว์ [2] หรือแม้แต่ด้านเภสัชกรรม [5] โดยยังไม่เคยมีการพบรายงานเกี่ยวกับผลกระทบในเชิงลบ (negative side-effect) ต่อผู้ใช้แต่อย่างใด [6] สารสกัดหลายชนิดจาก *Arthrospira* ยังมีประโยชน์ทางการแพทย์ เช่น สามารถเสริมภูมิคุ้มกันให้กับร่างกาย (Immunomodulation) [7], ช่วยต้านไวรัส [8], ช่วยลดโคเลสเตอรอลในกระแสเลือด [9], ต้านมะเร็ง [10], และบรรเทาอาการเป็นพิษจากสารหนู (Arsenic poisoning) [11] เป็นต้น นอกจากนี้ประโยชน์เหล่านี้แล้ว *Arthrospira* ยังสามารถถูกนำมาเพาะเลี้ยงกลางแจ้งเพื่อดูดซับ ก๊าซคาร์บอนไดออกไซด์ในอากาศผ่านกระบวนการสังเคราะห์แสงเพื่อช่วยลดภาวะโลกร้อน [12] รวมไปถึงการใช้ *Arthrospira* เป็นแหล่งผลิตออกซิเจนและอาหารให้กับนักบินในอวกาศของโครงการ MELISSA biological life support [13]

ด้วยความสามารถอันมากมายของ *Arthrospira* จึงทำให้หน่วยวิจัยต่างๆ ทั่วโลกได้ให้ความสนใจที่จะใช้ *Arthrospira* เพื่อเป็นเซลล์เจ้าบ้านในการผลิตสารต่างๆ และทำให้การหาสายพันธุ์ของ *Arthrospira* จากธรรมชาติที่มีความสามารถพิเศษในการสร้างหรือสะสมสารมูลค่าสูงนั้นเป็นที่ต้องการทั้งในหน่วยงานวิจัยและในภาคอุตสาหกรรม [14] วิธีการดั้งเดิมที่สามารถใช้ในการจำแนกสายพันธุ์ของจุลินทรีย์ใน ห้องปฏิบัติการจะทำได้โดยการเพาะเลี้ยงเชื้อที่เก็บสะสมได้จากแหล่งที่อยู่ตามธรรมชาติบนอาหารชนิดต่างๆ (culture media) และดูความแตกต่างผ่านลักษณะการแสดงออก (phenotype) หรือนำมาทดสอบความสามารถโดยใช้ชุดทดสอบทางชีวเคมี (Biochemical test) ซึ่งวิธีดังกล่าวนี้เป็นวิธีที่ต้องใช้เวลาและถึงจำเป็นต้องลงแรงและเสียค่าใช้จ่ายที่มากกว่านั้น วิธีดั้งเดิมยังไม่มี ความละเอียดเพียงพอที่จะจำแนกหรือแยกแยะสายพันธุ์ในระดับ strain ได้ [15] ซึ่งในปัจจุบันมีการยอมรับกัน อย่างแพร่หลายว่าจุลินทรีย์ชนิดเดียวกันนั้น (Same species) ก็สามารถที่จะมีความแตกต่างในเชิงชีววิทยาได้กันได้ (Intra species variation) [16]

อย่างไรก็ดีเนื่องจากปัจจุบันกลุ่มวิจัยภายใต้ความร่วมมือระหว่างคณะนักวิจัยของมหาวิทยาลัยเทคโนโลยี พระจอมเกล้าธนบุรี, มหาวิทยาลัยเชียงใหม่, Kazusa DNA Research Institute (ประเทศญี่ปุ่น) และศูนย์พันธุ วิศวกรรมและเทคโนโลยีแห่งชาติได้ทำการศึกษาหาลำดับเบสบนจีโนมของ *Arthrospira platensis* C1 เสร็จแล้ว [17] ดังนั้นจึงสามารถใช้ข้อมูลลำดับเบสบนจีโนมที่มีอยู่ในการค้นหาความหลากหลายในรหัสพันธุกรรมผ่านการ เปรียบเทียบกับข้อมูลจีโนมของ *Arthrospira* สายพันธุ์ต่างๆ ที่สามารถหาได้จากฐานข้อมูลสาธารณะเพื่อค้นหา ตำแหน่งจีโนมเครื่องหมาย (genomic marker) ที่สามารถใช้ในการจัดจำแนกสายพันธุ์ โดยใช้เทคนิคทางด้าน ชีวสารสนเทศ (Bioinformatics) ได้ ซึ่งผลที่ได้จะเป็นข้อมูลสำคัญที่จะช่วยให้นักวิจัยสามารถคัดเลือกตำแหน่ง จีโนมเครื่องหมายเพื่อใช้สำหรับการออกแบบชุดจัดจำแนกสายพันธุ์ในระดับรหัสพันธุกรรม (Genotyping) ซึ่งจะมี ความละเอียดมากกว่าวิธีแบบดั้งเดิม และช่วยลดขั้นตอนในการทำงานรวมถึงระยะเวลาและค่าใช้จ่ายในห้อง ปฏิบัติการได้ [18]

ในการทดลองนี้จะใช้ข้อมูลลำดับเบสบนจีโนมของ *Arthrospira platensis* C1 เพื่อเปรียบเทียบกับข้อมูล จีโนมของ *Arthrospira* สายพันธุ์อื่นๆ ที่มีอยู่ในฐานข้อมูลสาธารณะ เช่น NCBI Genome Database (<http://www.ncbi.nlm.nih.gov/genome/browse/>) ซึ่งเป็นฐานข้อมูลที่ ทำหน้าที่เก็บรวบรวมข้อมูลจีโนมของ สิ่งมีชีวิตต่างๆ โดยสำหรับกลุ่มแบคทีเรียนั้นในปัจจุบันมีข้อมูลจีโนมมากกว่า 14,000 สายพันธุ์ (ข้อมูลเดือน กันยายน 2555) และในจำนวนนี้เป็นข้อมูลของ *Arthrospira* สายพันธุ์ต่างๆ อยู่จำนวนหนึ่ง ในส่วนของการ วิเคราะห์เปรียบเทียบข้อมูลจีโนม (Comparative genomics) ด้วยคอมพิวเตอร์ โดยใช้เทคนิคทางด้านชีวสารสนเทศ

เทศ (Bioinformatics) จะทำโดยใช้โปรแกรมเปรียบเทียบลำดับเบสต่างๆ (Sequence similarity search) และจะเปรียบเทียบข้อมูลจีโนมของ *Arthrospira* ทั้งหมดที่รวบรวมได้แบบทั้งหมดเทียบกับทั้งหมด (All against all) เพื่อแยกแยะลำดับเบสบนจีโนมเป็นสองกลุ่มหลัก คือ บริเวณที่มีการอนุรักษ์ไว้ (conserved region) และบริเวณที่มีความจำเพาะเจาะจงกับแต่ละสายพันธุ์ (species/ strain specific region) บริเวณของจีโนมที่มีการอนุรักษ์ไว้จะมีการวิเคราะห์เพื่อหาความแตกต่างในระดับสปีชีส์ หรือ Single nucleotide polymorphism (SNP) ผ่านการเขียนโปรแกรมคอมพิวเตอร์ที่จะถูกสร้างขึ้นใหม่ในงานวิจัยนี้ ข้อมูลความแตกต่างของจีโนมในระดับสปีชีส์ (SNP) และบริเวณที่มีความจำเพาะเจาะจงกับแต่ละสายพันธุ์ (species/strain specific region) จะถูกจัดเก็บเป็นฐานข้อมูลเพื่อให้นักวิจัยหรือผู้สนใจใช้เป็นข้อมูลพื้นฐานในการคัดเลือกตำแหน่งจีโนมเครื่องหมายเพื่อพัฒนาชุดจัดจำแนกสายพันธุ์โดยใช้รหัสพันธุกรรมต่อไป

ซึ่งคณะทำงานของงานวิจัยนี้หวังว่าการค้นหาและศึกษาการความหลากหลายในระดับเบสบนจีโนมของ *Arthrospira* สายพันธุ์ต่างๆนั้นจะช่วยทำให้การออกแบบชุดจัดจำแนกสายพันธุ์ในระดับรหัสพันธุกรรมสำหรับ *Arthrospira* จะสามารถทำได้ง่ายขึ้น ซึ่งจะเป็นประโยชน์ต่อการนำไปใช้ในการคัดเลือกและจำแนกสายพันธุ์ที่เก็บรวบรวมได้จากแหล่งธรรมชาติต่างๆ และช่วยพัฒนาความสามารถในการใช้ *Arthrospira* เพื่อเป็นเซลล์เจ้าบ้านในการผลิตสารมูลค่าสูงต่างๆต่อไปในอนาคต

วัตถุประสงค์ของโครงการวิจัย

เพื่อรวบรวมข้อมูลจีโนมของ *Arthrospira* และวิเคราะห์หาความแตกต่างในระดับรหัสพันธุกรรมโดยความแตกต่างในระดับรหัสพันธุกรรมระหว่าง *Arthrospira* สายพันธุ์ต่างๆนี้จะสามารถนำมาใช้เพื่อหาตำแหน่งจีโนมเครื่องหมายเพื่อการจัดจำแนกสายพันธุ์ของ *Arthrospira* และพัฒนาความสามารถในการใช้ *Arthrospira* เพื่อเป็นเซลล์เจ้าบ้านในการผลิตสารมูลค่าสูงต่างๆต่อไป

ขอบเขตของโครงการวิจัย

การรวบรวมข้อมูลจีโนมของ *Arthrospira* จะทำโดยการค้นหาและดาวน์โหลดจากฐานข้อมูลจีโนมสาธารณะต่างๆ โดยข้อมูลจีโนมที่ได้มาจะมีการจัดเก็บในระบบฐานข้อมูลภายใน รวมถึงการตรวจสอบและเพิ่มเติมข้อมูลจีโนมใหม่ๆที่อาจจะพบเพิ่มเติมในอนาคต ความแตกต่างในระดับรหัสพันธุกรรม ระหว่าง *Arthrospira* สายพันธุ์ต่างๆจะทำได้โดยใช้โปรแกรมเปรียบเทียบลำดับเบสต่างๆ (Sequence similarity search tools) จากนั้นข้อมูลความแตกต่างในระดับรหัสพันธุกรรมทั้งหมดจะถูกจัดเก็บในระบบฐานข้อมูลและนำเสนอให้นักวิจัยและผู้สนใจผ่านระบบเว็บแอปพลิเคชันต่อไป

ทฤษฎีและกรอบแนวความคิดของโครงการวิจัย

การจัดจำแนกสายพันธุ์ของสิ่งมีชีวิตนั้นเป็นเครื่องมือที่สามารถนำมาใช้ในงานวิจัยทางด้านต่างๆได้ เช่น การคัดเลือกและปรับปรุงสายพันธุ์, การศึกษาโรคระบาดวิทยาเพื่อหาสายพันธุ์ต้น กำเนิดและศึกษาการพัฒนาของเชื้อโรครวมไปถึงการศึกษาวิวัฒนาการของสิ่งมีชีวิตเป็นต้น การจัดจำแนกสายพันธุ์นั้นสามารถทำได้ในห้องปฏิบัติการโดยการเพาะเลี้ยงเชื้อสายพันธุ์ต่างๆที่สนใจบนอาหารเลี้ยง เชื้อหลากหลายชนิดเพื่อดูความแตกต่างผ่านลักษณะการแสดงออกที่ปรากฏให้เห็น (Phenotype) เช่น รูปร่างและสีของ Colonies หรือนำมาทดสอบความสามารถทางชีววิทยาโดยใช้ชุดทดสอบทางชีวเคมี (Biochemical test) แต่อย่างไรก็ดีวิธีดังกล่าวนี้มีข้อด้อย ตรงที่ต้องใช้เวลานานและต้องใช้ค่าใช้จ่ายมากในการเตรียมวัสดุต่างๆ นอกจากนั้นสิ่งมีชีวิตตามธรรมชาติหลายชนิดก็ยังไม่สามารถนำมาเพาะเลี้ยงในห้องปฏิบัติการได้ และวิธีดังกล่าวก็ยังไม่มีความละเอียดเพียงพอที่จะนำไปจำแนกหรือแยกแยะสายพันธุ์ในระดับ Strain ของสิ่งมีชีวิตชนิดเดียวกันได้

ในปัจจุบันมีสิ่งมีชีวิตหลากหลายชนิด (โดยเฉพาะในกลุ่มของแบคทีเรีย) ที่นักวิจัยได้ทำการอ่านลำดับเบสบนจีโนมแล้ว ประกอบกับการเติบโตอย่างรวดเร็วของปริมาณข้อมูลจีโนมในฐานข้อมูลจีโนมต่างๆ ทำ

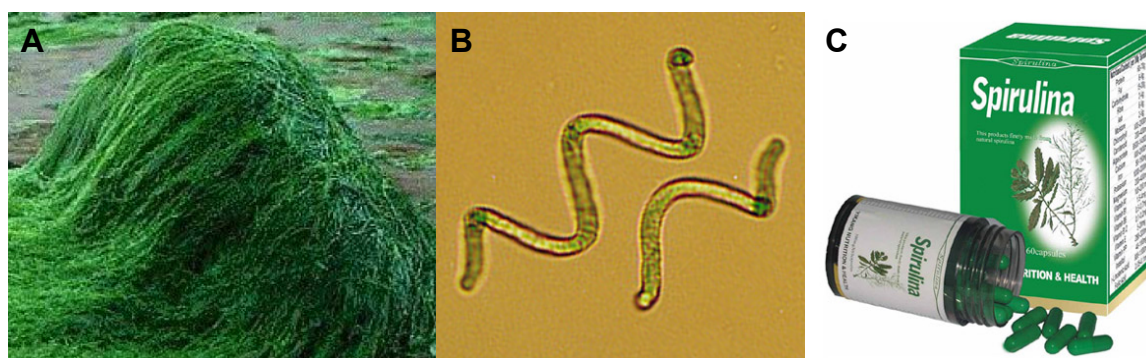
ให้มีความสนใจที่จะใช้ข้อมูลรหัสพันธุกรรมในการจัดจำแนกสายพันธุ์ของสิ่งมีชีวิตชนิดต่างๆ เพราะการจำแนกโดยใช้ความแตกต่างในระดับรหัสพันธุกรรมนั้นสามารถทำได้ง่ายและรวดเร็วกว่าวิธีดั้งเดิมที่กล่าวไปในข้างต้น ตลอดจนยังสามารถแยกแยะสายพันธุ์ต่างๆ ของสิ่งมีชีวิตชนิดเดียวกันได้อีกด้วย [18]

ถึงแม้จะเป็นที่ทราบกันอย่างแพร่หลายว่า *Arthrospira* นั้นมีองค์ประกอบที่เป็นประโยชน์อยู่เป็นจำนวนมากและยังมี *Arthrospira* อยู่หลากหลายสายพันธุ์ที่หน่วยงานวิจัยต่างๆ ทั่วโลกได้นำไปอ่านลำดับเบสบนจีโนมและเผยแพร่ข้อมูลจีโนมแล้ว แต่ก็ยังไม่มีการใช้ประโยชน์จากข้อมูลจีโนมเหล่านี้เพื่อศึกษาหาความแตกต่างในระดับรหัสพันธุกรรมและใช้เป็นข้อมูลในการหาตำแหน่งบนจีโนมที่เหมาะสมในการนำไปพัฒนาเป็นชุดจัดจำแนกสายพันธุ์ของ *Arthrospira* ที่สามารถพบได้ในแหล่งที่อยู่ตามธรรมชาติต่างๆ ดังนั้นงานวิจัยนี้จึงเล็งเห็นผลประโยชน์อันมาก มายที่จะได้รับการวิเคราะห์ข้อมูลจีโนมผ่านการเปรียบเทียบลำดับเบสบนจีโนมของ *Arthrospira* สายพันธุ์ต่างๆ โดยใช้วิธีการทางชีวสารสนเทศเพื่อหาความแตกต่างในระดับพันธุกรรม เพราะข้อมูลลักษณะนี้จะเป็นพื้นฐานในการศึกษาและช่วยในการจัดจำแนกสายพันธุ์ ของ *Arthrospira* ซึ่งจะนำไปสู่การพัฒนาความสามารถในการคัดเลือกสายพันธุ์ของ *Arthrospira* อย่างมีประสิทธิภาพในอนาคตได้

การทบทวนวรรณกรรม/สารสนเทศที่เกี่ยวข้อง

Arthrospira และประโยชน์

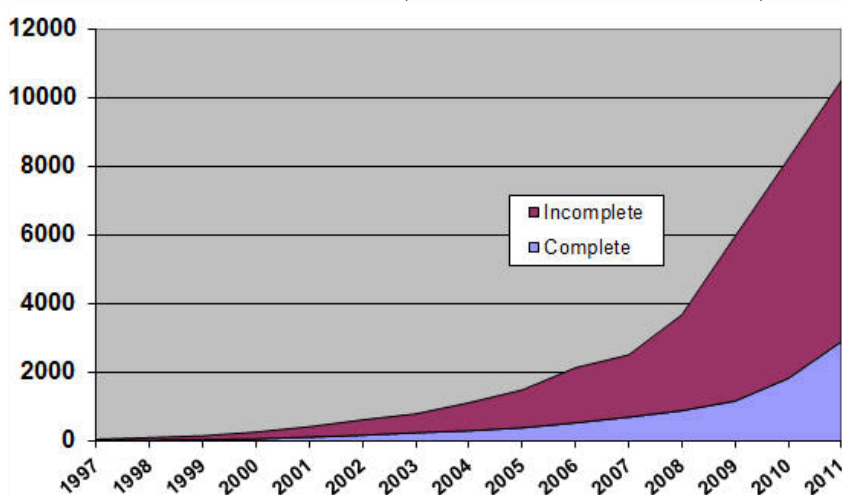
Arthrospira เป็นสาหร่ายสีเขียวแกมน้ำเงิน (Blue-Green Algae) ที่จัดอยู่ใน Division Cyanophyta (รูปที่ 1A) เป็นจุลินทรีย์ชนิด Prokaryote ที่มีร่างเซลล์เป็นสายยาวขดวนเป็นเกลียว ความยาวประมาณ 30-500 ไมโครเมตร (รูปที่ 1B) [1] *Arthrospira* มีความสามารถในการสังเคราะห์แสงได้เหมือนพืชเนื่องจากมีโครงสร้างที่คล้าย Chloroplast นอกจากนี้ *Arthrospira* ยังถูก นำไปใช้อย่างกว้างขวางในภาคธุรกิจ [3] โดยเฉพาะในอุตสาหกรรมอาหารเสริมสุขภาพตลอดจน อุตสาหกรรมทางด้านเภสัชกรรมและเครื่องสำอางค์ (รูปที่ 1C) ทั้งนี้เนื่องจากใน *Arthrospira* มีปริมาณโปรตีนสูงมากกว่าร้อยละ 70 ของน้ำหนักแห้ง ซึ่งมากกว่าสาหร่ายชนิดอื่นๆ และยังมีสารที่สำคัญอีกหลากหลายชนิด เช่น กรดไขมันจำเป็นแกมมาลิโนเลนิก (Gamma Linolenic Acid: GLA) ที่สามารถช่วยลดคอเลสเตอรอลในเลือดซึ่งเป็นสาเหตุหนึ่งของโรคหัวใจ [4], เบต้า-แคโรทีน ซึ่งเป็นสารต่อต้านมะเร็ง [19], ไฟโคไซยานิน (Phycocyanin) ที่มีบทบาทในการเพิ่มประสิทธิภาพของระบบภูมิคุ้มกัน [20] ตลอดจนคลอโรฟิลล์, เกลือแร่, และวิตามินต่างๆ [5] เป็นต้น นอกจากนี้ *Arthrospira* ยังมีความสำคัญในอุตสาหกรรมอาหาร สัตว์ [2] โดยถูกใช้เป็นส่วนประกอบอาหารเพื่อ เพิ่มการเจริญเติบโตและอัตราการรอดชีวิตของสัตว์วัยอ่อนตลอดจนช่วยเพิ่มสีส้มสำหรับปลาสวยงามและเพิ่มภูมิคุ้มกัน เป็นต้น



รูปที่ 1 A) *Arthrospira* ที่พบในธรรมชาติ B) เซลล์ของ *Arthrospira* ในระดับ Microscopic C) ตัวอย่างอาหารเสริม ที่ผลิตจาก *Arthrospira*

ข้อมูลจีโนมของ *Arthrospira*

ปัจจุบันความเจริญก้าวหน้าทางเทคโนโลยีสำหรับอ่านลำดับนิวคลีโอไทด์ในจีโนมของสิ่งมีชีวิต (Genome sequencing technology) ได้ทำให้ปริมาณข้อมูลจีโนมในฐานะข้อมูลต่างๆมีการเติบโตอย่างก้าวกระโดดโดยเฉพาะใน 8 ปีที่ผ่านมา (หลังปี ค.ศ. 2006) [21] ดังที่แสดงอยู่ในรูปที่ 2 ในที่นี้นอกจากข้อมูลจีโนมของ *Arthrospira platensis* C1 ที่ทางกลุ่มวิจัยภายใต้ความร่วมมือระหว่างคณะนักวิจัยของมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี, มหาวิทยาลัยเชียงใหม่, Kazusa DNA Research Institute (ประเทศญี่ปุ่น) และศูนย์พันธุวิศวกรรมและเทคโนโลยีแห่งชาติได้ทำการอ่านลำดับนิวคลีโอไทด์แล้ว ยังมีจีโนมของ *Arthrospira* สายพันธุ์อื่นๆที่มีความสำคัญถูกอ่านลำดับนิวคลีโอไทด์แล้ว ทั้งที่กำลังอ่านอยู่และอ่านเสร็จสิ้นเรียบร้อยแล้ว เช่นกัน (ดังที่แสดงอยู่ในตารางที่ 1) ในจำนวนนี้พบว่าข้อมูลจีโนมของ *Arthrospira* ที่ถูกบรรจุลงไปในฐานข้อมูลและเปิดเผยต่อสาธารณชนแล้วทั้งหมด 4 สายพันธุ์ ซึ่งประกอบด้วย *Arthrospira* sp. PCC 8005, *Arthrospira platensis* Paraca, *Arthrospira platensis* NIES-39 และ *Arthrospira maxima* CS-328 (กันยายน 2555) และเมื่อรวมกับข้อมูลจีโนมของสายพันธุ์ *Arthrospira platensis* C1 ทำให้มีข้อมูลจีโนมของ *Arthrospira* ถึง 5 สายพันธุ์ซึ่งจะเป็นข้อมูลอันล้ำค่าสำหรับการวิเคราะห์หาความแตกต่างในระดับรหัสพันธุกรรมระหว่าง *Arthrospira* สายพันธุ์เหล่านี้ต่อไป



รูปที่ 2 การเติบโตของโปรเจกต์อ่านลำดับนิวคลีโอไทด์ในจีโนมสิ่งมีชีวิต ใน GOLD Genome database

ตารางที่ 1 ข้อมูลจีโนมของ *Arthrospira* สายพันธุ์ต่างๆ

<i>Arthrospira</i> Genomes	Size (kb)	Project Status	Genbank accession	Sequencing Organization
<i>Arthrospira</i> sp. PCC 8005	6145	Draft	ADDH00000000	Belgian Nuclear Research Center, Belgium
<i>Arthrospira platensis</i> Paraca	4997	Draft	ACSK00000000	University of Applied Sciences of Western Switzerland, Switzerland
<i>Arthrospira platensis</i> NIES-39	6692	Complete	AP011615	Institute of Technology and Evaluation, NITE, Japan
<i>Arthrospira platensis</i> C1	6089	Draft	PRJNA67617	King Mongkut's University of Technology Thonburi, Thailand
<i>Arthrospira platensis</i> AGB-AP02	NA	In progress	NA	Ocean University of China, China
<i>Arthrospira platensis</i> ABT	NA	In progress	NA	Beijing Genome Institute, China
<i>Arthrospira platensis</i>	NA	In progress	NA	Human Genome Center, China
<i>Arthrospira maxima</i> CS-328	6003	Draft	ABYK00000000	US DOE Joint Genome Institute, USA

โปรแกรมสำหรับเปรียบเทียบข้อมูลจีโนม

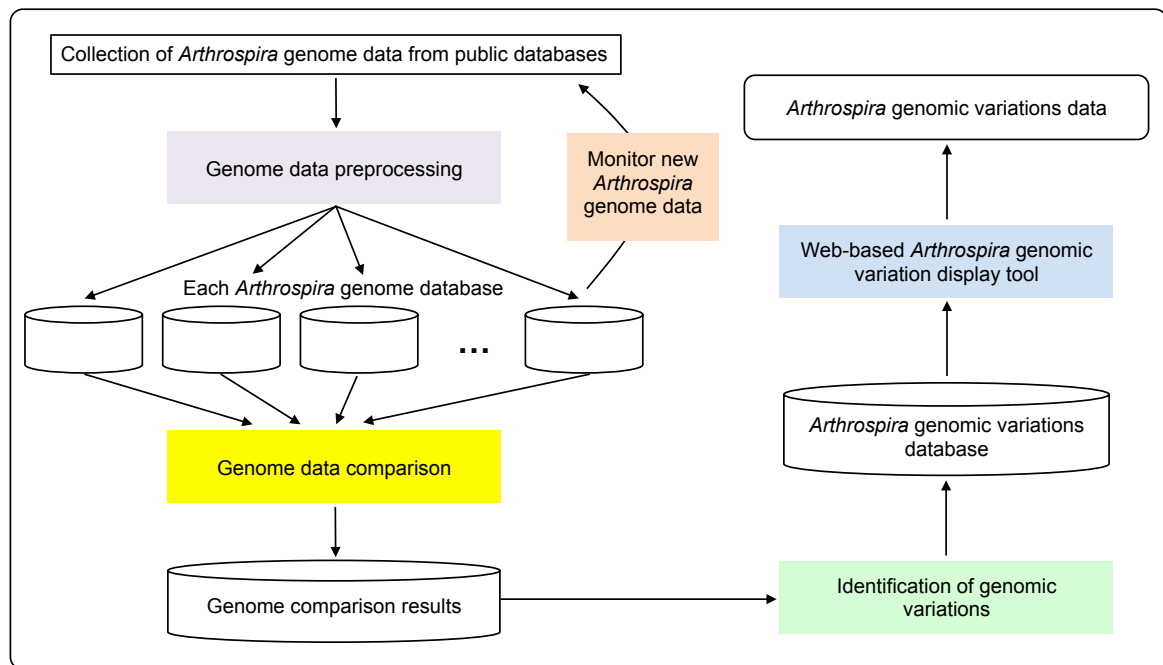
การเปรียบเทียบลำดับเบสจีโนมนั้นสามารถทำได้โดยการใช้โปรแกรมเปรียบเทียบลำดับเบสต่างๆ (Sequence similarity search tool) เช่น NCBI Blast+ [22], MAUVE [23] หรือ Clustal Omega [24] เป็นต้น โดยโปรแกรมต่างๆเหล่านี้มักจะถูกออกแบบให้ทำงานในรูปแบบของ Command-line และจะใช้เทคนิคในการเปรียบเทียบข้อมูลลำดับเบสที่ต่างกันไป ยกตัวอย่าง เช่น NCBI Blast+ จะทำการเปรียบเทียบข้อมูลลำดับเบสโดยใช้ Heuristic search และจะเป็นการเปรียบเทียบแบบ Local alignment algorithm เพื่อสร้างเป็น Pairwise alignment [22] ส่วน MAUVE และ Clustal Omega นั้นจะเปรียบเทียบข้อมูลลำดับเบสแบบกลุ่ม (Multiple sequence alignment) โดย MAUVE จะใช้วิธีที่เรียกว่า Recursive anchor search [23] และ Clustal Omega จะใช้วิธี Hidden Markov Model [24] แต่ละโปรแกรมนั้นจะมีข้อกำหนดเกี่ยวกับรูปแบบไฟล์ที่จะใช้เป็น Input และลักษณะไฟล์ที่จะเป็น Output ที่แตกต่างกัน ผู้ใช้จึงจำเป็นต้องทำความเข้าใจเกี่ยวกับรูปแบบไฟล์ต่างๆ เพื่อที่จะสามารถใช้โปรแกรมเหล่านี้ได้อย่างมีประสิทธิภาพ

ประโยชน์ที่คาดว่าจะได้รับ

จะได้เผยแพร่ผลงานวิจัยในรูปแบบสิ่งตีพิมพ์ในวารสารและงานประชุมวิชาการ และได้วิธีการทางชีวสารสนเทศสำหรับวิเคราะห์หาความแตกต่างระหว่างจีโนม รวมถึงระบบเว็บแอปพลิเคชันที่สามารถแสดงข้อมูลความแตกต่างที่พบในจีโนมของ *Arthrospira* สายพันธุ์ต่างๆได้ ซึ่งจะเป็นประโยชน์แก่นักวิจัยที่จะนำข้อมูลไปใช้ในการออกแบบวิธีการจัดจำแนกสายพันธุ์ของ *Arthrospira* ต่อไป

บทที่ 2 วิธีการดำเนินการวิจัย

วิธีการวิเคราะห์ข้อมูลจีโนมของ *Arthrospira* หลากหลายพันธุ์ เพื่อหาความแตกต่างบนลำดับนิวคลีโอไทด์ของแต่ละจีโนมสามารถแบ่งได้เป็น 5 ขั้นตอนหลักซึ่งประกอบด้วย 1) การเก็บรวบรวมข้อมูลจีโนม (Genome data collection), 2) การเตรียมข้อมูลจีโนม (Genome data preprocessing), 3) การเปรียบเทียบข้อมูลจีโนม (Genome data comparison), 4) การวิเคราะห์หาความแตกต่าง (Identification of genomic variations) และ 5) การพัฒนาระบบแสดงความแตกต่างในระดับจีโนม (Genomic variation display tool) ดังที่แสดงอยู่ในรูป 3



รูปที่ 3 วิธีการดำเนินการวิจัยเพื่อหาความแตกต่างในระดับจีโนมระหว่าง *Arthrospira* สายพันธุ์ต่างๆ

1 การรวบรวมข้อมูลจีโนมของ *Arthrospira* จากฐานข้อมูลจีโนมสาธารณะต่างๆ (Genome data collection)

ข้อมูลจีโนมของ *Arthrospira* สายพันธุ์ต่างๆที่จะถูกนำมาใช้เปรียบเทียบเพื่อหาความแตกต่างในระดับจีโนมและหาตำแหน่งจีโนมเครื่องหมาย (Genomic markers) จะถูกเก็บรวบรวมจากฐานข้อมูลจีโนมสาธารณะชั้นนำต่างๆ ดังตัวอย่างที่แสดงอยู่ในตารางที่ 2 โดยฐานข้อมูลเหล่านี้ได้ อนุญาตให้นักวิจัยและผู้สนใจสามารถเข้าปค้นหาและดาวน์โหลดข้อมูลจีโนมของสิ่งมีชีวิตสายพันธุ์ต่างๆได้ผ่านระบบเว็บแอปพลิเคชันที่ทางผู้พัฒนาฐานข้อมูลนั้นๆสร้างขึ้นมาโดยไม่ต้องเสียค่าใช้จ่ายแต่อย่างใด ยกตัวอย่างเช่น ฐานข้อมูล NCBI Genome Database ที่ในปัจจุบันมีข้อมูลจีโนมของสิ่งมีชีวิตกลุ่มแบคทีเรียมากกว่า 14,000 สายพันธุ์ (ข้อมูลเดือนกันยายน 2555) โดยผู้ใช้สามารถเข้าถึงข้อมูลจีโนมเหล่านี้ ผ่านทางเว็บไซต์ (<http://www.ncbi.nlm.nih.gov/genome/browse/>) หรือทาง FTP Server (<ftp.ncbi.nlm.nih.gov/genomes/>) ข้อมูลดิบที่ได้จากฐานข้อมูลต่างๆนี้จะถูกนำไปเข้าสู่ขั้นตอนการเตรียมข้อมูลเพื่อการวิเคราะห์ต่อไป

ตารางที่ 2 ตัวอย่างฐานข้อมูลจีโนมสาธารณะที่สามารถเข้าไปดาวน์โหลดข้อมูลจีโนมได้

Database	Description	URL	Ref.
NCBI Genome Database	Contains genome sequence and annotation data for both complete and draft genomes	http://www.ncbi.nlm.nih.gov/genome/browse/	[25]
GOLD Genome	Contains comprehensive access to information	http://www.genomesonli	[26]

Database	regarding genome and metagenome sequencing projects	ne.org	
Ensembl Genomes	Provides access to genomes of non-vertebrate species (bacteria, protists, fungi, plants and invertebrate metazoa).	http://www.ensemblgenomes.org/	[27]
Comprehensive Microbial Resource	Contains information on all of the publicly available, complete prokaryotic genomes	http://cmr.jcvi.org/	[28]
Integrated Microbial Genomes	Encyclopedia of Bacteria and Archaea genomes	http://img.jgi.doe.gov/cgi-bin/geba/main.cgi	[29]

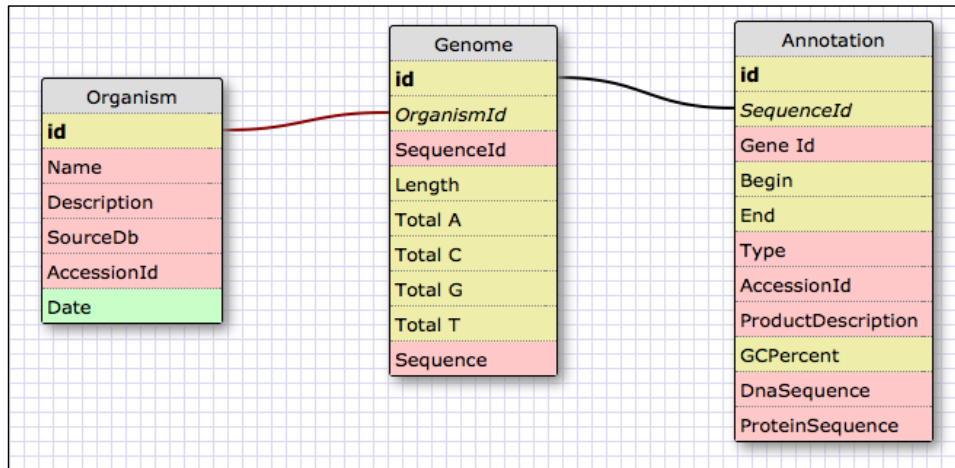
2 การเตรียมข้อมูลจีโนม (Genome data preprocessing)

ข้อมูลจีโนมที่ถูกดาวน์โหลดมาจากรฐานข้อมูลต่าง ๆ นั้นมักจะอยู่ในรูปแบบไฟล์ที่ไม่เหมือนกันดังตัวอย่างที่แสดงอยู่ในตารางที่ 3 โดยไฟล์แต่ละประเภทจะมีการจัดเก็บข้อมูลจีโนมที่แตกต่างกันไป ทำให้มีความจำเป็นที่จะต้องนำเอาไฟล์จีโนมที่ได้มาทั้งหมดมาบริหารจัดการเพื่อดึงเฉพาะข้อมูลที่มีประโยชน์ต่อโจทย์วิจัยและจัดเก็บให้อยู่ในรูปแบบเดียวกันในระบบฐานข้อมูล โดยขั้นตอนนี้จะช่วยให้การดึงเอาข้อมูลต่างๆจากแต่ละจีโนมเพื่อทำการเปรียบเทียบนั้นทำได้ง่ายและสะดวกรวดเร็วมากขึ้น

ตารางที่ 3 ตัวอย่างรูปแบบไฟล์จีโนมที่ได้จากการดาวน์โหลดข้อมูลจีโนมจากฐานข้อมูลสาธารณะ

Name	Description	More info
embl	A rich format for storing sequences and their associated meta-information, feature coordinates, and annotations	http://www.ebi.ac.uk/embl/Documentation/User_manual/usrman.html
fasta	A text-based format for representing either nucleotide sequences or peptide sequences, in which nucleotides or amino acids are represented using single-letter codes	http://en.wikipedia.org/wiki/FASTA_format
genbank	A rich format for storing sequences and associated annotations	http://www.ncbi.nlm.nih.gov/Sitemap/samplerecord.html
seqxml	an XML Schema to describe biological sequences	http://seqxml.org/xml/Main.html

ในขั้นตอนการเตรียมข้อมูลจีโนม หรือ Preprocess นี้ ภาษาโปรแกรมมิ่ง Perl และ BioPerl Module (<http://www.bioperl.org/>) จะถูกใช้ในการเขียนชุดคำสั่งเพื่ออ่านและดึงข้อมูลจากไฟล์จีโนมทั้งหมดและจัดเก็บลงไปในระบบฐานข้อมูล Relational database ดังตัวอย่างที่แสดงอยู่ในรูปที่ 4 โดยระบบ Relational database นี้จะถูกพัฒนาโดยใช้โปรแกรม MySQL (<http://www.mysql.com/>) สำหรับข้อมูลจากไฟล์จีโนมที่จำเป็นต่อการวิเคราะห์หาความแตกต่างประกอบด้วย a) ลำดับนิวคลีโอไทด์ของจีโนม b) ลำดับนิวคลีโอไทด์และลำดับกรดอะมิโนของยีนทั้งหมดที่พบในจีโนม และ c) ข้อมูลเกี่ยวกับหน้าที่ของแต่ละยีน เป็นต้น นอกจากนี้ทางทีมวิจัยจะมีการตรวจสอบข้อมูลจีโนมของ *Arthrospira* ในฐานข้อมูลสาธารณะอยู่เป็นระยะๆ (เช่น เดือนละหนึ่งครั้ง) เพื่อรวบรวมข้อมูลจีโนมของ *Arthrospira* สายพันธุ์อื่นๆที่อาจจะถูกเพิ่มเข้ามาในฐานข้อมูลในอนาคตได้



รูปที่ 4 ตัวอย่าง Entity-relationship model ของระบบฐานข้อมูลเพื่อจัดเก็บข้อมูลจีโนม

3 การเปรียบเทียบข้อมูลจีโนม (Genome data comparison)

ข้อมูลจีโนมในฐานข้อมูลภายในที่ได้จากขั้นตอน Preprocessing จะถูกนำมาเปรียบเทียบกับกันโดยใช้โปรแกรมสำเร็จรูปต่างๆดังตัวอย่างที่แสดงอยู่ในตารางที่ 4 ข้อมูลจีโนมจะถูกนำมาเปรียบเทียบแบบทั้งหมดเทียบกับทั้งหมด (All against all) ทั้งในระดับนิวคลีโอไทด์ของจีโนมและนิวคลีโอไทด์และกรดอะมิโนของยีน โดยวิธีการเปรียบเทียบจะทำทั้งแบบคู่ (Pairwise sequence alignment) ผ่านโปรแกรม เช่น BLAST+ และแบบกลุ่ม (Multiple sequence alignment) ผ่าน โปรแกรมเช่น Clustal Omega, MAUVE หรือ MUSCLE โปรแกรมสำเร็จรูปทั้งหมดที่ทางทีมวิจัยจะนำมาใช้จะต้องเป็นโปรแกรมที่มีการพัฒนาอย่างต่อเนื่องและเป็นที่ยอมรับว่าสามารถนำมาใช้ได้ดีกับข้อมูลลักษณะนี้ (State of the art software) โดยโปรแกรมต่างๆจะถูกติดตั้งและใช้ใน Command-line environment ซึ่งจะช่วยให้ทางทีมงานวิจัยสามารถพัฒนาชุดคำสั่งคอมพิวเตอร์ (Computational protocol) เพื่อเรียกใช้โปรแกรมต่างๆและรวบรวมผลได้อย่างเป็นระบบผ่านภาษาโปรแกรมมิ่ง Perl โดยวิธีดังกล่าวจะช่วยให้การวิเคราะห์ข้อมูลจีโนมจำนวนมากสามารถทำได้ อย่างมีประสิทธิภาพเพราะจะสามารถสั่งให้คอมพิวเตอร์ทำตามขั้นตอนที่กำหนดไว้ล่วงหน้าได้อย่างอัตโนมัติ หลังจากนั้นผลการเปรียบเทียบทั้งหมดจะถูกจัดเก็บในระบบฐานข้อมูลเพื่อการวิเคราะห์ และแบ่งแยกความแตกต่างในลำดับนิวคลีโอไทด์ที่พบบนแต่ละจีโนมในขั้นตอนต่อไป

ตารางที่ 4 ตัวอย่างโปรแกรมสำเร็จรูปสำหรับเปรียบเทียบข้อมูลจีโนม

Name	Description	Download URL	Ref.
BLAST+	Finds regions of local similarity between nucleotide or protein sequences and a sequence database	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LAT-EST/	[22]
Clustal Omega	For creating multiple sequence alignments of protein or DNA sequences using Hidden Markov Model method	http://www.clustal.org/	[24]
MAUVE	For constructing multiple genome alignments in the presence of large-scale evolutionary events such as rearrangement and inversion	http://gel.ahabs.wisc.edu/mauve/	[23]
MUSCLE	For creating multiple sequence alignments of protein or DNA sequences using Progressive/iterative alignment method	http://www.ebi.ac.uk/Tools/msa/muscle/	[30]

4 การหาความแตกต่างที่พบในจีโนมของ *Arthrospira* สายพันธุ์ต่างๆ (Identification of genomic variations)

ผลการเปรียบเทียบข้อมูลจีโนมของ *Arthrospira* หลายสายพันธุ์ที่ได้จากโปรแกรมต่างๆจะถูกวิเคราะห์เพื่อหาความแตกต่างในลำดับนิวคลีโอไทด์ที่พบในแต่ละจีโนมโดยความแตกต่างที่สามารถนำมาใช้เป็นข้อมูลเพื่อคัดเลือกตำแหน่งจีโนมเป้าหมาย (Genomic marker) ที่ทางกลุ่มวิจัยสนใจสามารถแบ่งออกได้เป็นสองแบบ คือ ความแตกต่างในระดับสปีส์ หรือ SNPs (Single Nucleotide Polymorphisms) ที่พบอยู่บนบริเวณจีโนมที่ถูกอนุรักษ์ไว้ (Conserved genomic region) และ ความแตกต่างในระดับยีนที่พบบนบริเวณจีโนมที่มีความจำเพาะเจาะจงกับแต่ละจีโนม (Variable genomic region) เนื่องจากว่าผลที่ได้จากการเปรียบเทียบข้อมูลจีโนมแบบทั้งหมดเทียบกับทั้งหมด (all against all) โดยใช้โปรแกรมต่างๆนั้นจะมีอยู่เป็นจำนวนมาก จึงมีความจำเป็นที่จะต้องพัฒนาโปรแกรมคอมพิวเตอร์เพื่อช่วยวิเคราะห์และหาความแตกต่างอย่างเป็นระบบ โดยทางทีมงานวิจัยจะใช้ภาษาโปรแกรมมิ่ง Perl และ BioPerl Module เพื่ออ่านไฟล์ที่ได้จากโปรแกรมเปรียบเทียบข้อมูลจีโนมต่างๆ และวิเคราะห์หาความแตกต่างทั้งสองแบบต่อไป ผลที่ได้จากขั้นตอนนี้จะถูกเก็บในระบบฐานข้อมูลเพื่อแสดงผลต่อนักวิจัยหรือผู้สนใจต่อไป

5 การแสดงความแตกต่างในระดับจีโนมผ่านการใช้เว็บแอปพลิเคชัน (Web-based genomic variation display tool)

ข้อมูลความแตกต่างในระดับจีโนมระหว่าง *Arthrospira* สายพันธุ์ต่างๆที่ได้จากงานวิจัยนี้จะถูกนำเสนอสู่นักวิจัยและผู้สนใจผ่านระบบเว็บแอปพลิเคชันที่ทางทีมงานจะสร้างขึ้น โดยคาดว่าเว็บแอปพลิเคชันนี้จะต้องสามารถแสดงถึงตำแหน่งและลำดับเบสตรงบริเวณที่มีความแตกต่างควบคู่กับข้อมูลหน้าที่ของยีน ซึ่งข้อมูลลักษณะนี้จะมีประโยชน์สำหรับการคัดเลือกตำแหน่งบนจีโนมที่จะนำมาใช้เป็นตำแหน่งจีโนมเป้าหมาย (Genomic marker) เพื่อการคัดเลือกสายพันธุ์ของ *Arthrospira* ที่พบในธรรมชาติต่อไป การนำเสนอข้อมูลผ่านเว็บยังช่วยให้นักวิจัยภายนอกอื่นๆสามารถเข้าถึงข้อมูลนี้ได้ง่าย ซึ่งจะช่วยผลักดันให้งานวิจัยที่เกี่ยวข้องกับการหาหรือคัดเลือกสายพันธุ์ของ *Arthrospira* ผ่านการใช้ข้อมูลจีโนมสามารถทำได้อย่างมีประสิทธิภาพและรวดเร็วมากขึ้น

บทที่ 3 ผลการวิจัยและวิจารณ์

การรวบรวมข้อมูลจีโนมของ *Arthrospira* จากฐานข้อมูลจีโนมสาธารณะต่างๆ (Genome data collection)

ข้อมูลจีโนมของ *Arthrospira* สายพันธุ์ต่างๆ ที่ถูกนำมาใช้เปรียบเทียบเพื่อหาความแตกต่างในระดับจีโนม และหาตำแหน่งจีโนมเครื่องหมาย (Genomic markers) ถูกเก็บรวบรวมจากฐานข้อมูล NCBI Genome Database (<http://www.ncbi.nlm.nih.gov/genome/browse/>) โดยข้อมูลจีโนมของ *Arthrospira* ทั้งหมดที่ดาวโหลดมาได้จะแสดงอยู่ใน **Error! Reference source not found. ตารางที่ 5** ซึ่งจะมีอยู่ทั้งหมด 5 จีโนมที่มีการเผยแพร่ข้อมูลสู่สาธารณะแล้ว ซึ่งประกอบด้วย *Arthrospira maxima* CS-328, *Arthrospira platensis* C1, *Arthrospira platensis* NIES-39, *Arthrospira platensis* str. Paraca, *Arthrospira* sp. PCC 8005 ข้อมูลจีโนมที่ดาวโหลดมาได้ทั้งหมดอยู่ในรูปแบบของไฟล์ Genbank ซึ่งในแต่ละไฟล์จะมีทั้งข้อมูลลำดับเบสของ จีโนม รวมไปถึงข้อมูลตำแหน่งของยีน, ผลผลิตของยีน, ลำดับเบสของยีน, และลำดับกรดอะมิโนของยีน

ตารางที่ 5 ข้อมูลจีโนมของ *Arthrospira* ที่ได้จากฐานข้อมูล NCBI Genome Database

Accession ID	Genome Name	Pubmed ID
NC_016640	<i>Arthrospira platensis</i> NIES-39	20203057
NZ_ABYK01000001	<i>Arthrospira maxima</i> CS-328	21890670
NZ_CM001632	<i>Arthrospira platensis</i> C1	22675597
CAFN01000001	<i>Arthrospira</i> sp. PCC 8005	-
ACSK03000001	<i>Arthrospira platensis</i> str. Paraca	-

การเตรียมข้อมูลจีโนม (Genome data preprocessing)

เพื่อให้การเข้าถึงข้อมูลที่เป็นที่อยู่ในข้อมูลจีโนมของ *Arthrospira* นั้นสามารถทำได้ง่ายและรวดเร็วขึ้น ไฟล์จีโนมที่ได้มาทั้งหมดถูกนำมาบริหารจัดการเพื่อดึงเฉพาะข้อมูลที่มีประโยชน์ต่อโจทย์วิจัยและจัดเก็บให้อยู่ในรูปแบบเดียวกัน โดยในขั้นตอนการ เตรียมข้อมูลจีโนม หรือ Preprocess นี้ ภาษาโปรแกรมมิ่ง Perl (<http://www.perl.org/>) และ BioPerl Module (<http://www.bioperl.org/>) ถูกใช้ในการเขียนชุดคำสั่งเพื่ออ่านและดึงข้อมูลจากไฟล์จีโนมทั้งหมด โดยข้อมูลหลักที่ถูกดึงออกมาจากแต่ละจีโนมจะประกอบด้วย

1. ลำดับเบสของจีโนม หรือ contig โดยจะจัดเก็บในรูปแบบไฟล์ Fasta
2. ลำดับเบสของยีน โดยจะจัดเก็บในรูปแบบไฟล์ Fasta
3. ลำดับโปรตีนของผลผลิตของยีน โดยจะจัดเก็บในรูปแบบไฟล์ Fasta
4. รายการยีนที่พบในจีโนมทั้งหมด โดยจะจัดเก็บในรูปแบบไฟล์ Tab-delimited
5. รายการจีโนม หรือ contig ทั้งหมด โดยจะจัดเก็บในรูปแบบไฟล์ Tab-delimited

โดยผลจากการเตรียมข้อมูลจีโนมถูกสรุปอยู่ในตารางที่ 6 ข้อมูลในตารางที่ 6 นั้นแสดงให้เห็นว่า *Arthrospira* สายพันธุ์ต่างๆ นั้นจะมีขนาดจีโนมที่มีความยาวรวมอยู่ที่ประมาณหกล้านกว่าเบส *Arthrospira platensis* NIES-39 เป็นสายพันธุ์ที่มีขนาดจีโนมที่ใหญ่ที่สุด และ *Arthrospira maxima* CS-328 เป็นสายพันธุ์ที่มีขนาดจีโนมเล็กที่สุด จำนวนยีนที่พบในแต่ละจีโนมจะอยู่ในช่วง 5,485 – 6,676 ยีนสำหรับ *Arthrospira platensis* str. Paraca นั้นถึงแม้สายพันธุ์นี้จะมีขนาดจีโนมที่ใหญ่เป็นอันดับสองรองจาก *Arthrospira platensis* NIES-39 แต่อย่างไรก็ตาม จำนวนยีนที่ถูกระบุมาในจีโนมนี้ก็กลับมีปริมาณที่น้อยที่สุด ตรงจุดนี้ก็อาจจะเป็นไปได้ว่าทางผู้ผลิตข้อมูลจีโนมของ *Arthrospira platensis* str. Paraca อาจจะมีการระบุตำแหน่งยีนที่พบในจีโนมนี้อย่างไม่ครบสมบูรณ์เท่าที่ควรนัก

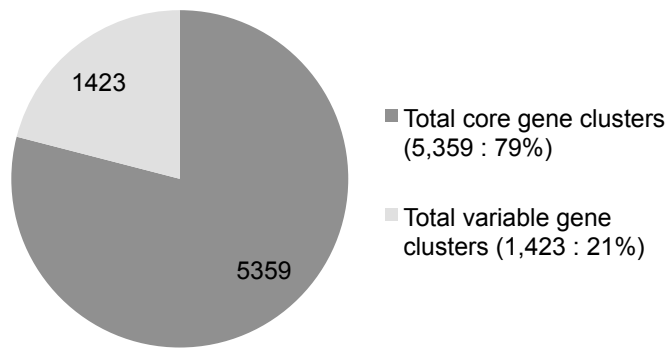
การวิเคราะห์ข้อมูลต่างๆที่มีการใช้จีโนมของ *Arthrospira platensis str. Paraca* จึงจำเป็นที่จะต้องคำนึงถึงจุดๆนี้ด้วย

ตารางที่ 6 สรุปข้อมูลยีนส์ที่พบในจีโนมของ *Arthrospira* ที่รวบรวมได้

Genome Name	Total number of contigs	Total contig lengths (bp)	GC Content (%)	Number of genes	Number of protein coding genes	Number of rRNA coding gene	Number of tRNA coding genes
<i>Arthrospira platensis</i> NIES-39	1	6788435	44	6676	6630	6	40
<i>Arthrospira maxima</i> CS-328	129	6003314	45	5728	5690	2	36
<i>Arthrospira platensis</i> C1	1	6089210	44	6153	6108	6	39
<i>Arthrospira</i> sp. PCC 8005	692	6166364	45	5996	5951	4	41
<i>Arthrospira platensis str. Paraca</i>	268	6501886	44	5485	5439	6	40

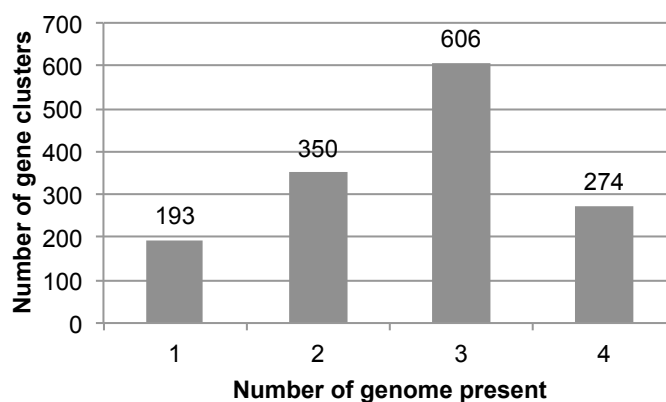
การเปรียบเทียบข้อมูลจีโนม (Genome data comparison)

ข้อมูล *Arthrospira* จีโนมและข้อมูลยีนส์ที่พบในจีโนมที่ได้จากขั้นตอน Preprocessing นั้นถูกนำมาเปรียบเทียบกัน แบบทั้งหมดเทียบกับทั้งหมด (All against all) ทั้งในระดับนิวคลีโอไทด์ของจีโนมและนิวคลีโอไทด์และกรดอะมิโนของยีน โดย วิธีการเปรียบเทียบจะทำทั้งแบบคู่ (Pairwise sequence alignment) ผ่านโปรแกรม BLAST+ โดยโปรแกรมต่างถูกติดตั้งและ ใช้ใน Command-line environment ซึ่งช่วยให้ทางทีมงานวิจัยสามารถพัฒนาชุดคำสั่งคอมพิวเตอร์ (Computational protocol) เพื่อเรียกใช้โปรแกรมต่างๆและรวบรวมผลได้อย่างเป็นระบบผ่านภาษาโปรแกรมมิ่ง Perl โดยวิธีดังกล่าวช่วยให้การวิเคราะห์ข้อมูลจีโนมจำนวนมากสามารถทำได้อย่างมีประสิทธิภาพ สำหรับข้อมูลจีโนมของ *Arthrospira* ที่รวบรวมมาได้ทั้งหมด 5 จีโนมนั้นมีจำนวนยีนส์รวมเท่ากับ 30,038 ยีนส์ โดยยีนส์ทั้งหมดนี้ถูกนำมาเปรียบเทียบแบบ All against all ทั้งในระดับนิวคลีโอไทด์และโปรตีน ซึ่งในขั้นตอนนี้ข้อมูลลำดับเบสนั้นได้ถูกนำมาเปรียบเทียบผ่านโปรแกรม BLAST+ ทั้งหมด 360,912,5776 ครั้ง เพื่อให้ได้มาซึ่งข้อมูล Cluster of Orthologous groups ที่เป็นข้อมูลการจัดกลุ่มยีนส์ที่เป็นยีนส์เดียวกันให้อยู่ในกลุ่มเดียวกัน โดยข้อมูล Cluster of Orthologous groups ดังกล่าวจะช่วยให้ทราบถึงจำนวนยีนส์ที่ต่างกันที่พบในจีโนมของ *Arthrospira* ทั้ง 5 สายพันธุ์ และยังสามารถช่วยให้ทราบได้ว่า มียีนส์ใดบ้างที่พบในทุกๆจีโนม และในขณะเดียวกันมียีนส์ใดบ้างที่พบเฉพาะในบางจีโนม ผลจากการเปรียบเทียบแสดงให้เห็นว่ายีนส์ทั้งหมด 30,038 ยีนส์ที่พบในจีโนมของ *Arthrospira* ทั้ง 5 สายพันธุ์ สามารถนำมาจัดกลุ่มได้ทั้งหมด 6,782 กลุ่มยีนส์ โดย 5,359 (79%) กลุ่มยีนส์นั้นเป็นยีนส์ที่พบในทั้ง 5 จีโนม หรือที่เรียกกันสั้นๆว่ากลุ่ม Core ยีนส์ และ 1,423 กลุ่มยีนส์ที่เหลือหรือประมาณ 21% มีการพบแค่ในบางจีโนมหรือที่เรียกกันสั้นๆว่ากลุ่ม Variable ยีนส์ ดังที่แสดงอยู่ในรูปที่ 5



รูปที่ 5 ผลของการจัดกลุ่มยีนทั้งหมด 30,038 ยีนที่พบในจีโนมของ *Arthrospira* ทั้ง 5 สายพันธุ์ แสดงให้เห็นว่ายีนทั้งหมดสามารถแบ่งออกได้เป็น 6,782 กลุ่มยีน โดยมี 5,359 (79%) กลุ่มเป็นยีนที่พบในทุกๆ จีโนม และ 1,423 กลุ่มเป็นยีนที่พบเฉพาะในบางจีโนม

สำหรับ 1,423 กลุ่มเป็นยีนที่พบเฉพาะในบางจีโนมนั้นสามารถแบ่งย่อยออกเป็น 4 กลุ่มตามจำนวนจีโนมที่มีการพบของแต่ละยีนดังที่แสดงอยู่ใน**รูปที่ 6** ผลตรงนี้แสดงให้เห็นว่า ยังมียีนกว่า 193 ยีนที่พบว่าจะมีอยู่ในเพียงจีโนมใดจีโนมหนึ่งเท่านั้น โดยยีนดังกล่าวอาจจะเกี่ยวข้องโดยตรงกับคุณสมบัติพิเศษที่พบในแต่ละสายพันธุ์ของ *Arthrospira* โดยในจำนวนนี้พบว่ามียีนอยู่ตั้งแต่ 0 – 88 ยีนที่พบเฉพาะในจีโนมใดจีโนมหนึ่งดังที่แสดงอยู่ในตารางที่ 7 ในขณะที่ยีนข้อมูลในตารางที่ 7 ยังแสดงให้เห็นว่า *Arthrospira platensis str. Paraca* นั้นไม่มียีนที่พบเฉพาะในจีโนมนี้ แต่อย่างไรก็ดีผลดังกล่าวอาจจะยังไม่ค่อยน่าเชื่อถือนักเพราะว่าจีโนมของ *Arthrospira platensis str. Paraca* นั้นมีการค้นพบยีนในจีโนมน้อยกว่าที่ควรจะเป็นดังที่ได้กล่าวไว้แล้วในข้างต้น ด้วยเหตุนี้การที่ไม่พบยีนที่จำเพาะเจาะจงกับ *Arthrospira platensis str. Paraca* อาจจะไม่ได้หมายความว่าสายพันธุ์ดังกล่าวไม่มียีนเฉพาะ แต่อาจจะหมายความว่ากำหนัดตำแหน่งของยีนที่มีอยู่ในจีโนมอาจจะยังทำมาไม่ครอบคลุมถึงกลุ่มยีนดังกล่าว



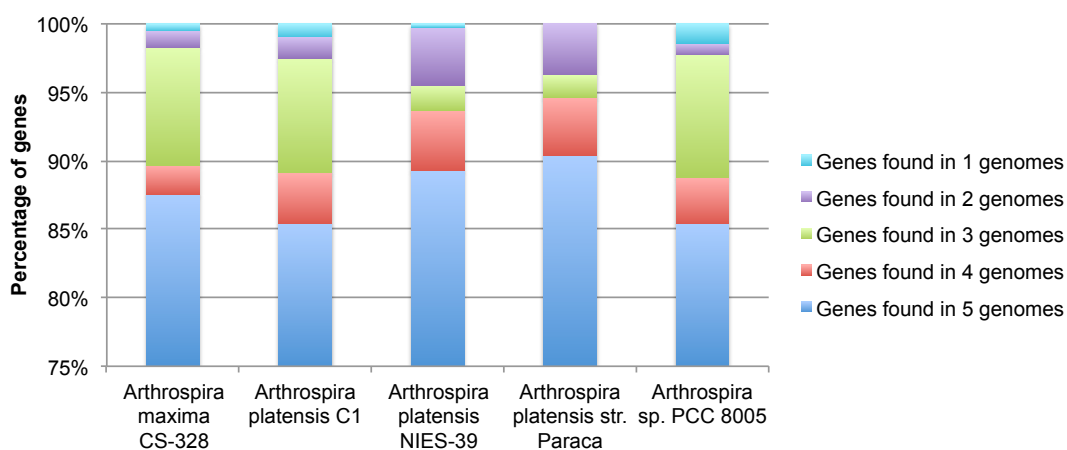
รูปที่ 6 การแบ่งกลุ่มของยีนที่พบเฉพาะในบางจีโนม ตามจำนวนจีโนมที่มีการพบยีนดังกล่าว

ตารางที่ 7 จำนวนยีนที่จำเพาะกับแต่ละจีโนม

Genome	Specific genes
<i>Arthrospira maxima</i> CS-328	30
<i>Arthrospira platensis</i> C1	60

<i>Arthrospira platensis</i> NIES-39	15
<i>Arthrospira platensis</i> str. Paraca	0
<i>Arthrospira</i> sp. PCC 8005	88
Total	193

หากลองวิเคราะห์ยีนส์ที่พบในแต่ละจีโนมของ *Arthrospira* นั้นพบกว่าประมาณ 85%-90% เป็นยีนส์ที่มีการพบในทุกๆจีโนม และประมาณ 10%-15% เป็นยีนส์ที่พบเฉพาะในบางจีโนมดังที่แสดงอยู่ในรูปที่ 7



รูปที่ 7 การแบ่งแยกยีนที่พบในแต่ละจีโนมของ *Arthrospira* ตามจำนวนจีโนมที่มีการพบยีนส์ดังกล่าว

ผลจากการวิเคราะห์แสดงให้เห็นว่าจีโนมของ *Arthrospira* แต่ละสายพันธุ์นั้นประกอบด้วยยีนส์ที่พบในทุกๆจีโนม และยีนส์ที่จำเพาะเจาะจงกับบางจีโนม ข้อมูลดังกล่าวแสดงให้เห็นถึงความหลากหลายในระดับยีนส์ ซึ่งจะเป็นข้อมูลสำคัญที่จะสามารถช่วยให้แยกแยะได้ว่า ยีนส์ไหนน่าจะเป็นยีนส์ที่มีความสำคัญกับลักษณะพิเศษของแต่ละสายพันธุ์ ในขณะที่ยีนส์ที่จำเพาะเจาะจงยังสามารถใช้เป็นเป้าหมายสำหรับการออกแบบชุดตรวจสอบสายพันธุ์ได้อีกด้วย

การแสดงความแตกต่างในระดับจีโนมผ่านการใช้เว็บแอปพลิเคชัน (Web-based genomic variation display tool)

ข้อมูลความแตกต่างในระดับจีโนมระหว่าง *Arthrospira* สายพันธุ์ต่างๆที่ได้จากงานวิจัยนี้ถูกนำเสนอสู่นักวิจัยและผู้สนใจผ่านระบบเว็บแอปพลิเคชันที่ทางทีมงานสร้างขึ้น ผู้ที่ต้องการค้นหาข้อมูลยีนส์ที่จำเพาะต่อ *Arthrospira* สายพันธุ์ *Arthrospira maxima* CS-328, *Arthrospira platensis* C1, *Arthrospira platensis* NIES-39, *Arthrospira platensis* str. Paraca หรือ *Arthrospira* sp. PCC 8005 สามารถเริ่มต้นโดยการเข้าไปที่ http://www1.sbi.kmutt.ac.th/~weerayuth/w1_57/ และจะพบเว็บไซต์ที่มีรูปแบบดังที่แสดงอยู่ในรูปที่ 8 สำหรับหน้าแรกของระบบเว็บแอปพลิเคชันจะประกอบด้วย 3 ส่วนหลักคือ 1) Summary, 2) Genome-specific genes, และ 3) Core genes that are found in all genomes ในส่วนของ Summary จะเป็นคำอธิบายของข้อมูลที่สามารถค้นหาได้จากเว็บแอปพลิเคชันนี้ ถัดลงมาในส่วนของ Genome-specific genes จะรวบรวมลิงค์ที่จะนำผู้ใช้ไปสู่หน้าที่แสดงตารางยีนส์ที่พบเฉพาะในจีโนมของ *Arthrospira* สายพันธุ์นั้นๆ สุดท้ายในส่วนของ Core genes that are found in all genomes จะมีลิงค์ที่จะนำผู้ใช้สู่หน้าที่แสดงถึงยีนส์ที่พบในจีโนมของ *Arthrospira* ทั้ง 5 สายพันธุ์



Summary

This page summarizes genomics diversity found in genomes of five different *Arthrospiras*. Specifically, five *Arthrospira* genomes (*Arthrospira maxima* CS-328, *Arthrospira platensis* C1, *Arthrospira platensis* NIES-39, *Arthrospira platensis* str. Paraca, *Arthrospira* sp. PCC 8005) were compared to identify genome-specific genes and core genes (genes shared by all genomes). The genome-specific genes can provide some clues on special characteristics of each genome. In addition, they can also be used as targets for molecular detection. The core genes on the other hand can be useful in identifying genes that define the genus or species of these *Arthrospira*. Multiple sequence alignments of homologous gene sequences are also provided and can be used to identify sequence variations (single nucleotide polymorphisms) that are found within the same gene across different genomes.

Genome-specific genes

- [Arthrospira maxima CS-328 Specific Genes](#)
- [Arthrospira platensis C1 Specific Genes](#)
- [Arthrospira platensis NIES-39 Specific Genes](#)
- [Arthrospira platensis str. Paraca Specific Genes](#)
- [Arthrospira sp. PCC 8005 Specific Genes](#)

Core genes that are found in all genomes

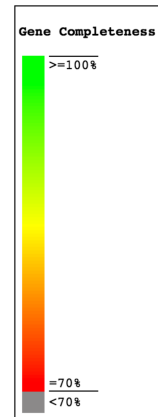
- [Core genes](#)

Pilot plant development and training insitute, King Mongkut's university of technology thonburi, 2014

รูปที่ 8 หน้าเริ่มต้นของระบบเว็บแอพลิเคชันสำหรับค้นหาข้อมูลที่จำเพาะกับ *Arthrospira* จำนวน 5 สายพันธุ์
http://www1.sbi.kmutt.ac.th/~weerayuth/w1_57/

ในส่วนของ Genome-specific genes นั้นหากผู้ใช้กดที่ลิงค์ใดลิงค์หนึ่งจะมีหน้าต่างใหม่ปรากฏขึ้นมา ซึ่งจะเป็นหน้าต่างที่แสดงตารางยีนส์ที่พบเฉพาะในจีโนมของ *Arthrospira* สายพันธุ์นั้นๆ ยกตัวอย่างเช่น หากผู้ใช้กดที่ลิงค์ “*Arthrospira maxima* CS-328 Specific Genes” จะมีหน้าต่างรูปที่ 9 ปรากฏขึ้น โดยหน้าต่างนี้จะแสดงตารางที่แต่ละแถวจะเป็นรายการยีนส์ ส่วนแต่ละคอลัมจะแสดงถึงการมีและไม่มีของยีนส์ดังกล่าวในแต่ละจีโนม โดยสีเขียวบ่งบอกถึงการมีและสีแดงบ่งบอกถึงการไม่มียีนส์นั้นๆ ในจีโนมที่แสดงอยู่บนหัวของแต่ละคอลัม เนื่องจากว่ารูปที่ 9 แสดงถึงยีนส์ที่พบเฉพาะในจีโนมของ *Arthrospira maxima* CS-328 ตารางดังกล่าวจะแสดงสีเขียวเฉพาะในคอลัมของ *Arthrospira maxima* CS-328 ผู้ใช้สามารถเข้าถึงข้อมูลเพิ่มเติมของยีนส์ในแต่ละจีโนมโดยการกดในช่องสีเขียวจะมีกล่องปรากฏขึ้นดังที่แสดงอยู่ในรูปที่ 10 โดยในกล่องนี้ผู้ใช้สามารถดาวน์โหลดลำดับนิวคลีโอไทด์ของยีนโดยการกดที่ [D] ดาวน์โหลดลำดับโปรตีนของผลผลิตของยีนโดยการกดที่ [N] นอกจากนั้นผู้ใช้อยังสามารถที่จะนำเอาลำดับโปรตีนของผลผลิตของยีนไปเปรียบเทียบกับฐานข้อมูลลำดับโปรตีนทั้งหมดในฐานข้อมูลของ NCBI ได้โดยการกดที่ [N] ซึ่งตรงนี้จะช่วยทำให้การศึกษาเพิ่มเติมเกี่ยวกับหน้าที่ของโปรตีนผ่านการเปรียบเทียบกับฐานข้อมูลโปรตีนทั้งหมดใน NCBI สามารถทำได้ง่ายและรวดเร็ว

#	Cluster ID	COG	Function Annotation	Expected gene Length (bp)	Arthrospira maxima CS-328	Arthrospira platensis C1	Arthrospira platensis NIES-39	Arthrospira platensis str. Paraca	Arthrospira sp. PCC 8005
1	p-cluster000233	Uncategorized	hypothetical protein	3882					
2	p-cluster000234	Uncategorized	hypothetical protein	237					
3	p-cluster001275	Uncategorized	hypothetical protein	2667					
4	p-cluster001276	Uncategorized	hypothetical protein	843					
5	p-cluster001362	Uncategorized	hypothetical protein	351					
6	p-cluster001459	Uncategorized	hypothetical protein	159					
7	p-cluster001475	COG0500 [QR]	hypothetical protein	726					
8	p-cluster001478	Uncategorized	putative hexapeptide transferase fa	1029					
9	p-cluster001479	Uncategorized	hypothetical protein	210					
10	p-cluster001481	COG0457 [R]	TFR repeat-containing protein	2256					
11	p-cluster001482	Uncategorized	hypothetical protein	1602					
12	p-cluster001483	Uncategorized	hypothetical protein	1023					
13	p-cluster001906	COG0535 [R]	Radical SAM domain protein	1068					
14	p-cluster001907	Uncategorized	hypothetical protein	837					
15	p-cluster001908	COG0535 [R]	Radical SAM domain protein	1023					



: Similar sequence present
 : Not present
 * : Within 300 bp from ends
 F : Possible frameshift
 I : Island genes

รูปที่ 9 หน้าตารางแสดงยีนส์และการพบหรือไม่พบในจีโนมต่างๆ

13	p-cluster001906	COG0535 [R]	Radical SAM domain protein	1068	[Close]
14	p-cluster001907	Uncategorized	hypothetical protein	837	Arthrospira maxima CS-328
15	p-cluster001908	COG0535 [R]	Radical SAM domain protein	1023	Gene ID
16	p-cluster002773	Uncategorized	hypothetical protein	315	AmoxDRAFT_1908 [D] [P] [N]
17	p-cluster002846	COG3972 [R]	hypothetical protein	2151	Length (bp) Contig ID (Length) Begin End %GC Product
					NZ_ABYK01000011 55979 54912 35 Radical SAM domain protein

[Close]
Arthrospira maxima CS-328
Gene ID
AmoxDRAFT_1908 [D] [P] [N]
Length (bp) Contig ID (Length) Begin End %GC Product
1068 NZ_ABYK01000011 55979 54912 35 Radical SAM domain protein

รูปที่ 10 กล้องแสดงข้อมูลยีนส์ในจีโนมที่ผู้ใช้สามารถเรียกดูได้โดยการกดที่ช่องมีปรากฏการมีของยีนส์

ในส่วนของ Core genes that are found in all genomes หรือยีนที่พบในจีโนมของ *Arthrospira* ทั้ง 5 สายพันธุ์นั้น หากผู้ใช้กดที่ลิงค์นี้ (รูปที่ 8 ด้านล่าง) จะปรากฏหน้าต่างที่เป็นตารางรายชื่อยีนต่างๆที่พบในทุกๆจีโนม ดังที่แสดงอยู่ในรูปที่ 11 จากตารางนี้จะแสดงให้เห็นว่าทุกช่องภายใต้คอลัมน์ของแต่ละจีโนมจะมีสีเขียวหรือสีอื่นๆ แสดงให้เห็นว่ายีนดังกล่าวมีการพบในจีโนมนั้นๆ นอกจากนั้นสีในช่องยังสามารถบอกถึงความหายากของยีนเมื่อเปรียบเทียบกับยีนเดียวกันที่ยาวที่สุดอีกด้วยตามรายละเอียดที่แสดงอยู่ในช่องแสดงความหมายของสีในด้านขวามือ ในขณะที่เดียวกันหากสีในช่องเป็นสีเทาอ่อน นั้นแสดงให้เห็นว่ายีนดังกล่าวไม่ได้ถูกกำหนดไว้ในจีโนม แต่มีบริเวณยีนที่มีลำดับเบส คล้ายกับยีนนั้นๆ

ตรงนี้จะช่วยให้ผู้ใช้สามารถเข้าไปวิเคราะห์เพิ่มเติมได้ว่ายีนดังกล่าวไม่ได้ถูกกำหนดเป็นยีนด้วยสาเหตุใด นอกจากนั้นยังช่วยไม่ให้เข้าใจผิดว่าไม่มียีนอยู่ แต่ในความเป็นจริงยีนดังกล่าวเพียงแค่มิได้ถูกกำหนดจากผู้ที่เป็นเจ้าของข้อมูลเท่านั้น หากผู้ใช้กดบนช่องภายใต้คอลัมน์ Cluster ID จะมีกล่องปรากฏขึ้นดังที่แสดงอยู่ในรูปที่ 12 ภายในกล่องนี้ผู้ใช้สามารถเข้าถึงข้อมูลลำดับเบสของยีนเดียวกันที่พบในจีโนมของ *Arthrospira* ทั้ง 5 สายพันธุ์ได้โดยการกดที่ลิงค์ “DNA fasta file” สำหรับลำดับของโปรตีนของยีนเดียวกันที่พบในทุกๆจีโนม ผู้ใช้สามารถดาวน์โหลดได้โดยการกดที่ลิงค์ “Protein fasta file” ซึ่งสองมุลลำดับทั้งสองชุดนี้จะช่วยให้ผู้สนใจสามารถเข้าถึงข้อมูลลำดับเบสของยีนเดียวกันได้อย่างรวดเร็ว นอกจากนั้นหากผู้ใช้กดที่ลิงค์ “DNA clustalw alignment” จะมีหน้าต่างที่แสดงถึงการทำ Multiple sequence alignment ของลำดับเบสของยีนเดียวกันจากจีโนมของ *Arthrospira* ทั้ง 5 สายพันธุ์ดังที่แสดงอยู่ในรูปที่ 13 ข้อมูล Multiple sequence alignment จะช่วยให้นักวิจัย

บทที่ 4 สรุปผลการดำเนินงานวิจัยและข้อเสนอแนะ

ในงานวิจัยนี้ความหลากหลายในระดับพันธุกรรมใน *Arthrospira* ที่มีข้อมูลจีโนมที่สามารถดาวน์โหลดได้จากฐานข้อมูลสาธารณะทั้งหมดได้ถูกนำมาศึกษาในเชิงลึก จากการค้นหาผ่านระบบฐานข้อมูลจีโนมสาธารณะต่างๆ พบว่ามี *Arthrospira* ทั้งหมด 5 สายพันธุ์ซึ่งประกอบด้วย *Arthrospira maxima* CS-328, *Arthrospira platensis* C1, *Arthrospira platensis* NIES-39, *Arthrospira platensis* str. Paraca และ *Arthrospira* sp. PCC 8005 ที่ได้ถูกอ่านลำดับเบสบนจีโนม รวมไปถึงกำหนดตำแหน่งยีนที่พบในจีโนมแล้ว ข้อมูลจีโนมดังกล่าวถูกนำมาเปรียบเทียบอย่างละเอียด เพื่อสรุปเป็นข้อมูลความหลากหลายในระดับยีนและในระดับเบส ผลของการเปรียบเทียบแสดงให้เห็นว่า ในจำนวนยีนที่พบในจีโนมของ *Arthrospira* ประมาณ 85-90% จะเป็นยีนที่พบได้ในทุกสายพันธุ์ และที่เหลืออีก 10-15% จะเป็นยีนที่มีการพบเพียงบางสายพันธุ์ นอกจากนั้นยังมียีนจำนวน 0-3% ที่เป็นยีนที่พบเฉพาะในจีโนมนั้นๆ ซึ่งยีนเหล่านี้จะเป็นยีนจำเพาะที่เกี่ยวข้องโดยตรงกับคุณลักษณะพิเศษของสายพันธุ์ต่างๆ ทั้งนี้ทั้งนั้นผลของการวิเคราะห์ยังแสดงให้เห็นว่า ไม่มีการพบยีนที่จำเพาะกับ *Arthrospira platensis* str. Paraca แต่อย่างไรก็ดีผลดังกล่าวไม่ได้หมายความว่า *Arthrospira platensis* str. Paraca ไม่มีลักษณะพิเศษ เพียงแต่ข้อมูลจีโนมของ *Arthrospira* สายพันธุ์นี้อาจจะมาพร้อมข้อมูลตำแหน่งยีนที่ไม่สมบูรณ์ ซึ่งมีความเป็นไปได้สูงเนื่องจากจำนวนยีนที่ถูกกำหนดมานั้นมีค่อนข้างน้อยเมื่อเปรียบเทียบกับขนาดของจีโนม นอกเหนือจากยีนจำเพาะแล้วนั้น ข้อมูลลำดับเบสของยีนที่พบในทุกๆจีโนมยังถูกนำมาเปรียบเทียบผ่านการทำ Multiple sequence alignment เพื่อแสดงถึงตำแหน่งเบสภายในยีนดังกล่าวที่มีความแตกต่างระหว่างสายพันธุ์ที่ต่างกัน โดยข้อมูลความหลากหลายในระดับรหัสพันธุกรรม ทั้งในระดับยีนและในระดับเบส นี่จะเป็นข้อมูลที่มีประโยชน์สำหรับการค้นหาตำแหน่งจีโนมเครื่องหมายสำหรับการจัดจำแนกสายพันธุ์ต่างๆของ *Arthrospira* ได้

เพื่อให้การเข้าถึงข้อมูลความหลากหลายในระดับรหัสพันธุกรรมของ *Arthrospira* ทั้งหมด 5 สายพันธุ์สามารถทำได้ง่ายและรวดเร็ว ทางกลุ่มวิจัยจึงได้มีการสร้างระบบเว็บแอปพลิเคชัน (Web-application) ขึ้น โดยระบบดังกล่าวมีการจัดข้อมูลต่างๆที่ได้จากงานวิจัยนี้ ไม่ว่าจะเป็นยีนที่เฉพาะในแต่ละจีโนม รวมไปถึงยีนที่พบในทุกๆจีโนม ออกเป็นหมวดหมู่ที่ชัดเจน เพื่อให้การเข้าถึงข้อมูลที่มีประโยชน์ต่างๆนี้สามารถทำได้ง่าย และรวดเร็ว นอกจากนี้ผู้ยังสามารถเข้าถึงข้อมูลพื้นฐานเช่น ลำดับเบสของยีน, ลำดับโปรตีนของผลผลิตของยีน, ลำดับเบสของยีนเดียวกันที่พบในหลายๆจีโนม และ Multiple sequence alignment ของลำดับเบสของยีนเดียวกันที่พบในจีโนมต่างๆ เป็นต้น ซึ่งผู้สนใจสามารถนำข้อมูลดังกล่าวไปต่อยอดเพื่อศึกษาเพื่อเติมเกี่ยวกับความเชื่อมโยงของความหลากหลายในระดับพันธุกรรม และลักษณะพิเศษของ *Arthrospira* สายพันธุ์ต่างๆ อันจะนำไปสู่ความรู้และความเข้าใจใหม่ๆของสิ่งมีชีวิตที่มีศักยภาพสูงอย่าง *Arthrospira*

เอกสารอ้างอิงของโครงการวิจัย

1. Rasmussen, U., & Svenning, M. M. (January 01, 1998). Fingerprinting of cyanobacteria based on PCR with primers derived from short and long tandemly repeated repetitive sequences. *Applied and Environmental Microbiology*, 64, 1, 265-72.
2. Dillon, J. C., Phuc, A. P., & Dubacq, J. P. (January 01, 1995). Nutritional value of the alga *Spirulina*. *World Review of Nutrition and Dietetics*, 77, 32-46.
3. Karkos, P. D., Leong, S. C., Sivaji, N., Karkos, C. D., & Assimakopoulos, D. A. (February 01, 2011). *Spirulina* in clinical practice: Evidence-based human applications. *Evidence-based Complementary and Alternative Medicine*, 2011.
4. Cheong, S. H., Kim, M. Y., Sok, D. E., Hwang, S. Y., Kim, J. H., Kim, H. R., Lee, J. H., ... Kim, M. R. (January 01, 2010). *Spirulina* prevents atherosclerosis by reducing hypercholesterolemia in rabbits fed a high-cholesterol diet. *Journal of Nutritional Science and Vitaminology*, 56, 1, 34-40.
5. Kay, R. A. (January 01, 1991). Microalgae as food and supplement. *Critical Reviews in Food Science and Nutrition*, 30, 6, 555-73.
6. Marles, R. J., Barrett, M. L., Barnes, J., Chavez, M. L., Gardiner, P., Ko, R., Mahady, G. B., ... Griffiths, J. (August 01, 2011). United states pharmacopeia safety evaluation of spirulina. *Critical Reviews in Food Science and Nutrition*, 51, 7, 593-604.
7. Hirahashi, T., Matsumoto, M., Hazeki, K., Saeki, Y., Ui, M., & Seya, T. (January 01, 2002). Activation of the human innate immune system by *Spirulina*: augmentation of interferon production and NK cytotoxicity by oral administration of hot water extract of *Spirulina platensis*. *International Immunopharmacology*, 2, 4, 423-34.
8. Hayashi, T., Hayashi, K., Maeda, M., & Kojima, I. (January 01, 1996). Calcium spirulan, an inhibitor of enveloped virus replication, from a blue-green alga *Spirulina platensis*. *Journal of Natural Products*, 59, 1, 83-7.
9. Ramamoorthy, A., & Premakumari, S. (January 01, 1996). Effect of Supplementation of *Spirulina* on Hypercholesterolemic Patients. *Journal of Food Science and Technology Mysore*, 33, 2, 124-127.
10. Mathew, B., Sankaranarayanan, R., Nair, P. P., Varghese, C., Somanathan, T., Amma, B. P., Amma, N. S., ... Nair, M. K. (January 01, 1995). Evaluation of chemoprevention of oral cancer with *Spirulina fusiformis*. *Nutrition and Cancer*, 24, 2, 197-202.
11. Misbahuddin, M., Maidul, I. A. Z. M., Khandker, S., Ifthaker-Al-Mahmud, ., Islam, N., & Anjumanara, . (January 01, 2006). Efficacy of *Spirulina* Extract Plus Zinc in Patients of Chronic Arsenic Poisoning: A Randomized Placebo-Controlled Study. *Clinical Toxicology (formerly Journal of Toxicology: Clinical Toxicology)*, 44, 2, 135-141.
12. Henrikson R. *Spirulina* World Food: How this micro algae can transform your health and our planet. Ronore Enterprises, Inc. PO Box 909, Hana, Maui, Hawaii 96718; 2010.
13. Hendrickx, L., et al., Microbial ecology of the closed artificial ecosystem MELiSSA (Micro-Ecological Life Support System Alternative): reinventing and compartmentalizing the Earth's food and oxygen regeneration system for long-haul space exploration missions. *Research in Microbiology*, 2006. 157: p. 77-86.

14. Ciferri, O., & Tiboni, O. (January 01, 1985). The biochemistry and industrial potential of *Spirulina*. *Annual Review of Microbiology*, 39, 503-26.
15. Li, W., Raoult, D., & Fournier, P.-E. (September 01, 2009). REVIEW ARTICLE: Bacterial strain typing in the genomic era. *Fems Microbiology Reviews*, 33, 5, 892-916.
16. Kittichotirat, W., Bumgarner, R. E., Asikainen, S., & Chen, C. (January 01, 2011). Identification of the pangenome and its components in 14 distinct *Aggregatibacter actinomycetemcomitans* strains by comparative genomic analysis. *Plos One*, 6, 7.)
17. Cheevadhanarak, S., Paithoonrangsarit, K., Prommeenate, P., Vorapreeda, T., Khannapho, C., Thammamongtham, C., Hongsthong, A., ... Kaneko, T. (April 02, 2012). Draft genome sequence of *Arthrospira platensis* C1 (PCC9438). *Standards in Genomic Sciences*, 6, 1, 43-53.
18. Didelot, X., Bowden, R., Wilson, D. J., Peto, T. E., & Crook, D. W. (January 01, 2012). Transforming clinical microbiology with bacterial genome sequencing. *Nature Reviews. Genetics*, 13, 9, 601-12
19. Shklar, G., & Schwartz, J. (January 01, 1988). Tumor necrosis factor in experimental cancer regression with alphatocopherol, beta-carotene, canthaxanthin and algae extract. *European Journal of Cancer & Clinical Oncology*, 24, 5, 839-50.
20. Reddy, M. C., Subhashini, J., Mahipal, S. V., Bhat, V. B., Srinivas, R. P., Kiranmai, G., Madyastha, K. M., ... Reddanna, P. (January 01, 2003). C-Phycocyanin, a selective cyclooxygenase-2 inhibitor, induces apoptosis in lipopolysaccharide-stimulated RAW 264.7 macrophages. *Biochemical and Biophysical Research Communications*, 304, 2, 385-92.
21. Mardis, E. R. (March 01, 2008). The impact of next-generation sequencing technology on genetics. *Trends in Genetics*, 24, 3, 133-141.
22. Johnson, M., Zaretskaya, I., Raytselis, Y., Merezuk, Y., McGinnis, S., & Madden, T. L. (January 01, 2008). NCBI BLAST: a better web interface. *Nucleic Acids Research*, 36, 5.
23. Darling, A. E., Mau, B., & Perna, N. T. (January 01, 2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *Plos One*, 5, 6.)
24. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., ... Higgins, D. G. (January 01, 2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7.
25. Wheeler, D. L., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvernin, V., Church, D. M., ... Yaschenko, E. (January 01, 2006). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, 34, 173.
26. Pagani, I., Liolios, K., Jansson, J., Chen, I. M., Smirnova, T., Nosrat, B., Markowitz, V. M., ... Kyrpides, N. C. (January 01, 2012). The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Research*, 40, 571-9.
27. Kersey, P. J., Staines, D. M., Lawson, D., Kulesha, E., Derwent, P., Humphrey, J. C., Hughes, D. S., ... Birney, E. (January 01, 2012). Ensembl Genomes: an integrative resource for genome-scale data from non-vertebrate species. *Nucleic Acids Research*, 40, 91-7.
28. Peterson, J. D., Umayam, L. A., Dickinson, T., Hickey, E. K., & White, O. (January 01, 2001). The Comprehensive Microbial Resource. *Nucleic Acids Research*, 29, 1, 123-5.

29. Markowitz, V. M., Chen, I. M., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., Ratner, A., ... Kyrpides, N. C. (January 01, 2010). The integrated microbial genomes system: an expanding comparative analysis resource. *Nucleic Acids Research*, 38, 382-90.
30. Edgar, R. C. (January 01, 2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32, 5, 1792-1797.