

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การจัดลำดับเบสหลายลำดับนั้นเป็นส่วนหนึ่งของศาสตร์ทางด้านชีวสารสนเทศศาสตร์ (Bioinformatics) และเป็นพื้นฐานของชีววิทยาระดับโมเลกุล (Molecular biology) ลำดับเบสแต่ละลำดับที่มีความใกล้เคียงกันทั้งทางโครงสร้างและองค์ประกอบจะแสดงถึงความใกล้เคียงกันทางสายพันธุ์ที่มี นำไปสู่การสร้างต้นไม้วิวัฒนาการ (Phylogenetic tree) และการเปรียบเทียบกันระหว่างลำดับเบสนั้นยังสามารถทำให้ได้ข้อมูลโครงสร้างในขั้นที่สอง และขั้นที่สาม (Secondary and tertiary structure) [1],[2]

การจัดเรียงลำดับเบสหลายลำดับเป็นปัญหาที่แก้ไขได้ยาก (NP-hard) เพราะลำดับเบสทั้งสองที่ต้องการเปรียบเทียบกันนั้นไม่ได้เหมือนกันทั้งหมด เนื่องจากลำดับเบสมีการเกิด การลบ (Deletion) และ การแทรก (Insertion) ในขั้นตอนวิวัฒนาการสิ่งมีชีวิตอยู่อย่างสม่ำเสมอตั้งแต่ในอดีต โดยขบวนการแทรกนั้นเป็นตัวที่ทำให้เบสในตำแหน่งนั้นเปลี่ยนไปจากเดิม

ในอดีตวิธีที่ใช้ในการแก้ปัญหาการจัดเรียงลำดับเบสหลายลำดับนั้นใช้กำหนดการพลวัต (Dynamic programming) [3] แต่วิธีนี้มีข้อเสียคือขนาดของปัญหาจะใหญ่ขึ้นตามจำนวนของความยาว และจำนวนของลำดับเบส ปัญหาการจัดเรียงลำดับเบสหลายลำดับจึงถูกจัดอยู่ในปัญหาที่แก้ได้ยาก จากปัญหาที่ได้รับทำให้ได้มีการนำวิทยาการศึกษาลำบาก (Heuristic) มาประยุกต์ใช้ ทำให้เกิดวิธีการจัดเรียงแบบก้าวหน้า (Progressive alignment algorithm) โดยที่วิธีนี้จะค่อยๆทำการจัดเรียงโดย ขั้นแรกได้มีการกะประมาณระยะทางของการวิวัฒนาการ (Evolutionary distance) ระหว่างทุกๆลำดับเบสต่อมาจึงนำลำดับที่ได้มาจัดเรียงตามลำดับความใกล้เคียงกันทางวิวัฒนาการ โปรแกรมที่ใช้วิธีนี้และเป็นที่รู้จักกันดีคือ Clustal V [4] , Clustal W [5] และ Clustal X [6] อย่างไรก็ตามวิธีขั้นตอนการจัดเรียงแบบก้าวหน้านั้นจะเกิดปัญหาในการเจอคำตอบที่อาจจะไม่ดีที่สุด (Local optima) จากปัญหาที่ได้รับทำให้มีการนำวิธีต่างๆมาใช้ รวมถึงวิธีการทำซ้ำ (Iterative) ซึ่งวิธีการทำซ้ำนั้นมีทั้งวิธีการจำลองการหลอมละลาย (Simulate annealing: SA) และขั้นตอนวิธีเชิงวิวัฒนาการ (Evolutionary Algorithm: EA) [7-15] ทั้งวิธีการจำลองการหลอมละลาย และวิธีขั้นตอนวิธีเชิงวิวัฒนาการนั้นมีประสิทธิภาพในการหาคำตอบที่มีขนาด

ของปัญหาใหญ่ได้ดี แต่วิธีการจำลองการหลอมละลาย กับการแก้ปัญหาการจัดเรียงลำดับเบสหลายลำดับนั้น จะเกิดการที่คำตอบนั้นจะเอนคำตอบเข้าไปสู่คำตอบที่อาจจะไม่ดีที่สุดได้ ซึ่งขั้นตอนวิธีเชิงวิวัฒน์จะมีวิธีในการที่จะหาคำตอบได้ดีกว่า อย่างไรก็ตามขั้นตอนวิธีเชิงวิวัฒน์ก็ยังมีปัญหาบางอย่าง เช่นการที่จะต้องใช้เวลาในการหาคำตอบจากขั้นตอนต่างๆ เช่น การสร้างกลุ่มของประชากรในตอนเริ่มต้น และต้องใช้เวลาในการที่จะค่อยๆปรับปรุงคำตอบเพื่อที่จะให้ได้คำตอบที่ดีที่สุด เป็นต้น

ในงานวิจัยส่วนมากมีการใช้สมการผลรวมคู่เบส (Sum-of-pairs) เป็นฟังก์ชันวัตถุประสงค์ บ่งบอกค่าความถูกต้องในการจัดลำดับเบสหลายลำดับ แต่ค่าการทำโทษที่ใช้กับสมการผลรวมคู่เบสมีการเปลี่ยนแปลงได้ตามค่าคงที่ในสมการ ดังนั้นในงานวิจัยนี้จึงได้นำขั้นตอนวิธีเชิงวิวัฒน์แบบหลายวัตถุประสงค์มาแก้ปัญหาในการกำหนดค่าคงที่ในการทำโทษให้กับสมการผลรวมคู่เบส

1.2 วัตถุประสงค์

งานวิจัยนี้มีจุดประสงค์ที่จะพัฒนาขั้นตอนวิธีเชิงวิวัฒน์แบบหลายวัตถุประสงค์ในการแก้ปัญหาการจัดลำดับเบสหลายลำดับ โดยโปรแกรมพยายามที่จะพัฒนาประสิทธิภาพของคำตอบให้ดีขึ้นกว่างานวิจัยที่มีอยู่

1.3 ขอบเขตงานวิจัย

1. ใช้ขั้นตอนวิธีเชิงวิวัฒน์แบบหลายวัตถุประสงค์แก้ปัญหาการจัดเรียงลำดับเบสหลายลำดับ
2. ใช้ข้อมูลจากฐานข้อมูลบาลีเบส (BaliBASE) [16]
3. พยายามหาคำตอบที่มีคุณภาพมากกว่างานวิจัยก่อน

1.4 ขั้นตอนและวิธีดำเนินงานวิจัย

1. ศึกษาโปรแกรมที่ใช้ขั้นตอนวิธีพันธุกรรม เช่น MSA-EA [7-9] หรือ SAGA [10] เป็นต้น
2. ออกแบบโครงสร้างของโปรแกรมว่าควรใช้วิธีอะไรบ้างในการสร้างขั้นตอนวิธีเชิงวิวัฒน์แบบหลายวัตถุประสงค์กับปัญหาการจัดลำดับเบสหลายลำดับ
3. เขียนโปรแกรมตามที่ได้คิดไว้ข้างต้น
4. ทดลองการทำงานกับข้อมูลจริง
5. วิเคราะห์และประเมินผล

1.5 ประโยชน์ที่คาดว่าจะได้รับ

สามารถนำโปรแกรมนี้ไปใช้เป็นอีกทางเลือกหนึ่งให้กับการแก้ปัญหาการจัดเรียงลำดับเบสหลายลำดับได้อย่างมีประสิทธิภาพ

1.6 ผลงานที่ตีพิมพ์จากวิทยานิพนธ์

ส่วนหนึ่งของวิทยานิพนธ์นี้ได้รับการตีพิมพ์เป็นบทความทางวิชาการในหัวข้อเรื่อง

“ขั้นตอนวิธีเชิงวิวัฒนาการดัดแปลงสำหรับการจัดเรียงลำดับเบสหลายลำดับ” โดย พศุทธิ์ สีเหลืองสวัสดิ์ และประภาส จงสถิตย์วัฒนา ในงานประชุมวิชาการ “The 8th National Computer Science and Engineering Conference” ซึ่งจัด ณ อ.หาดใหญ่ จ.สงขลา ในระหว่างวันที่ 21-22 ตุลาคม 2547

(Poster Presentation) “Multiple Objective Evolutionary Algorithm for Multiple Sequence Alignment”, GECCO2005, Washington DC, 25-29 June 2005.



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย