

การประกอบชิ้นส่วนดีเอ็นเอกลายเป็นปัญหา เนื่องจากการหาลำดับเบสของสายดีเอ็นเอที่มีความยาวจำนวนมากนั้น จำเป็นจะต้องทำการหาลำดับแบบช็อตกันโดยการทำให้ดีเอ็นเอมีชิ้นส่วนที่เล็กพอที่จะนำมาหาลำดับเบสได้ ซึ่งหลังจากนั้นนำชิ้นส่วนดีเอ็นเอที่ได้ลำดับเบสแล้วมาทำการประกอบกลับเข้าเป็นสายยาวอีกครั้ง ปัญหาในการประกอบชิ้นส่วนดีเอ็นเอนี้ อยู่ที่การหาอันดับและทิศทางที่ถูกต้องของชิ้นส่วนดีเอ็นเอแต่ละชิ้น และเมื่อจัดเรียงแล้วจะต้องเป็นลำดับที่สอดคล้องกับสายสายดีเอ็นเอดั้งเดิม วิธีการประกอบชิ้นส่วนดีเอ็นเอที่น่าเสนอนี้จะใช้กฎเมตาฮิวริสติก ประกอบด้วย 3 ขั้นตอนคือการหาคะแนนค่าซ้อนเหลื่อม ซึ่งได้จากวิธีสมิท-วอเตอร์แมน จากนั้นจะเป็นขั้นตอนการหาอันดับและทิศทางของชิ้นส่วนดีเอ็นเอ ซึ่งได้จากกฎฮิวริสติกเพื่อนบ้านที่ใกล้ที่สุดร่วมกับวิธีเอชไอเอส และนำค่าที่ได้มาจัดแถวลำดับชิ้นส่วนดีเอ็นเอให้เป็นคอนทิกด้วยโปรแกรมที่เขียนขึ้น โดยผลลัพธ์คอนทิกแต่ละเส้นที่ได้ สามารถพิจารณาเป็นชิ้นส่วนดีเอ็นเอย่อยเพื่อนำกลับไปคำนวณด้วยวิธีข้างต้นอีกครั้ง และผลลัพธ์จากการทดลองนี้ จะนำมาเทียบเคียงด้วยโปรแกรมแคปทริกกับสายดีเอ็นเอต้นแบบ

Abstract

192132

DNA fragment assembly is an important step in genome sequencing projects. The assembly problem generally arises during the sequencing of large strands of DNA where the strands are needed to be shotgun-replicated and broken into fragments that are small enough for sequencing. The assembly problem can thus be classified as a combinatorial optimization problem where the aim is to find the right order and orientation of each fragment in the ordering layout that leads to the formation of a consensus sequence that truly reflects the original DNA strands. This thesis presents a metaheuristic procedure for the task where the procedure contains three main steps: fragment overlap score calculation, fragment layout determination and contiguous sequence (contig) construction. The overlap score is obtained from a Smith-Waterman algorithm while the score-based fragment layout including the order and orientation of each fragment is generated by a nearest neighbor heuristic rule and an ant colony system algorithm. With the availability of the fragment layout, contigs can be constructed where each contig can also be viewed as a new DNA fragment. As a result, the metaheuristic procedure can be recursively applied to the resulting contig sets. The experimental results indicate that the performance of the metaheuristic procedure is comparable to that of a CAP3 assembly program.