

วิทยานิพนธ์นี้ได้กล่าวถึงการนำขั้นตอนวิธีระบบอาณานิคมมาใช้ในการประกอบชิ้นส่วนดีเอ็นเอ ซึ่งปัญหาในการประกอบชิ้นส่วนดีเอ็นเอจะเกิดขึ้นในระหว่างการจัดเรียงลำดับของสายดีเอ็นเอ ที่มีขนาดยาว โดยสายดีเอ็นเอจำเป็นต้องถูกจำลองแบบซ้ำกัน แล้วทำให้เป็นชิ้นส่วนที่มีขนาดเล็กพอที่จะนำมาจัดเรียงลำดับได้ ปัญหาการประกอบชิ้นส่วนดีเอ็นเอสามารถจัดให้เป็นปัญหาการหาค่าเหมาะที่สุดเชิงการจัดได้อย่างหนึ่ง โดยมีวัตถุประสงค์ที่จะหาอันดับ และทิศทาง ที่ถูกต้องของชิ้นส่วนดีเอ็นเอแต่ละชิ้น ในลำดับการจัดเรียงตัว ซึ่งจะนำไปสู่การสร้างลำดับสอดคล้อง ที่บ่งบอกถึงสายดีเอ็นเอดั้งเดิมได้ ในปัญหาหนึ่งคอนทิกนั้น จะมีชิ้นส่วนดีเอ็นเออย่างน้อยสองชิ้นส่วนที่มีการซ้อนเหลื่อมกัน และครอบคลุมแต่ละเบสของลำดับการจัดเรียงตัว และสำหรับปัญหาหลายคอนทิกนั้น จะมีบางตำแหน่งในลำดับการจัดเรียงตัวที่ไม่มีการซ้อนเหลื่อมของชิ้นส่วนหรือมีช่องว่าง เกิดขึ้น วิธีการประกอบที่ได้นำเสนอมีด้วยกันสองขั้นตอน คือ ขั้นตอนการประกอบชิ้นส่วน และขั้นตอนการประกอบคอนทิก โดยในขั้นตอนแรกความเป็นไปได้ของการจัดวางชิ้นส่วนเข้าด้วยกันจะถูกกำหนดโดยขั้นตอนวิธีสมิท-วอเตอร์แมน จากนั้นจะใช้ขั้นตอนวิธีระบบอาณานิคม ในการสร้างลำดับการจัดเรียงตัว ซึ่งผลลัพธ์ที่ได้จะถูกเรียกว่า คอนทิก จากนั้นคอนทิกจะถูกนำมาประกอบเข้าด้วยกันโดยอาศัยกฎฮิวริสติกเพื่อนบ้านใกล้เคียงที่สุด จากผลการจำลองพบว่าประสิทธิภาพโดยรวมของเทคนิคที่ใช้วิธี ACS+NNH ให้ผลที่ดีกว่าวิธี NNH นอกจากนี้ยังพบว่าเทคนิคที่ใช้จะให้จำนวนคอนทิกเทียบเท่ากับที่ได้จากโปรแกรม CAP3 และให้ผลเฉลี่ยที่มีคุณภาพสูงกว่าในปัญหาที่มีขนาดใหญ่ จึงได้เสนอแนวทางการรวมขั้นตอนวิธีระบบอาณานิคมเข้ากับโปรแกรม CAP3 เพื่อเป็นปรับปรุงประสิทธิภาพต่อไป

This research presents the use of an ant colony system algorithm (ACS) in DNA fragment assembly. 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 optimisation problem where the aim is to find the right order and orientation of each fragment in the ordering sequence that leads to the formation of a consensus sequence that truly reflects the original DNA strands. In a single-contig problem, there are at least two overlapping fragments covering each location of a base in the ordering sequence. On the other hand, there are no overlapping fragments at certain locations of the ordering sequence in multiple-contig problem. The assembly procedure proposed is composed of two stages: fragment assembly and contig assembly. In the fragment assembly stage, a possible alignment between fragments is determined with the use of a Smith-Waterman algorithm where the fragment ordering sequence is created using the ACS algorithm. The resulting contigs are then assembled together using a nearest neighbour heuristic (NNH) rule. The results indicate that in overall the performance of the combined ACS+NNH technique is superior to that of the NNH search. Furthermore, the solutions produced by both CAP3 and proposed techniques contain a comparable number of contigs while the quality of the combined ACS+NNH solutions is higher than that of the CAP3 solutions when the problem size is large. A recommendation for combining the CAP3 program with the ACS algorithm has also been given.