

วัชรพัช เมตตานันท์ 2552: อัลกอริทึมแบบเวลาเชิงเส้นสำหรับปัญหาการคัดลอกหลาย
ยีน ปริญาวิทยาสตรมหาบัณฑิต (วิศวกรรมคอมพิวเตอร์) สาขาวิศวกรรม
คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์ อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก:
ผู้ช่วยศาสตราจารย์จิตรีทัศน์ ฝึกเจริญผล, Ph.D. 36 หน้า

การสร้างต้นไม้วิวัฒนาการ เป็นหนึ่งในเป้าหมายที่สำคัญที่สุดในการวิจัยทางชีววิทยา
ด้านที่เกี่ยวข้องกับการวิวัฒนาการ ในวิทยานิพนธ์ฉบับนี้ เราสนใจประเด็นหนึ่งเกี่ยวกับการ
พิจารณาความสัมพันธ์ในเชิงวิวัฒนาการ โดยดูจากความสัมพันธ์ของยีน ในรูปแบบที่เรียกว่า
ต้นไม้ยีน ซึ่งยีนแต่ละชนิดจะให้ความสัมพันธ์ระหว่างสิ่งมีชีวิตแตกต่างกัน ความขัดแย้งตรงนี้
เกิดได้จากกระบวนการที่เรียกว่า การคัดลอกยีน ซึ่งเป็นกระบวนการที่ผิดปกติ Guigo, Muchnik,
และ Smith ได้นิยามปัญหาการคัดลอกหลายยีนขึ้น ซึ่งเป็นการหาจำนวนการคัดลอกหลายยีนที่
น้อยที่สุดที่อธิบายความขัดแย้งของต้นไม้สปีชีส์และต้นไม้ยีนทั้งหมดได้ เร็วๆนี้ Bansal และ
Eulenstein ได้เสนออัลกอริทึมแรกที่สามารถแก้ปัญหาคัดลอกหลายยีนนี้ได้ ซึ่งอัลกอริทึมนั้น
ใช้เวลาทำงานเป็นกำลังสองของขนาดข้อมูลเข้า ในวิทยานิพนธ์ฉบับนี้เราได้เสนออัลกอริทึมใหม่
สำหรับปัญหาการคัดลอกหลายยีน ซึ่งเป็นอัลกอริทึมแรกที่ใช้เวลาทำงานเชิงเส้นบนขนาดข้อมูล
เข้า และยังได้เสนออัลกอริทึมที่ใช้เวลาทำงานใกล้เคียงเชิงเส้น แต่นำไปใช้งานจริงได้ง่ายกว่า ซึ่ง
เมื่อทำการทดลอง ได้ผลว่าอัลกอริทึมที่ใช้เวลาใกล้เคียงเชิงเส้นนี้ยังทำงานได้เร็วกว่าอัลกอริทึม
เดิมอย่างน้อย 2 เท่า

Vacharapat Mettanant 2009: A Linear – Time Algorithm for the Multiple Gene Duplication Problem. Master of Engineering (Computer Engineering), Major Field: Computer Engineering, Department of Computer Engineering. Thesis Advisor: Assistant Professor Jittat Fackcheroenphol, Ph.D. 36 pages.

Reconstructing the Tree of Life is one of the most important research goals in evolutionary biology. In this thesis, we consider one of the issues regarding inconsistency among the trees reconstructed using sequences from different genes, called gene trees. This inconsistency is often a result of gene duplication. Given a species tree and a set of gene trees, the Multiple Gene Duplication Problem as formulated by Guigo, Muchnik, and Smith asks for an assignment of gene duplication events on the species tree node to minimize the number of duplication episodes. Recently, Bansal and Eulenstein give the first efficient algorithm for solving that problem. Their algorithm runs in quadratic time. In this thesis we give the first linear-time algorithm for the problem. Our algorithm can be seen as a faster implementation of Bansal and Eulenstein's method. We also present a more practical algorithm that runs in near-linear time. Experiments on synthetic data show a very good improvement on the running time (at least twice faster).