

Thesis Title	Conditional Random Field for Regulatory Regions Prediction in Rice Genome	
Author	Ms. Piyachat Udomwong	
Degree	Master of Science (Bioinformatics)	
Thesis Advisory Committee	Dr.Vannajan Sanghiran Lee	Advisor
	Assoc. Prof. Dr. Jeerayut Chaijaruwanich	Co-advisor
	Assoc. Prof. Dr. Somboon Anuntalabhochai	Co-advisor

ABSTRACT

For the regulation of gene expression at transcription level in rice genome, the regulatory elements play their role as triggers of transcription initiation. Identification of their locations on DNA sequence builds up comprehension of the regulation of gene expression in the rice genome. According to an effort toward the comprehension, the sequenced rice genome has required the annotation combined between computational and laboratorial methods. This study proposes a computational model named as conditional random field to predict different regulatory elements in rice genome based on their specific characteristics. The assumptions of this proposed method are the location of regulatory element regions associating with their neighboring regions and their biochemical properties consisting of types of base, bond strength, and radical content. This study investigates CpG Islands, core promoters and *cis*-regulatory

elements. A conditional random field-based model distinguishes CpG Island class I genes from other classes in the rice genome. It is applied to locate core promoters consisting of TATA and CCAAT boxes in plants. A hybrid framework of expectation maximization (EM) and conditional random field-based methods is proposed to discover and search for shared *cis*-regulatory elements of co-regulated genes in rice genome. The empirical results of each regulatory element measured by statistical measurements present not only the existing relationship between regulatory element regions and their neighboring regions but also improved performance of prediction when biochemical properties are integrated with the models.

ชื่อเรื่องวิทยานิพนธ์

คอนดิชันนอลแรนดอมฟีลด์เพื่อการทำนายบริเวณรีกูลาร์
ทอริในจีโนมข้าว

ผู้เขียน นางสาว ปิยะฉัตร

อุดมวงษ์

ปริญญา วิทยาศาสตร์มหาบัณฑิต (ชีวสารสนเทศศาสตร์)

คณะกรรมการที่ปรึกษาวิทยานิพนธ์ ดร.

วรรณจันทร์ แสงหิรัญ ติ

อาจารย์ที่ปรึกษาหลัก

รศ.ดร. จิรยุทธ ไชยจารุณิช

อาจารย์ที่ปรึกษาร่วม

รศ.ดร. สมบูรณ์ อนันตลาโภชัย

อาจารย์ที่ปรึกษาร่วม

บทคัดย่อ

การควบคุมการแสดงออกของยีนในระดับการ คัดลอกรหัสของจีโนมข้าว นั้น รีกูลาร์ทอริอีลี-
เมนต์มีบทบาทสำคัญต่อการเป็นจุดเริ่มต้นของกระบวนการถอดรหัส ดังนั้นการระบุบริเวณของรีกู-
ลาร์ทอริอีลีเมนต์บนลำดับสายดีเอ็นเอ จึงเป็นการสร้างความเข้าใจเกี่ยวกับการ คัดลอกรหัสในจีโนม
ข้าวเนื่องมาจากความพยายามที่จะทำความเข้าใจ เกี่ยวกับการ คัดลอกรหัสนี้เอง จึงได้นำลำดับสายดี-
เอ็นเอที่ได้รับการตรวจสอบ ลำดับเบสแล้วมาทำการระบุบริเวณของ รีกูลาร์ทอริอีลีเมนต์ ซึ่งขั้นตอน
นั้นก็ประกอบไปด้วย กระบวนการเชิงการคำนวณและกระบวนการในห้องปฏิบัติการ วิทยานิพนธ์
ฉบับนี้ได้เสนอแบบจำลองเชิงการคำนวณบนพื้นฐานของคอนดิชันนอลแรนดอมฟีลด์ เพื่อทำนายรีกู-
ลาร์ทอริอีลีเมนต์ต่างชนิดกัน ตามลักษณะเฉพาะ ในจีโนมข้าว สมมติฐานของการศึกษาวิจัยนี้คือ
บริเวณ รีกูลาร์ทอริ อีลีเมนต์ มีความสัมพันธ์กับบริเวณรอบข้าง และคุณสมบัติชีวเคมี โดยคุณสมบัติ
ดังกล่าวประกอบด้วย ชนิดของเบส, ความแข็งแรงของพันธะ และ เรดิคอลคอนเทนท์ การศึกษาวิจัยนี้
ได้แสดงให้เห็นถึงการพัฒนา ระเบียบวิธีการเรียนรู้บนพื้นฐานของคอนดิชันนอลแรนดอมฟีลด์ ให้มี
ความสอดคล้องกับ สมมติฐานที่ตั้งไว้ การศึกษาวิจัยได้ทำการตรวจสอบ บริเวณของรีกูลาร์ทอริอีลี-
เมนต์ในจีโนมข้าว 3 ชนิด ซึ่งประกอบไปด้วย ซีพีจีไอแลนด์, คอร์โปรโมเตอร์และ ซีต-รีกูลาร์ทอริอี-
ลีเมนต์ การทดลองได้ออกแบบเพื่อให้ตรวจสอบความถูกต้องของการทำนายบริเวณ รีกูลาร์ทอริอีลี-
เมนต์ และความ ความสัมพันธ์ที่มึนัย กับคุณสมบัติทาง ชีวเคมี ในบริเวณดังกล่าว หลังจากนั้นได้

ประเมินประสิทธิภาพของการทำนาย ด้วยวิธีการวัดประสิทธิภาพทางสถิติ ซึ่งไม่เพียงแต่แสดงให้เห็นถึงความที่มีนัยระหว่าง บริเวณ รีกูลาร์ทอรี อีลีเมนต์ และคุณสมบัติทาง ชีวเคมี แต่ยังคงปรับปรุงประสิทธิภาพของการทำนายด้วยเมื่อรวมคุณสมบัติทางชีวเคมีเข้าไปกับแบบจำลอง



ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved