



การวิเคราะห์แผนที่ความสัมพันธ์ยื่น *Plant U-box*
ที่เกี่ยวข้องกับลักษณะน้ำตาลในอ้อย

โดย

นางสาวกฤติยา นิลแก้ว

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต (เทคโนโลยีชีวภาพ)
สาขาวิชาเทคโนโลยีชีวภาพ
คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์
ปีการศึกษา 2563
ลิขสิทธิ์ของมหาวิทยาลัยธรรมศาสตร์

การวิเคราะห์แผนที่ความสัมพันธ์ใน *Plant U-box*
ที่เกี่ยวข้องกับลักษณะน้ำตาลในอ้อย

โดย

นางสาวกฤติยา นิลแก้ว



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต (เทคโนโลยีชีวภาพ)
สาขาวิชาเทคโนโลยีชีวภาพ
คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์
ปีการศึกษา 2563
ลิขสิทธิ์ของมหาวิทยาลัยธรรมศาสตร์

ASSOCIATION MAPPING OF *PLANT U-BOX GENE* FOR SUGAR
RELATED TRAITS IN SUGARCANE

BY

MISS KRITTIYA NINKAEO



A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF SCIENCE (BIOTECHNOLOGY)

DEPARTMENT OF BIOTECHNOLOGY
FACULTY OF SCIENCE AND TECHNOLOGY
THAMMASAT UNIVERSITY

ACADEMIC YEAR 2020

COPYRIGHT OF THAMMASAT UNIVERSITY

มหาวิทยาลัยธรรมศาสตร์
คณะวิทยาศาสตร์และเทคโนโลยี

วิทยานิพนธ์

ของ

นางสาวกฤติยา นิลแก้ว

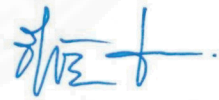
เรื่อง

การวิเคราะห์แผนที่ความสัมพันธ์ใน Plant U-box ที่เกี่ยวข้องกับลักษณะน้ำตาลในอ้อย

ได้รับการตรวจสอบและอนุมัติ ให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต (เทคโนโลยีชีวภาพ)

เมื่อ วันที่ 15 ธันวาคม พ.ศ. 2563

ประธานกรรมการสอบวิทยานิพนธ์



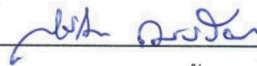
(รองศาสตราจารย์ ดร.ธีระชัย ชนานันต์)

กรรมการและอาจารย์ที่ปรึกษาวิทยานิพนธ์



(รองศาสตราจารย์ ดร.กิตติพัฒน์ อุโฆษกิจ)

กรรมการสอบวิทยานิพนธ์



(ดร.สิทธิโชค ตั้งภัสสรเรื่อง)

คณบดี



(รองศาสตราจารย์ ดร.ณัฐชนนท์ หงส์วรสิทธิ์)

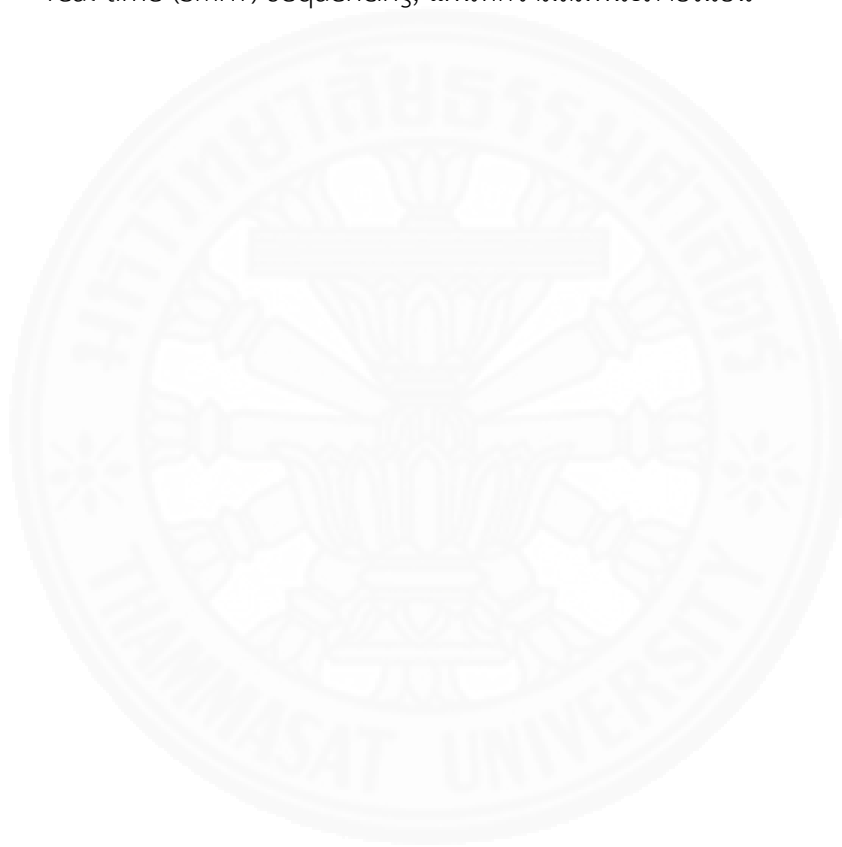
หัวข้อวิทยานิพนธ์	การวิเคราะห์แผนที่ความสัมพันธ์ยีน <i>Plant U-box</i> ที่เกี่ยวข้องกับลักษณะน้ำตาลในอ้อย
ชื่อผู้เขียน	นางสาวกฤติยา นิลแก้ว
ชื่อปริญญา	วิทยาศาสตรมหาบัณฑิต (เทคโนโลยีชีวภาพ)
สาขาวิชา/คณะ/มหาวิทยาลัย	สาขาวิชาเทคโนโลยีชีวภาพ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์	รองศาสตราจารย์ ดร.กิตติพัฒน์ อุโฆษกิจ
ปีการศึกษา	2563

บทคัดย่อ

แผนที่ความสัมพันธ์ภายในยีนเป็นเครื่องมือที่มีประสิทธิภาพในการศึกษาพันธุกรรมของลักษณะที่เกี่ยวข้องกับน้ำตาลในอ้อย งานวิจัยนี้ศึกษาความไม่สมดุลลิงเกจ (linkage disequilibrium; LD) และแผนที่ความสัมพันธ์ของยีน *U-box domain-containing protein35 (SaPUB35)* ที่เกี่ยวข้องกับลักษณะน้ำตาล ในตัวอย่างอ้อยที่มีความหลากหลายทางพันธุกรรมจำนวน 141 ตัวอย่าง ประกอบด้วยอ้อยพันธุ์ปรับปรุงจำนวน 73 ตัวอย่างและอ้อยที่เป็นเชื้อพันธุกรรมนำเข้าจำนวน 68 ตัวอย่าง เทคโนโลยี PacBio single molecule real-time (SMRT) sequencing สามารถระบุเครื่องหมายสลับ (single nucleotide polymorphism; SNP) ได้จำนวน 126 SNPs และอินเดล (Insertion/Deletion; InDel) จำนวน 50 InDels ภายในยีน *SaPUB35* ซึ่งมีขนาด 5731 คู่เบส โดยพบว่าความแปรปรวนลำดับเบสในประชากรอ้อยเชื้อพันธุกรรมนำเข้ามากกว่าประชากรอ้อยพันธุ์ปรับปรุง การวิเคราะห์ LD ระหว่างคู่เครื่องหมายโดยใช้เกณฑ์ค่ากำลังสองของค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างเครื่องหมายสองตำแหน่งที่ 0.1 พบว่าในประชากรพันธุ์ปรับปรุงมีขอบเขต LD 1500 คู่เบส แต่มีความเสื่อมลงไปที่ 1400 คู่เบสในประชากรเชื้อพันธุกรรมนำเข้า แสดงเป็นนัยถึงอิทธิพลของ selective sweep ซึ่งคัดเลือกการกลายที่มีผลดีในระหว่างกระบวนการปรับปรุงพันธุ์อ้อย เครื่องหมาย EST-SSR และ ILP ถูกนำมาใช้ในการประเมินโครงสร้างประชากร (population structure; Q) และความเป็นเครือญาติ (relative kinship; K) เพื่อควบคุมผลบวกแฝงในการวิเคราะห์แผนที่ความสัมพันธ์ระหว่างเครื่องหมายกับลักษณะที่เกี่ยวข้องกับน้ำตาล พบว่าหลังจากควบคุม Q และ K แล้วสามารถระบุเครื่องหมาย SNP จำนวน 2 SNPs และเครื่องหมาย InDel จำนวน 2 InDels จากบริเวณอินตรอนบนยีน 4 อินตรอนที่มีความสัมพันธ์อย่างมีนัยสำคัญกับ

ลักษณะที่เกี่ยวข้องกับน้ำตาลทั้ง 4 ลักษณะในอ้อยปลูกและอ้อยต่อจากทั้งสองแปลงการทดลอง ผลการวิจัยแสดงให้เห็นว่ารูปแบบแอลลีลในอินตรอนของยีน *SaPUB35* มีความเกี่ยวข้องกับลักษณะที่เกี่ยวข้องกับน้ำตาลอย่างมาก และยังมีศักยภาพในการเป็นเครื่องหมายช่วยคัดเลือกในโครงการปรับปรุงพันธุ์อ้อย

คำสำคัญ: อ้อย, เครื่องหมายสลับ, ความไม่สมดุลของลิงเกจ, เทคโนโลยี PacBio single molecule real-time (SMRT) sequencing, แผนที่ความสัมพันธ์ภายในยีน



Thesis Title	ASSOCIATION MAPPING OF <i>PLANT U-BOX</i> GENE FOR SUGAR RELATED TRAITS IN SUGARCANE
Author	Miss Krittiya Ninkaeo
Degree	Master of Science (Biotechnology)
Department/Faculty/University	Biotechnology Faculty of Science and Technology Thammasat University
Thesis Advisor	Associate Professor Kittipat Ukoskit, Ph.D.
Academic Year	2020

ABSTRACT

Candidate gene association mapping provides a powerful tool for dissection of genetic architecture of sugar-related traits in sugarcane. In this study, linkage disequilibrium (LD) and association mapping of *U-box domain-containing protein35* gene (*SaPUB35*) for sugar-related traits were conducted in 141 diverse sugarcane accessions consisting of 73 breeding lines and 68 exotic sugarcane accessions. PacBio single molecule, real-time (SMRT) sequencing was performed to identify 126 single nucleotide polymorphisms (SNPs) and 50 Insertion/deletions (InDels) in 5731 bp of *SaPUB35*. The exotic accessions have more variants than the breeding line accessions. Pairwise LD analysis revealed that mean LD, at a threshold of squared allele frequency correlations (r^2) = 0.1, extends to at least 1500 bp in the breeding line accessions but is eroded to 1400 bp in exotic accessions. This implies a signature of a selective sweep for a beneficial mutation during selection processes of sugarcane breeding. Subsequently, the population structure (Q) and relative kinship (K) were determined by genotyping with EST-SSR and ILP markers. After controlling Q and K, two SNPs and two InDels in four introns were significant associated with four sugar related traits in the planted cane crop and the first ratoon in two locations. The results demonstrated that allelic variations of the introns in *SaPUB35* was closely related sugar

related traits and have a potential for marker-assisted selection in sugarcane breeding programs.

Keywords: Sugarcane, Single nucleotide polymorphism (SNP), Linkage disequilibrium, PacBio single molecule real-time (SMRT) sequencing, Candidate gene association mapping.



กิตติกรรมประกาศ

วิทยานิพนธ์นี้สำเร็จลุล่วงได้อย่างสมบูรณ์ เนื่องจากความกรุณา และการสนับสนุนของ รองศาสตราจารย์ ดร. กิตติพัฒน์ อุโฆษกิจ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ซึ่งให้ทั้งความรู้ คำแนะนำ และแนวทางในการแก้ปัญหาในทุก ๆ ขั้นตอนของดำเนินงานวิจัย และขอขอบคุณ รองศาสตราจารย์ ดร. ธีระชัย ธนานันต์ และดร. สิทธิโชค ตั้งภัสสรเรือง ซึ่งได้ให้ข้อชี้แนะและคำปรึกษาต่าง ๆ จนกระทั่งวิทยานิพนธ์นี้ดำเนินลุล่วงไปได้ด้วยดี

ขอขอบคุณ บริษัทมิตรผลวิจัย พัฒนาอ้อยและน้ำตาล จำกัด และศูนย์พันธุวิศวกรรมและเทคโนโลยีชีวภาพแห่งชาติ สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ ที่ได้มอบทุนสนับสนุนการดำเนินงานวิจัย

ขอขอบคุณทุนบัณฑิตเรียนดีเพื่อศึกษาต่อระดับบัณฑิตศึกษา คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ ตามบันทึกข้อตกลงเลขที่ ทบ 09/2560

ขอขอบคุณผู้ช่วยสอน มหาวิทยาลัยธรรมศาสตร์ ประจำปี 2561

ขอขอบคุณคณาจารย์สาขาวิชาเทคโนโลยีชีวภาพทุกท่าน ที่ได้ให้ความรู้ คำปรึกษา และคำแนะนำตลอดการศึกษาที่ผ่านมา รวมถึงเจ้าหน้าที่และนักวิทยาศาสตร์ในสาขาวิชาเทคโนโลยีชีวภาพทุกท่านที่คอยช่วยเหลือและสนับสนุนตลอดด้วยดีเสมอมา

สุดท้ายนี้ขอกราบขอบพระคุณ คุณพ่อ คุณแม่ และครอบครัว รวมถึงเพื่อน ๆ รุ่นพี่ รุ่นน้องในสาขาวิชาเทคโนโลยีชีวภาพที่ให้อำลั่งใจและสนับสนุนผู้วิจัยมาโดยตลอด

นางสาวกฤติยา นิลแก้ว

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	(1)
บทคัดย่อภาษาอังกฤษ	(3)
กิตติกรรมประกาศ	(5)
สารบัญตาราง	(9)
สารบัญภาพ	(10)
รายการสัญลักษณ์และคำย่อ	(11)
บทที่ 1 บทนำ	1
1.1 ความสำคัญและที่มา	1
1.2 วัตถุประสงค์	3
1.3 ประโยชน์ที่คาดว่าจะได้รับ	3
บทที่ 2 วรรณกรรมและงานวิจัยที่เกี่ยวข้อง	4
2.1 อ้อย	4
2.1.1 ลักษณะทางพันธุศาสตร์	4
2.1.2 การวิเคราะห์ลำดับเบสจีโนมอ้อย	5
2.2 โปรตีน Plant U-box (PUB)	7
2.3 เครื่องหมายสลับ (single nucleotide polymorphism; SNP)	9
2.3.1 เครื่องหมายสลับในพืชพอลิพลอยด์	11
2.3.2 การพัฒนาเครื่องหมายสลับในอ้อย	12

2.4 การวิเคราะห์ลำดับเบสด้วยเทคโนโลยี PacBio SMRT Sequencing	13
2.4.1 การวิเคราะห์ลำดับเบสในอ้อยด้วยเทคโนโลยี PacBio SMRT Sequencing	16
2.5 แผนที่ความสัมพันธ์ (association mapping)	17
2.5.1 ความไม่สมดุลลิงเกจ (linkage disequilibrium; LD)	18
2.5.2 การวิเคราะห์แผนที่ความสัมพันธ์ในอ้อย	21
2.5.3 การวิเคราะห์แผนที่ความสัมพันธ์ภายในยีน (candidate gene association mapping)	22
บทที่ 3 วิธีการวิจัย	24
3.1 ตัวอย่างอ้อย และการสกัดดีเอ็นเอ	24
3.2 การเก็บข้อมูลพีโนไทป์ และการวิเคราะห์ค่าทางสถิติ	24
3.3 การออกแบบไพรเมอร์	25
3.4 การเพิ่มปริมาณดีเอ็นเอและการทำให้บริสุทธิ์	26
3.4.1 การเพิ่มปริมาณดีเอ็นเอ	26
3.4.2 การทำผลิตภัณฑ์พีซีอาร์ให้บริสุทธิ์	27
3.5 การค้นหาเครื่องหมายสลับ (SNP) และอินเดล (Insertion/Deletion; InDel)	27
3.6 การวิเคราะห์ความไม่สมดุลของลิงเกจ (linkage disequilibrium; LD) และการเสื่อมถอยของความไม่สมดุลของลิงเกจ (LD decay)	27
3.7 การวิเคราะห์แผนที่ความสัมพันธ์ภายในยีน (candidate gene association mapping)	28
บทที่ 4 ผลการวิจัยและอภิปรายผล	29
4.1 การวิเคราะห์ทางสถิติของข้อมูลพีโนไทป์	29
4.2 การวิเคราะห์ลำดับเบสด้วยเทคโนโลยี PacBio SMRT Sequencing	34
4.3 ความแปรปรวนของเครื่องหมาย SNP และ InDel	35

4.4 การไม่สมดุลของลิงเกจ (linkage disequilibrium; LD) และการเสื่อม ถอยของความไม่สมดุลของลิงเกจ (LD decay)	37
4.5 การวิเคราะห์แผนที่ความสัมพันธ์	38
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ	46
รายการอ้างอิง	47
ภาคผนวก	
ภาคผนวก ก	61
ภาคผนวก ข	64
ประวัติผู้เขียน	67



สารบัญตาราง

ตารางที่	หน้า
2.1 ตัวอย่างและหน้าที่ของโปรตีน PUB ที่พบในพืช	9
4.1 ค่าเฉลี่ย ช่วงค่า ค่าความแปรปรวนจีโนไทป์ ค่าความแปรปรวนฟีโนไทป์ และค่าอัตราพันธุกรรมแบบกว้าง (broad-sense heritability; H^2) ของลักษณะที่เกี่ยวข้องกับลักษณะน้ำตาลทั้ง 4 ลักษณะ ในอ้อยปลูก (plant cane) และอ้อยตอ (ratoon) ของแปลงการทดลองมหาวิทยาลัยขอนแก่น (KKU) และแปลงการทดลองศูนย์นวัตกรรมและการวิจัยมิตรผล (MPT)	31
4.2 ค่าเฉลี่ย ช่วงค่า และค่าความแปรปรวนฟีโนไทป์ ของลักษณะที่เกี่ยวข้องกับลักษณะน้ำตาลทั้ง 4 ลักษณะ ในอ้อยปลูก (plant cane) และอ้อยตอ (ratoon) ระหว่างตัวอย่างอ้อยเชื้อพันธุกรรมนำเข้าและอ้อยพันธุ์พ่อแม่ของอ้อยพันธุ์ปรับปรุงในประเทศไทย	32
4.3 ค่าเฉลี่ย ช่วงค่า และค่าความแปรปรวนฟีโนไทป์ของลักษณะดัชนีความเข้มของสีใบ (leaf color index) และประสิทธิภาพการสังเคราะห์ด้วยแสง (photosynthetic efficiency (fv/fm) ระหว่างตัวอย่างอ้อยเชื้อพันธุกรรมนำเข้าและอ้อยพันธุ์ปรับปรุงในประเทศไทย	33
4.4 ค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันระหว่างลักษณะที่เกี่ยวข้องกับน้ำตาลทั้ง 4 ลักษณะในอ้อยปลูกและอ้อยตอของทั้ง 2 แปลงการทดลอง	33
4.5 จำนวนเครื่องหมาย SNP และ InDel บนยีน SaPUB35 ที่ได้จากการค้นหาด้วยวิธี GATK	36
4.6 เครื่องหมาย SNP และ InDel ที่มีความสัมพันธ์กับลักษณะที่เกี่ยวข้องกับลักษณะน้ำตาลของอ้อยปลูก (plant) และอ้อยตอ (1st ratoon) แปลงการทดลองมหาวิทยาลัยขอนแก่น (KKU) แปลงการทดลองศูนย์นวัตกรรมและการวิจัยมิตรผล (MPT) และรวมแปลงการทดลอง (combined)	43

สารบัญภาพ

ภาพที่	หน้า
2.1 การวิเคราะห์หาลำดับเบสของจีโนมโมโนพลอยด์	6
2.2 เทคโนโลยี PacBio SMRT sequencing	15
2.3 หลักการของการไม่สมดุลของลิงค์เกจ (LD) และความสัมพันธ์ระหว่างขอบเขต LD กับความหนาแน่นของเครื่องหมาย	21
3.1 การออกแบบไพรเมอร์จาก cDNA ของยีน <i>SaPUB35</i>	26
4.1 การกระจายตัวของเครื่องหมาย SNP และ InDel บนยีน <i>SaPUB35</i>	36
4.2 แผนภาพการกระจายตัวของค่า r^2 ระหว่างคู่เครื่องหมาย	38
4.3 การกระจายตัวของค่า p-value ที่ได้จากโมเดลทดสอบ 4 แบบที่ใช้วิเคราะห์ความสัมพันธ์ระหว่างเครื่องหมาย SNP และ InDel กับลักษณะที่เกี่ยวข้องกับน้ำตาลทั้ง 4 ลักษณะของอ้อยปลูก (A-D) และอ้อยตอ (E-H) แปลงการทดลองมหาวิทยาลัยขอนแก่น	41
4.4 การกระจายตัวของค่า p-value ที่ได้จากโมเดลทดสอบ 4 แบบที่ใช้วิเคราะห์ความสัมพันธ์ระหว่างเครื่องหมาย SNP และ InDel กับลักษณะที่เกี่ยวข้องกับศูนย์นวัตกรรมและการวิจัยมิตรผล	41
4.5 การกระจายตัวของค่า p-value ที่ได้จากโมเดลทดสอบ 4 แบบที่ใช้วิเคราะห์ความสัมพันธ์ระหว่างเครื่องหมาย SNP และ InDel กับลักษณะที่เกี่ยวข้องกับน้ำตาลทั้ง 4 ลักษณะของอ้อยปลูก (A-D) และอ้อยตอ (E-H)	42
4.6 แผนภาพ box plot แสดงอิทธิพลของเครื่องหมาย SNP และ InDel ที่สัมพันธ์กับลักษณะที่เกี่ยวข้องกับลักษณะน้ำตาลในอ้อยปลูก	44
4.7 แผนภาพ box plot แสดงอิทธิพลของเครื่องหมาย SNP และ InDel ที่สัมพันธ์กับลักษณะที่เกี่ยวข้องกับลักษณะน้ำตาลในอ้อยตอ	45

รายการสัญลักษณ์และคำย่อ

สัญลักษณ์/คำย่อ	คำเต็ม/คำจำกัดความ
Brix	Total soluble solid
POL	Sucrose content
CCS	Commercial cane sugar
SNP	Single nucleotide polymorphism
InDel	Insertion/Deletion
SSR	Simple sequence repeat
EST	Expressed sequence tag
EST-SSR	Expressed sequence tag-simple sequence repeat
PCR	Polymerase chain reaction
SSR	Simple sequence repeat
GWAS	Genome wide association study
MAS	Marker-assisted selection
MTA	Marker-trait association
GLM	General linear model
MLM	Mixed linear model
LD	Linkage disequilibrium
QTL	Quantitative trait loci
PUB	Plant U-box
HG	Homologous group
bp	Base pair
cM	centimorgan

บทที่ 1

บทนำ

1.1 ความสำคัญและที่มา

อุตสาหกรรมอ้อยและน้ำตาลทรายของไทยมีการเจริญเติบโตอย่างต่อเนื่อง และมีแนวโน้มขยายตัวเพิ่มขึ้น โดยปัจจุบันประเทศไทยเป็นผู้ส่งออกน้ำตาลจัดเป็นอันดับสองของประเทศที่ส่งออกน้ำตาลมากที่สุดในระดับโลกรองจากบราซิล และยังเป็นผู้ส่งออกน้ำตาลที่สำคัญที่สุดในอาเซียนอีกด้วย (สำนักพัฒนาอุตสาหกรรมอ้อย น้ำตาลทรายและอุตสาหกรรมต่อเนื่อง, 2560) แต่จากรายงานของสำนักงานพัฒนาอ้อยและน้ำตาลทราย ประจำปี 2563 พบว่าคุณภาพของน้ำตาลของประเทศไทยนั้นอยู่ที่ 12.68 ซี.ซี.เอส. ซึ่งค่อนข้างต่ำเมื่อเปรียบเทียบกับประเทศคู่แข่ง เพื่อเพิ่มมูลค่าทางเศรษฐกิจและปริมาณผลผลิต จึงต้องมีการปรับปรุงพันธุ์อ้อยเพื่อให้ได้พันธุ์ที่มีปริมาณผลผลิตและความหวานสูง ในการปรับปรุงพันธุ์อ้อยแบบดั้งเดิมนั้นกว่าจะได้พันธุ์อ้อยที่มีลักษณะตามความต้องการใช้เวลาไม่ต่ำกว่า 10 ปี เนื่องจากต้องคัดเลือกลูกผสมจากลักษณะฟีโนไทป์ ต้องใช้ทั้งระยะเวลาในการดำเนินการ พื้นที่เพาะปลูกและค่าใช้จ่ายสูง จึงมีการนำเครื่องหมายดีเอ็นเอมาช่วยในการคัดเลือกลูกผสมที่ได้จากการปรับปรุงพันธุ์ เนื่องจากช่วยลดทั้งระยะเวลาและค่าใช้จ่าย แต่มีประสิทธิภาพและความแม่นยำในการคัดเลือกพันธุ์อ้อยที่มีลักษณะตามที่ต้องการได้มากขึ้น

เครื่องหมายสเนิป (single nucleotide polymorphism; SNP) เป็นเครื่องหมายดีเอ็นเอที่กำลังเป็นที่นิยมศึกษาและพัฒนาเพื่อใช้เป็นเครื่องหมายคัดเลือกในโครงการปรับปรุงพันธุ์พืช ความแปรปรวนของพันธุกรรมนี้เกิดจากการกลายและรีคอมบิเนชัน ทำให้เกิดความแตกต่างของเบสเพียงหนึ่งตำแหน่ง จึงมีความแปรปรวนและการกระจายตัวทั่วจีโนมสูง (Agarwal, Shrivastava, & Padh, 2008) สามารถประยุกต์ใช้กับงานทางด้านปรับปรุงพันธุ์ได้อย่างกว้างขวาง ทั้งการศึกษาความหลากหลายทางพันธุกรรม การวิเคราะห์ความไม่สมดุลลิงเกจ (linkage disequilibrium; LD) และการเสื่อมถอยของความไม่สมดุลลิงเกจ (LD decay) การสร้างแผนที่พันธุกรรม การวิเคราะห์ความสัมพันธ์กับลักษณะฟีโนไทป์ รวมถึงเป็นเครื่องหมายในการคัดเลือก จึงมีการศึกษาและพัฒนา SNP ในพืชหลายชนิด เช่น ข้าวสาลี ข้าวฟ่าง และอ้อย (Alqudah et al., 2020; McIntyre et al., 2015; Nelson et al., 2011; Parida et al., 2016)

วิธีการพัฒนาเครื่องหมายดีเอ็นเอภายในยีนที่สัมพันธ์กับลักษณะที่สนใจเรียกว่า candidate gene association mapping เป็นการตรวจสอบพอลิมอร์ฟิซึมภายในยีนที่สนใจซึ่งเป็นยีนที่มีความเกี่ยวข้องหรือส่งผลกับการแสดงออกของลักษณะที่สนใจ เพื่อระบุเครื่องหมายดีเอ็นเอที่เป็นสาเหตุ

ของความแปรปรวนของพีโนไทป์ โดยอาศัยพื้นฐานของความไม่สมดุลลิงเกจหรือ LD ซึ่งเป็นความสัมพันธ์แบบไม่สุ่มระหว่างแอลลีลที่อยู่ต่างตำแหน่งกัน เพื่อตรวจสอบการมี LD กันระหว่างเครื่องหมายดีเอ็นเอกับแอลลีลภายในยีนที่ควบคุมลักษณะปริมาณว่ามีผลต่อการแสดงออกของลักษณะพีโนไทป์หรือไม่ จึงถือได้ว่าเป็นวิธีการศึกษาลักษณะที่สนใจอย่างละเอียด และส่วนใหญ่มักใช้เครื่องหมาย SNP เนื่องจากเป็นเครื่องหมายดีเอ็นเอที่มีความหนาแน่นและกระจายตัวทั่วจีโนม ทำให้การตรวจสอบความสัมพันธ์ระหว่างแอลลีลของยีนกับลักษณะที่สนใจมีประสิทธิภาพมากขึ้น อีกทั้งการศึกษาเครื่องหมาย SNP เป็นเรื่องที่สะดวกขึ้น เนื่องจากการพัฒนาของเทคโนโลยีการวิเคราะห์ลำดับเบสยุคใหม่ หนึ่งในเทคโนโลยีนี้คือเทคโนโลยี PacBio SMRT sequencing ซึ่งเป็นเทคโนโลยีที่วิเคราะห์ดีเอ็นเอธรรมชาติโมเลกุลเดี่ยวแบบตามเวลาจริง โดยตรวจจับการเข้าสู่สายของนิวคลีโอไทด์ที่ติดฉลากฟลูออเรสเซนต์ ขณะที่เอนไซม์ดีเอ็นเอพอลิเมอเรสกำลังสังเคราะห์ดีเอ็นเอในหลุมขนาดเล็บบนแผ่นเซลล์ SMRT ข้อดีของเทคโนโลยีนี้คือสามารถวิเคราะห์ลำดับเบสสายยาวเฉลี่ยได้มากถึง 10-18 Kb และเนื่องจากดีเอ็นเอต้นแบบเป็นแบบวงแหวนทำให้สามารถวิเคราะห์ลำดับเบสวนซ้ำได้ 10-15 รอบของโมเลกุลเดิม ช่วยลดอัตราการคลาดเคลื่อนของการวิเคราะห์ลำดับเบสได้

จากการศึกษาก่อนหน้าของ Khanbo (2020) ในการพัฒนาเครื่องหมาย Intron length polymorphisms (ILP) จากยีนในกลุ่ม Plant U-box (*PUB*) พบว่ามีเครื่องหมาย ILP ที่มีความสัมพันธ์กับค่าความเข้มข้นของของแข็งที่ละลายอยู่ในน้ำอ้อย (Brix) ซึ่งเป็นลักษณะที่มีความเกี่ยวข้องกับลักษณะน้ำตาลในอ้อย (กำลังเตรียมต้นฉบับตีพิมพ์) จากการศึกษาเชิงเปรียบเทียบยีนนี้ในจีโนมของข้าวฟ่างซึ่งเป็นพืชที่มีบรรพบุรุษร่วมกันกับอ้อย (Jannoo et al., 2007) พบว่ามีความคล้ายคลึงกับยีน *U-box domain-containing protein 35* ในข้าวฟ่าง โปรตีน U-box ประกอบด้วยโดเมน U-box ที่ด้านปลายคาร์บอกซิลและโดเมน serine/threonine kinase ที่ด้านปลายอะมิโน จึงจัดอยู่ในคลาส IV ของโปรตีน PUB family ในปัจจุบันมีการทำนายโปรตีน PUB และศึกษาหน้าที่ของโปรตีน PUB ในพืชชั้นสูงหลายชนิด พบว่ามีความเกี่ยวข้องกับหลายกระบวนการ เช่นการตอบสนองต่อความเครียดทั้งแบบ abiotic (ความแล้ง อุณหภูมิ และความเค็ม) (Byun et al., 2017; Wang et al., 2020) และ biotic (เชื้อก่อโรค และเชื้อรา) (Ryu et al., 2019; Wang et al., 2015) การแบ่งและขยายขนาดของเซลล์ การผสมตัวเองไม่ติด การพัฒนาละอองเรณู รวมถึงอาจจะมีข้องเกี่ยวกับกระบวนการเมแทบอลิซึมของแป้งและน้ำตาลด้วย (Chen et al., 2019; Zhang, Wang, Wang, Wang, & Gao, 2020)

งานวิจัยนี้จึงทำการวิเคราะห์ความไม่สมดุลของลิงเกจจากเครื่องหมาย SNP และ InDel ภายในยีน *U-box domain-containing protein 35* (*SaPUB35*) ของอ้อยด้วยเทคโนโลยี PacBio SMRT sequencing เพื่อทำความเข้าใจในความไม่สมดุลลิงเกจภายในยีนที่เกี่ยวข้องกับลักษณะ

น้ำตาลในอ้อย อันเกิดจากการคัดเลือกในโครงการปรับปรุงพันธุ์อ้อย และวิเคราะห์แผนที่ความสัมพันธ์ระหว่างเครื่องหมายดีเอ็นเอทั้งสองชนิดภายในยีน *SaPUB35* กับลักษณะที่เกี่ยวข้องกับความหวานในอ้อยด้วยวิธี candidate gene association mapping

1.2 วัตถุประสงค์

1.2.1 ระบุเครื่องหมาย SNP และ InDel ภายในยีน *SaPUB35* ที่เกี่ยวข้องกับลักษณะน้ำตาล โดยใช้เทคโนโลยี Pacbio SMRT sequencing

1.2.2 วิเคราะห์ความไม่สมดุลลิงเกจและการเสื่อมถอยของความไม่สมดุลลิงเกจภายในยีน *SaPUB35* ที่เกี่ยวข้องกับลักษณะน้ำตาล

1.2.3 วิเคราะห์แผนที่ความสัมพันธ์ความสัมพันธ์ระหว่างเครื่องหมาย SNP และ InDel ภายในยีน *SaPUB35* ที่เกี่ยวข้องกับลักษณะน้ำตาลกับลักษณะน้ำตาลในอ้อย

1.3 ประโยชน์ที่คาดว่าจะได้รับ

ได้เครื่องหมายดีเอ็นเอที่มีความสัมพันธ์กับลักษณะน้ำตาลในอ้อย ที่จะสามารถช่วยในการคัดเลือกพันธุ์อ้อยที่มีปริมาณผลผลิตและความหวานสูงจากกระบวนการปรับปรุงพันธุ์อ้อยได้

บทที่ 2

วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

2.1 อ้อย

2.1.1 ลักษณะทางพันธุศาสตร์

อ้อย (*Saccharum* spp. L.) จัดอยู่ในวงศ์หญ้า (Poaceae) เช่นเดียวกับข้าว ข้าวโพด และข้าวฟ่าง พันธุ์อ้อยสมัยใหม่เกิดจากการผสมระหว่างอ้อยพันธุ์ *S. officinarum* กับอ้อยพันธุ์ *S. spontaneum* โดยที่ *S. officinarum* เป็นอ้อยพันธุ์ปลูกที่มีความสำคัญและมีการปลูกอย่างกว้างขวาง เนื่องจากความสามารถในการสะสมน้ำตาลซูโครสไว้ที่ลำต้น มีความหวานสูง ลำต้นมีขนาดใหญ่ (มีขนาดเส้นผ่าศูนย์กลางมากกว่า 3 เซนติเมตร) มีเส้นใยน้อย แต่เป็นอ้อยที่ต้องปลูกในสภาพภูมิอากาศที่จำเพาะ มีความต้องการน้ำและดินที่อุดมสมบูรณ์ อีกทั้งยังไม่มีความต้านทานโรค ขณะที่ *S. spontaneum* เป็นพันธุ์ป่าที่มีความแข็งแรง มีประสิทธิภาพในการแตกกอปรับตัวเข้ากับสภาพแวดล้อมที่ไม่เหมาะสมได้ดี อีกทั้งยังมียีนที่เกี่ยวข้องกับการต้านทานโรคและความเครียดอีกด้วย (Cheavegatti-Gianotto et al., 2011)

S. officinarum เป็นอ้อยพันธุ์ที่มีจีโนมเป็นแบบ auto-octaploid มีโครโมโซม 80 แท่ง ($2n = 8x = 80$) ขณะที่ *S. spontaneum* เป็นอ้อยพันธุ์ที่มีจีโนมเป็นแบบ autopolyploid มีโครโมโซมตั้งแต่ 40 -128 แท่ง ($2n = 4x-12x = 40-128$) (Raboin et al., 2006) ในการผสมข้ามระหว่างอ้อย 2 พันธุ์นี้เพื่อทำให้ *S. officinarum* ได้รับยีนต้านทานโรคจาก *S. spontaneum* หลังจากผสมแล้วนำลูกผสมรุ่น F1 ผสมกลับไปยัง *S. officinarum* หลาย ๆ ชั่วรุ่น เพื่อทำให้ได้พันธุ์อ้อยที่มีระดับน้ำตาลสูง มีความทนทานต่อโรคเพิ่มขึ้น เพิ่มผลผลิต เพิ่มความสามารถในการไว้ตอและเพิ่มความสามารถในการปรับตัวต่อสภาพแวดล้อมที่ไม่เหมาะสม (D'Hont, 2005) ส่งผลให้จีโนมของพันธุ์อ้อยสมัยใหม่มีความซับซ้อน โครโมโซมเป็นแบบออโตโพลีพลอยดี (autopolyploidy) และอนิวพลอยดี (aneuploidy) มีจำนวนโครโมโซมตั้งแต่ 100-130 แท่ง และจีโนมมีขนาดประมาณ 10,000 Mb (R570, $2n=115$) (D'Hont & Glaszmann, 2001)

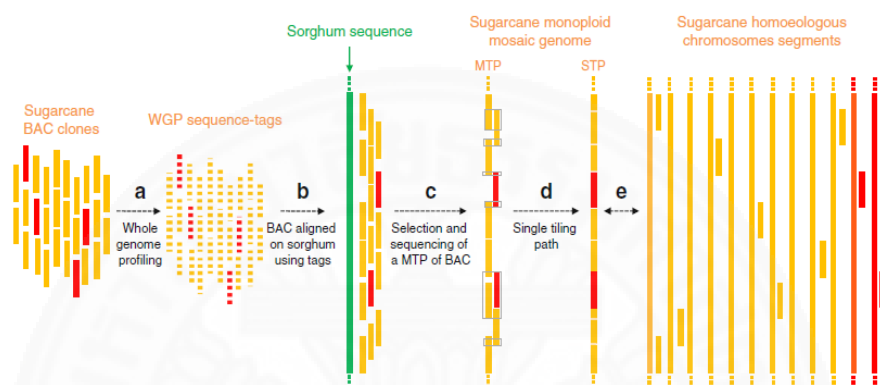
จากการศึกษาโครงสร้างจีโนมของพันธุ์อ้อยสมัยใหม่ของ D'Hont et al. (1996) ด้วยเทคนิค *in situ* hybridization จากการใช้เทคนิค genomic *in situ* hybridization (GISH) ของอ้อยพันธุ์ R570 ($2n \approx 115$) พบว่าจีโนมประกอบด้วยโครโมโซมของ *S. officinarum* 80%, *S. spontaneum* 10% และ recombinant chromosomes 10% และในปี 2005 (D'Hont, 2005)

ทำการศึกษาคู่มือสร้างจีโนมของอ้อยพันธุ์ NCo376 ($2n \approx 112$) พบว่าประกอบด้วยโครโมโซมของ *S. officinarum* 70%, *S. spontaneum* 20% และ recombinant chromosomes 10%

2.1.2 การวิเคราะห์ลำดับเบสจีโนมอ้อย

เนื่องจากอ้อยและข้าวฟ่าง (*Sorghum bicolor*) มีบรรพบุรุษร่วมกันเมื่อประมาณ 8-9 ล้านปีก่อน (Jannoo et al., 2007) จึงทำให้มีวิวัฒนาการและความสัมพันธ์ทางเครือญาติที่ใกล้ชิดกัน อีกทั้งขนาดจีโนมโมโนพลอยด์ของอ้อย (730 Mb) มีขนาดใกล้เคียงกับจีโนมของข้าวฟ่าง (800-900 Mb) จากจีโนมข้าวฟ่างที่มีขนาดเล็กทำให้สามารถหาลำดับเบสของทั้งจีโนมและสามารถใช้เป็นลำดับเบสต้นแบบในการประกอบลำดับเบสของอ้อยได้ (Garsmeur et al., 2018) จึงมีการวิเคราะห์ลำดับเบสจีโนมของอ้อยโดยใช้จีโนมข้าวฟ่างเป็นจีโนมอ้างอิง โดย Okura, de Souza, de Siqueira Tada, and Arruda (2016) วิเคราะห์หาลำดับเบสและประกอบลำดับเบส (assembled) ของโคลน bacteria artificial chromosome (BAC) ของอ้อยพันธุ์ SP80-3280 จำนวน 178 โคลนด้วยเทคโนโลยี Illumina HiSeq2000 และ PacBio platforms และจากการเปรียบเทียบกับลำดับเบสโครโมโซมของข้าวฟ่าง พบว่าสามารถประกอบลำดับเบสได้ความยาว 19 Mb Vilela et al. (2017) ใช้วิธีจีโนมิกส์เชิงเปรียบเทียบ (comparative genomics) ในการศึกษาการเรียงลำดับของยีน *LFY* (Leafy) ยีน *PHYC* (phytochrome C) และยีน *TOR Kinase* (target of rapamycin) ซึ่งเป็นยีนหลักที่เกี่ยวข้องกับการพัฒนาของพืช โดยหาลำดับเบสของ homo/homeologous BAC จำนวน 27 โคลนในอ้อยพันธุ์ R570 แล้วเปรียบเทียบการเรียงลำดับของยีนบนโครโมโซมกับ orthologous region ของข้าวฟ่างและข้าว พบว่ายีน *TOR Kinase* เป็นยีนที่ไม่มีการเรียงลำดับคล้ายกับการเรียงลำดับของยีนบนโครโมโซมของข้าวฟ่างและข้าว ยีน *LFY* มีการเรียงลำดับคล้ายกับการเรียงลำดับของยีนบนโครโมโซมของข้าวฟ่าง ขณะที่ยีน *PHYC* มีการอนุรักษ์องค์ประกอบของยีนและลำดับการเรียงตัวของยีนสูงระหว่างอ้อย ข้าวฟ่างและข้าว และ Garsmeur et al. (2018) ศึกษาโครงสร้างจีโนมโมโนพลอยด์ (monoploid genome) ในอ้อยพันธุ์ R570 จากการเปรียบเทียบโคลน BAC ของอ้อยพันธุ์ R570 กับจีโนมของข้าวฟ่าง เพื่อระบุ minimum tiling path (MTP) โดยใช้เทคโนโลยี whole genome profiling (WGPTM technology) พบว่ามี MTP จำนวน 4660 โคลนที่ซ้อนทับกับบริเวณที่มียีนมากที่สุด (gene-rich part) บนจีโนมข้าวฟ่าง จากการตัดแต่งปลายของ MTP เพื่อสร้าง single tiling path (STP) แล้วประกอบลำดับเบส STP (ดังแสดงในภาพที่ 2.1) พบว่าสามารถประกอบได้ความยาวลำดับเบสรวม 382 Mb และในปีเดียวกันนั้น Zhang et al. (2018) ทำการวิเคราะห์หาลำดับเบสแฮพลอยด์ของ *S. spontaneum*, AP85-441 จากการระบุตำแหน่งของยีนบนโครโมโซมคู่เหมือนสามารถระบุตำแหน่งยีนได้จำนวน 35,525 ยีน

จากการติดตามการเปลี่ยนตำแหน่งของโครโมโซม 4 แห่ง พบว่าการลดจำนวนโครโมโซมพื้นฐานจาก 10 แห่ง เป็น 8 แห่ง ของ *S. spontaneum* เกิดจากการรวมตัวของโครโมโซมบรรพบุรุษคู่เหมือนจำนวน 2 แห่งของข้าวฟ่าง คือโครโมโซมแห่งที่ 5 และแห่งที่ 8 ซึ่งโครโมโซมของ *S. spontaneum* ในอ้อยสมัยใหม่มีการกระจายตัวแบบสุ่มในจีโนมของ AP85-441 แสดงให้เห็นถึงการรวมตัวแบบสุ่มระหว่างโครโมโซมคู่เหมือนใน *S. spontaneum*



ภาพที่ 2.1 การวิเคราะห์หาลำดับเบสของจีโนมโมโนพลอยด์โดยอาศัยพื้นฐานของการอนุรักษ์ลำดับการเรียงตัวของยีนที่คล้ายกัน (synteny) และลำดับการเรียงตัวของยีนเหมือนกันในโครโมโซมเดียวกัน (collinearity) ภายในโครโมโซมคู่เหมือนของอ้อยและข้าวฟ่าง a) สร้าง whole genome profiling (WGP) sequence tags จากโคลน BACs อ้อยพันธุ์ R570 b) เปรียบเทียบ WGP sequence tags กับลำดับเบสของข้าวฟ่าง เพื่อระบุตำแหน่งของ BACs ของอ้อยบนจีโนมข้าวฟ่าง c) กำหนดและวิเคราะห์หาลำดับเบสของ minimum tiling path ของ BAC (MTP) ที่สอดคล้องกับจีโนมโมโนพลอยด์ของอ้อย d) ตัดแต่งส่วนที่เหลือของลำดับเบสของโคลน BAC เพื่อสร้าง single tiling path (STP) e) คอนทิกของ BAC บน STP sequence อยู่ในโฮมิโอโลกัสโครโมโซมที่แตกต่างกันสี่สีและสีแดงแทนโครโมโซมของ *S. officinarum* และ *S. spontaneum* ตามลำดับ จาก “A mosaic monoploid reference sequence for the highly complex genome of sugarcane,” โดย O. Garsmeur, G. Droc, R. Antonise, J. Grimwood, B. Potier, K. Aitken, J. Jenkins, G. Martin, C. Charron, H. Catherine, et al., 2018, *Nature communications*, 9(1), p. 3.

นอกจากการศึกษาและวิเคราะห์ลำดับเบสของจีโนมแล้ว ยังมีการวิเคราะห์ลำดับเบสของจีโนมในไมโทคอนเดรียในอ้อย โดย Shearman et al. (2016) ใช้เทคโนโลยี PacBio RSII ในการหาลำดับเบสและประกอบลำดับเบสจีโนมของไมโทคอนเดรียของอ้อยพันธุ์ขอนแก่น 3 (KK3)

พบว่าจากการประกอบลำดับเบสที่มีขนาดมากกว่า 30 Kb ประกอบด้วย 20 คอนทิก (contig) ซึ่งรวมส่วนคอนทิกของไมโทคอนเดรียจำนวน 4 contigs และคอนทิกของคลอโรพลาสต์จำนวน 2 contigs โดยคอนทิกของไมโทคอนเดรียทั้ง 4 contigs สามารถรวมตัวเป็น circular chromosome ที่แตกต่างกัน 2 แบบ คือโครโมโซมวงใหญ่ มีขนาด 3,000,778 bp และโครโมโซม วงเล็กมีขนาด 144,698 bp ทางผู้วิจัยพบว่าโครโมโซมทั้ง 2 วงที่มีขนาดต่างกันนี้ไม่มีลำดับเบสซ้ำร่วมกันเลย ซึ่งลำดับเบสซ้ำนี้คือส่วนที่จะทำให้เกิดรีคอมบิเนชัน (recombination) จึงเป็นหลักฐานว่าจีโนมไมโทคอนเดรียในอ้อยนั้นไม่ได้เป็น master circle และ minicircles เหมือนกับจีโนมของไมโทคอนเดรียพืชทั่วไป แต่เป็น DNA circle 2 วง ที่แยกออกจากกันอย่างสมบูรณ์

2.2 โปรตีน Plant U-box (PUB)

โปรตีน PUB จัดอยู่ในกลุ่ม U-box protein family ประกอบด้วยกลุ่มของ E3 ubiquitin ligase ซึ่งเป็นหนึ่งในเอนไซม์ที่สำคัญในระบบยูบิควิติน-โปรตีโอโซม (ubiquitin/26S proteasome system; UPS) ในปัจจุบันมีการศึกษาและทำนาย plant U-box (PUB) proteins ในพืชหลายชนิด ทั้งในอะราบิโดพซิส (Bae & Kim, 2014), ข้าว , ถั่วเหลือง (Trujillo, 2018), ข้าวบาร์เลย์ (Ryu et al., 2019), ข้าวสาลี (Zhang et al., 2020) และฝ้าย (Lu et al., 2020) และจากการศึกษาของ Khanbo (2020) ในการพัฒนาเครื่องหมาย intron length polymorphism (ILP) จากยีนในกลุ่ม *Plant U-box (PUB)* พบว่าเครื่องหมาย ILP ที่พัฒนามาจากยีนนี้มีความสัมพันธ์กับค่าความเข้มข้นของของแข็งที่ละลายอยู่ในน้ำอ้อย (Brix) ซึ่งเป็นลักษณะที่เกี่ยวข้องกับลักษณะน้ำตาลในอ้อย (กำลังเตรียมต้นฉบับตีพิมพ์) จากการศึกษาเชิงเปรียบเทียบยีนนี้ในจีโนมของข้าวฟ่าง ซึ่งเป็นพืชที่มีบรรพบุรุษร่วมกับอ้อย (Jannoo et al., 2007) พบว่ามีความคล้ายคลึงกับยีน *U-box domain-containing protein 35* ในข้าวฟ่าง โดยโปรตีน U-box domain-containing protein 35 ประกอบด้วยโดเมน U-box ซึ่งเป็นลำดับกรดอะมิโนที่มีความอนุรักษ์สูงความยาวประมาณ 70 กรดอะมิโนที่ด้านปลายคาร์บอกซิลและโดเมน serine/threonine kinase ที่ด้านปลายอะมิโน จึงจัดอยู่ในคลาส IV ของโปรตีน PUB (Ryu et al., 2019) โดยโดเมน U-box มีหน้าที่ในการย้ายยูบิควิตินให้กับโปรตีนเป้าหมายในกระบวนการเติมยูบิควิตินให้แก่โปรตีน (protein ubiquitination) เพื่อให้โปรตีนนั้นถูกกำจัดออกไปซึ่งเป็นหนึ่งในกระบวนการที่สำคัญในระบบ UPS และโดเมน serine/threonine kinase ทำหน้าที่ในการเร่งปฏิกิริยาการเติมหมู่ฟอสเฟตให้กับกรดอะมิโนกลุ่มเซอรีน/ทรีโอนีน (serine/threonine)

จากการศึกษาหน้าที่ของโปรตีน PUB ในพืชชั้นสูง พบว่ามีความเกี่ยวข้องกับหลายกระบวนการ เช่นการตอบสนองต่อความเครียดทั้งแบบ abiotic (ความแล้ง อุณหภูมิ และความเค็ม) (Byun et al., 2017; Ryu et al., 2019; Wang et al., 2020) และ biotic (เชื้อก่อโรค และเชื้อรา) (Wang et al., 2015) การแบ่งและขยายขนาดของเซลล์ การผสมตัวเองไม่ติด การพัฒนาละอองเรณู (Chen et al., 2019; Zhang et al., 2020) จากการศึกษาอื่นที่ตอบสนองต่อความแล้งในข้าวของ Qin et al. (2020) โดยการดัดแปลงพันธุกรรมในข้าวโดยทำให้ยีน *OsPUB67* ไม่มีการแสดงออก (knock out) ด้วยระบบ CRISPR/Cas9 และยังคงดัดแปลงพันธุกรรมด้วยการทำให้ยีนนี้มีการแสดงออกที่มากขึ้น (overexpression) ด้วย พบว่ายีน *OsPUB67* ถูกชักนำให้มีการแสดงออกโดยความเครียดแบบ abiotic ทั้งความแล้ง อุณหภูมิต่ำ ความเค็ม รวมทั้ง abscisic acid และยังพบว่าภายใต้สภาวะแล้งนั้นข้าวที่ได้รับการดัดแปลงให้ยีน *OsPUB67* มีการแสดงออกมากขึ้นมีความทนแล้งมากกว่าข้าวที่ยีน *OsPUB67* ไม่มีการแสดงออก ผลจากการศึกษานี้จึงคาดว่ายีน *OsPUB67* อาจจะมีเกี่ยวข้องกับทนต่อความแล้งได้มากขึ้น การศึกษาของ Zhang et al. (2020) พบว่ายีนที่ควบคุมกระบวนการเมแทบอลิซึมของน้ำตาลและแป้งมีระดับการถอดรหัสลดลงอย่างมีนัยสำคัญในอะราบิดอพซิสที่มีการกลายของยีน *atpub4* แต่ในอะราบิดอพซิสที่มีการกลายของยีน *atpub4* ที่ได้รับการโคลนยีน *TaPUB* ที่ได้จาก cDNA ของอับละอองเรณูที่กำลังพัฒนาในข้าวสาลีพบว่าอะราบิดอพซิสสามารถกลับมามีระดับการถอดรหัสของยีนที่ควบคุมกระบวนการเมแทบอลิซึมของน้ำตาลและแป้งที่ใกล้เคียงกับอะราบิดอพซิสที่มีแอลลีลปกติ (wild type) บ่งชี้ได้ว่า *TaPUB4* อาจจะช่วยการพัฒนาของละอองเรณูโดยควบคุมการสะสมแป้งในละอองเรณู และจากการศึกษาอื่น *OsPUB73* ในข้าวของ Chen et al. (2019) โดยการทำให้ยีน *OsPUB73* ไม่มีการแสดงออก (knock out) ด้วยระบบ CRISPR/Cas9 แล้วศึกษาเซลล์วิทยาและทรานสคริปโตมของยีนที่เกี่ยวข้องกับการพัฒนาของละอองเรณูและการเป็นหมันในข้าวเพื่อระบุยีนที่แสดงออกแตกต่างกันระหว่างตัวอย่างที่ไม่มีการแสดงออกของยีน *OsPUB73* และตัวอย่างที่มีแอลลีลปกติในระหว่างกระบวนการไมโอซิส แล้ววิเคราะห์ KEGG (Kyoto Encyclopedia of Genes and Genomes) พบว่ายีนที่เกี่ยวข้องกับหลายกระบวนการมีการแสดงออกน้อยลง เช่นกระบวนการเมแทบอลิซึมของกลีเซอโรลิวัด กระบวนการสังเคราะห์กรดไขมันไม่อิ่มตัวและพลาโนนอยด์ กระบวนการย่อยสลายกรดไขมัน รวมถึงกระบวนการเมแทบอลิซึมของน้ำตาลและแป้ง จากการศึกษาเหล่านี้แสดงให้เห็นว่าโปรตีน PUB นั้นอาจจะมีเกี่ยวข้องกับกระบวนการเมแทบอลิซึมของน้ำตาลและแป้งในพืช

ตารางที่ 2.1 ตัวอย่างและหน้าที่ของโปรตีน PUB ที่พบในพืช

Species	PUB name	Function in plants	Reference
<i>Arabidopsis thaliana</i>	<i>PUB22 and PUB23</i>	Negatively Regulate Drought Tolerance	Zhao et al. (2017)
<i>Arabidopsis thaliana</i>	<i>PUB13</i>	Regulates Cell Death, Defense, and Flowering Time	Li et al. (2012)
<i>Brassicaceae oleracea</i>	<i>BoPUB7</i>	negatively regulates in pollen germination	Lian et al. (2019)
<i>Oryza sativa</i> L.	<i>OsPUB73</i>	male reproductive development	Chen et al. (2019)
<i>Oryza sativa</i> L.	<i>OsPUB67</i>	positively involved in drought tolerance	Qin et al. (2020)
<i>Oryza sativa</i> L.	<i>OsPUB2</i> and <i>OsPUB3</i>	positive regulation of low temperature stress response	Byun et al. (2017)
<i>Oryza sativa</i> L.	<i>OsPUB15</i>	plant cell death and innate immunity	Wang et al. (2015)
<i>Triticum aestivum</i> L.	<i>TaPUB4</i>	sucrose-starch metabolism in anthers	Zhang et al. (2020)
<i>Triticum aestivum</i> L.	<i>TaPUB1</i>	salt stress tolerance	Wang et al. (2020)

2.3 เครื่องหมายสนิป (single nucleotide polymorphism; SNP)

การตรวจสอบและการวิเคราะห์ความแปรปรวนทางพันธุกรรมได้นั้นสามารถช่วยให้เข้าใจพื้นฐานทางด้านโมเลกุลของพืชได้มากขึ้น ปัจจุบันมีการพัฒนาเครื่องหมายที่ใช้ดีเอ็นเอเป็นตัวบ่งชี้ทางพันธุกรรมโดยตรง หรือที่รู้จักกันในฐานะเครื่องหมายดีเอ็นเอ หรือเครื่องหมายโมเลกุล (molecular marker) ซึ่งเป็นส่วนของดีเอ็นเอที่บ่งชี้ความแตกต่างทางพันธุกรรมในระดับจีโนม เครื่องหมายดีเอ็นเอมีข้อได้เปรียบหลายประการคือมีความเสถียรภาพกว่า สามารถใช้ตรวจสอบ

เนื้อเยื่อทั้งหมดได้โดยไม่ต้องคำนึงถึงระยะการเจริญเติบโต การเปลี่ยนแปลงและการพัฒนาของพืช อีกทั้งยังไม่มีผลกระทบจากสิ่งแวดล้อมอีกด้วย และหนึ่งในเครื่องหมายดีเอ็นเอที่กำลังเป็นที่สนใจและได้รับความนิยมคือเครื่องหมายสเนป (single nucleotide polymorphism; SNP) ซึ่งเป็นความแปรปรวนหรือการเปลี่ยนแปลงของเบสเพียงหนึ่งตำแหน่งที่เกิดจากการกลายและรีคอมบิเนชัน จึงมีความหนาแน่นและกระจายตัวทั่วจีโนม อีกทั้งยังเป็นเครื่องหมายดีเอ็นเอที่มีความแปรปรวนมากที่สุด (Agarwal et al., 2008) สามารถแบ่งได้เป็น 2 กลุ่ม ตามการแทนที่ของชนิดเบสของนิวคลีโอไทด์คือการแทนที่แบบทรานสิชัน (transition) เป็นการแทนที่ระหว่างเบสพิวรีน และการแทนที่ระหว่างเบสไพริมิดีน (C/T และ A/G) และการแทนที่แบบทรานส์เวอร์ชัน (transversion) เป็นการแทนที่ระหว่างเบสพิวรีนกับเบสไพริมิดีน (C/G, A/T, C/A, และ T/G) โดยทั่วไปจะพบ SNP จำนวนมากในบริเวณที่ไม่มีการถอดรหัส (non-coding regions) ส่วน SNP ที่พบภายในบริเวณที่มีการถอดรหัส (coding region) อาจจะมีผลหรือไม่มีผลต่อการเปลี่ยนแปลงของลำดับกรดอะมิโน (non-synonymous or synonymous) ก็ได้ (Sunyaev et al., 1999) SNP ที่พบบริเวณยีนที่มีการถอดรหัสไปเป็นกรดอะมิโน (coding sequence) สามารถทำให้เกิดการเปลี่ยนแปลงของกรดอะมิโนซึ่งมีผลต่อความแปรปรวนของลักษณะฟีโนไทป์ (Jehan & Lakhanpaul, 2006) จากคุณสมบัติการกระจายตัวทั่วจีโนมสูงนี้ ทำให้มีประสิทธิภาพในการสร้างแผนที่พันธุกรรมที่มีความหนาแน่นสูงและมีประสิทธิภาพในการตรวจสอบความสัมพันธ์ระหว่างแอลลีลของยีนกับฟีโนไทป์ อีกทั้งยังสามารถทำซ้ำแล้วให้ผลเหมือนเดิมสูง (reproducibility) มีตำแหน่งที่จำเพาะบนโครโมโซม และยังสามารถแสดงออกเป็นแบบข่มร่วมอีกด้วย (codominant marker) (Bhatramakki et al., 2002; Rafalski, 2002)

ในสปีชีส์ที่เป็นดิพลอยด์ (diploid) ตัวอย่างจะมียีนอยู่ในสภาพเฮเทอโรไซกัส (heterozygous) ทำให้ความแตกต่างระหว่างนิวคลีโอไทด์ของยีนที่ควบคุมลักษณะเดียวกันซึ่งมีตำแหน่งเดียวกันบนโครโมโซมคู่เหมือน อาจเกิดจาก SNP และวัดการถ่ายทอดความแปรปรวนนี้ได้โดยตรงในรุ่นลูกหลาน หลัก ๆ แล้ว DNA ของตัวอย่างในประชากรจะมีความเหมือนกัน หากพันธุกรรมของตัวอย่างในประชากรมีความแตกต่างกันสามารถนิยามได้ว่าเป็น SNP และสามารถประเมินความหลากหลายทางพันธุกรรม (genetic diversity) จากความถี่ของ SNP ได้โดยตรง (Oraguzie, Gardiner, Rikkerink, & Silva 2007) การศึกษาความถี่ของ SNP เช่น ในมนุษย์พบ SNP จำนวน 3.1 ล้านตำแหน่งหรือพบ 1 ตำแหน่งในทุกๆ 1000 เบส (Sachidanandam R et al., 2001) ในพืชพบ SNP กระจายทั่วจีโนมที่มีความหนาแน่นสูง ตัวอย่างเช่นในจีโนมข้าวพบ SNP 1 ตำแหน่งทุกๆ 107 เบส (Huang et al., 2009) ข้าวโพดพบ SNP ในทุกๆ 200 เบส (Gore et al., 2009) ในถั่วเหลืองพบ SNP ทุกๆ 370 เบส (Choi et al., 2007) และ *Arabidopsis* พบ SNP ในทุกๆ 500 เบส (Atwell et al., 2010) แต่ภายใต้บางเงื่อนไขส่งผลต่อความถี่ของ SNP อย่างการ

บังคับการผสมพันธุ์ อย่างเช่นการผสมกลับไปยังพันธุ์พ่อแม่ การผสมระหว่างรุ่นลูกที่มีพ่อแม่เดียวกัน หรือการผสมพันธุ์กันระหว่างตัวอย่างที่มีความสัมพันธ์ทางเครือญาติต่ำ ส่งผลให้ความหลากหลายทางพันธุกรรมและความถี่ของ SNP ลดลง โดยเงื่อนไขดังกล่าวอาจเกิดขึ้นเนื่องจากการลดลงของประชากร หรือการแยกตัวของประชากรโดยธรรมชาติ (เรียกว่า founder effect) และสำหรับพืชที่ใช้เพาะปลูกซึ่งมีพันธุกรรมแคบมีผลต่อการลดลงของความหลากหลายทางพันธุกรรมในระดับนิวคลีโอไทด์

2.3.1 เครื่องหมายสปีในพืชพอลิพลอยด์

เนื่องจาก SNP เป็นเครื่องหมายที่มีบทบาทสำคัญในการเกิดวิวัฒนาการของความหลากหลายทางพันธุกรรม จึงถูกนำมาใช้ในการสร้างแผนที่พันธุกรรม (genetic map) ซึ่งมีความหนาแน่นสูง ระบุตำแหน่งของยีน (gene localization) การวิเคราะห์หน้าที่จีโนม (functional genomics) วิเคราะห์ความสัมพันธ์กับฟีโนไทป์ ด้วยวิธี linkage mapping และ linkage disequilibrium (LD) mapping ใช้เป็นเครื่องหมายในการคัดเลือก (molecular marker-assisted selection) ประเมินความสัมพันธ์ทางพันธุกรรมระหว่างตัวอย่าง และใช้ในการระบุสปีชีส์ (Oraguzie et al., 2007; Wu, Zhou, Huang, Wang, & Xu, 2019) SNP จึงกลายเป็นเครื่องหมายที่ได้รับความนิยมอย่างมากโดยเฉพาะในพืชดิพลอยด์ (diploid) เพราะสามารถระบุ SNP ได้ง่ายเนื่องจากรูปแบบของจีโนไทป์ของ SNP แต่ละตำแหน่งเป็นไปได้มากที่สุดแค่ 3 แบบ คือ homozygote ของ reference allele, heterozygote, และ homozygote ของ alternative allele แต่การระบุ SNP ในพืชพอลิพลอยด์ (polyploid) ยังคงมีความท้าทาย แม้ว่าในปัจจุบันจะมีการพัฒนาเทคโนโลยีการวิเคราะห์ลำดับเบสยุคใหม่ขึ้นมาเพื่อการวิเคราะห์และสร้างข้อมูลลำดับเบส ซึ่ง SNP ในแต่ละตำแหน่งของพืช polyploid นั้น สามารถมีรูปแบบของจีโนไทป์ได้มากกว่า 3 แบบ โดยในพืชที่มีจีโนมเป็นแบบ tetraploid อย่างถั่วลิสง และมันฝรั่ง สามารถมีจีโนไทป์ได้ถึง 5 แบบ ในข้าวโอ๊ตและข้าวสาลีซึ่งเป็นพืช hexaploid มี จีโนไทป์ได้ถึง 7 แบบ ในพืชที่มีจีโนมเป็นแบบ octaploid อย่างอ้อย และสตรอว์เบอร์รี่นั้น สามารถมีจีโนไทป์ได้มากถึง 9 แบบ และรูปแบบของจีโนไทป์สามารถมีได้มากถึง 13 แบบในพืชที่มีจีโนมเป็น dodecaploid และในการระบุ SNP ของพืช polyploid ทั้งแบบ allopolyploid และ autopolyploid เองก็ยังคงมีความแตกต่างกัน โดยในพืชที่เป็น allopolyploid หลาย ๆ ชนิด พบว่าพันธุกรรมมีความเหมือนพืช diploid ทำให้สามารถระบุ SNP ได้ง่ายกว่า โดยอาศัยการวิเคราะห์หา SNP จาก homeologous SNP ซึ่งตรงกันข้ามกับการตรวจสอบ SNP ในพืช autopolyploid ซึ่งจะมีความซับซ้อนกว่า เนื่องจากการวิเคราะห์ genotype ต้องประเมินจำนวนซ้ำของแอลลีล (allele copy number) เรียกว่า allele dosage หรือ dosage ด้วย ตัวอย่างเช่น auto-

hexaploid ที่มีแอลลีล a เป็น reference allele และแอลลีล A เป็น alternative allele มีรูปแบบของ allelic combinations ที่เป็นไปได้ 7 แบบคือ nulliplex (aaaaaa), simplex (aaaaaA), duplex (aaaaAA), triplex (aaaAAA), quadruplex (aaAAAA), pentaplex (aAAAAA), และ hexaplex (AAAAAA) (Clevenger, Chavarro, Pearl, Ozias-Akins, & Jackson, 2015; Oraguzie et al., 2007)

มีการศึกษา SNP ในพืช polyploid ทั้ง allopolyploid และ autopolyploid ด้วยเทคโนโลยีการวิเคราะห์ลำดับเบสที่แตกต่างกัน โดยในปี 2019 Kim et al. (2019) ระบุ SNP ด้วยเทคนิค genotyping-by-sequencing (GBS) ในสตรอว์เบอร์รี่พันธุ์แท้จำนวน 89 ตัวอย่าง จากการวิเคราะห์ลำดับเบสเพื่อค้นหา SNP สามารถระบุ SNP ได้ 20,923 SNP สุดท้ายแล้วคัดเลือก SNP จำนวน 2863 SNP ไปวิเคราะห์ความหลากหลายทางพันธุกรรมและโครงสร้างประชากรต่อ และในปีเดียวกันนั้น Anagbogu et al. (2019) วิเคราะห์ SNP ในกาแฟของไนจีเรียด้วยเทคนิค GBS สามารถระบุ SNP ได้ 433,048 ตำแหน่ง แล้วนำข้อมูลที่ได้อันวิเคราะห์ความหลากหลายทางพันธุกรรมและทำการจัดกลุ่มกาแฟของไนจีเรียใหม่

2.3.2 การพัฒนาเครื่องหมายสลับในอ้อย

Cordeiro, Elliott, McIntyre, Casu, and Henry (2006) ศึกษาและระบุ SNP ของอ้อยภายใน ESTs (expressed sequence tags) จำนวน 69 ESTs โดยใช้เทคนิคการวิเคราะห์ลำดับเบสไพโรฟอสเฟต (pyrophosphate sequencing) พบว่าแต่ละลำดับเบส EST มี SNP จำนวน 9 SNP หรือโดยเฉลี่ยแล้วพบ SNP 1 ตำแหน่งในทุก ๆ 50 เบสของลำดับเบสที่มีการถอดรหัส และจากตัวอย่างอ้อยจำนวน 9 ตัวอย่าง ลำดับเบสของ contig จำนวน 19 contigs พบ SNP จำนวน 58 SNPs Parida et al. (2016) ศึกษาและวิเคราะห์จีโนมโทปของ SNP ในวิถีน้ำตาลและยีนต้านทานโรคของอ้อย โดยใช้วิธีการหาลำดับเบสจากชิ้นดีเอ็นเอที่ได้จากการเพิ่มปริมาณด้วยเทคนิค PCR (cloned amplicons) และ CAPs (cleaved amplified polymorphic sequences) assays พบว่า SNP มีการกระจายตัวในบริเวณที่ไม่มีการแปลรหัสไปเป็นกรดอะมิโน (non-coding untranslated regions; UTRs) ด้านปลาย 3' (3'UTRs) มากกว่าด้านปลาย 5' (5' UTRs) และบริเวณของยีนที่มีการแปลรหัสไปเป็นกรดอะมิโน (coding sequences) และเกิดการแทนที่เบสแบบทรานสลิชัน (transition substitutions) ระหว่าง C กับ T มากกว่าการเกิดแบบทรานสเวอร์ชัน (transversions) และจากการตรวจสอบ SNP ด้วย CAPs assays พบ SNP จำนวน 16 SNPs ในวิถีน้ำตาล 4 วิถี และ 10 SNPs ในยีนต้านทานโรค 9 ยีน McIntyre et al. (2006) ระบุแอลลีลใน sucrose phosphate synthase family gene III ของอ้อย จากการวิเคราะห์หาลำดับเบสขนาด 400 นิวคลีโอไทด์จากยีนนี้ โดยใช้

อ้อยพันธุ์ Q165^A และอ้อยพันธุ์ U76-514 สามารถระบุ SNP ได้จำนวน 10 SNPs ซึ่งในจำนวนนี้มี 7 SNPs ที่พบในอ้อยทั้ง 2 พันธุ์ ต่อมา McIntyre et al. (2015) ศึกษาต่อโดยระบุแอลลีลในยีนตระกูลซูโครสฟอสเฟตซินเทสทั้ง 5 ตระกูลของอ้อยพันธุ์ Q165^A และอ้อยพันธุ์ U76-514 พบว่าสามารถระบุ SNP และ/หรือ Insertion/Deletion (InDel) ได้จำนวน 41, 10, 10, 11, และ 8 แอลลีล จาก sucrose phosphate synthase family gene I, II, III, IV, และ V ตามลำดับ Yang et al. (2017) ทำการระบุ SNP ด้วยเทคนิค GBS แล้วทำการเปรียบเทียบโปรแกรมที่ใช้ในการค้นหา SNP พบว่าสามารถระบุ SNP ได้ตั้งแต่ 16,652 ถึง 57,952 SNPs ขึ้นอยู่กับโปรแกรมที่ใช้ค้นหาและพบว่าโปรแกรม Tassel และ Genome Analysis Tool Kit (GATK) ให้ผลในการค้นหา SNP ได้ดีกว่าโปรแกรมอื่น Balsalobre et al. (2017) ระบุ SNP และ InDel ด้วยเทคนิค GBS ได้จำนวน 7,049 เครื่องหมาย เพื่อนำไปสร้าง integrated genetic linkage map ร่วมกับเครื่องหมาย TRAP และ microsatellite

2.4 การวิเคราะห์ลำดับเบสด้วยเทคโนโลยี PacBio SMRT Sequencing

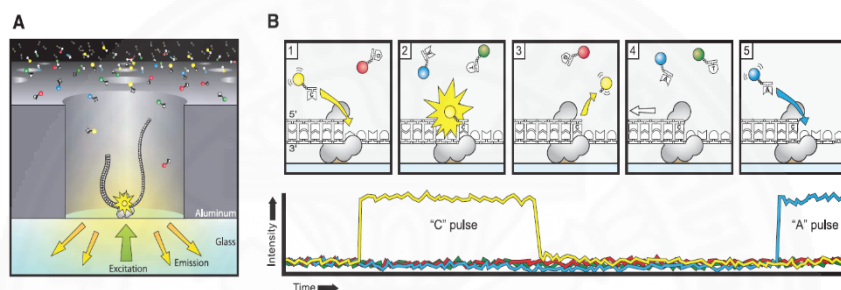
จากข้อจำกัดของขนาดลำดับเบสที่ได้จากการวิเคราะห์ลำดับเบสด้วยวิธีของแซงเกอร์ Sanger, Nicklen, and Coulson (1977) ซึ่งจัดเป็นเทคโนโลยีการวิเคราะห์ลำดับเบสยุคที่หนึ่ง ที่สามารถวิเคราะห์ลำดับเบสได้ขนาดเพียง 500 ถึง 800 คู่เบส (basepair; bp) จึงทำให้เกิดการพัฒนาเทคโนโลยีการวิเคราะห์ลำดับเบสยุคใหม่ขึ้นมา เทคโนโลยีการวิเคราะห์ลำดับเบสยุคใหม่เรียกได้ว่าเป็นเทคโนโลยีที่สามารถวิเคราะห์ลำดับเบสของดีเอ็นเอได้พร้อม ๆ กันจำนวนมาก มีความรวดเร็วขึ้น ต้นทุนต่ำลง และขั้นตอนการเตรียมดีเอ็นเอสายต้นแบบน้อยกว่าวิธีของแซงเกอร์ ดีเอ็นเอสายต้นแบบที่ใช้ในเทคโนโลยีการวิเคราะห์ลำดับเบสยุคใหม่ ได้จากการเพิ่มปริมาณดีเอ็นเอด้วยปฏิกิริยาลูกโซ่พอลิเมอเรส (*in vitro* library preparation) ทำให้ได้ข้อมูลลำดับเบสจำนวนมาก นอกจากนี้ยังสามารถหาลำดับเบสพร้อม ๆ กันจากชิ้นดีเอ็นเอ (DNA fragment) จำนวนมหาศาลได้อีกด้วย (Shendure & Ji, 2008) ปัจจุบันเทคโนโลยีการวิเคราะห์ลำดับเบสยุคใหม่อยู่ในยุคที่สองและสาม โดยเทคโนโลยีการวิเคราะห์ลำดับเบสในยุคที่สอง มีรูปแบบที่นิยมใช้กันอยู่ 4 รูปแบบ คือ 1) 454/Roche genome sequencer 2) Illumina Genome Analyzer 3) Applied Biosystems SOLiDTM Sequencer 4) Ion Torrent/ Life technology (กิตติพัฒน์ อุโฆษกิจ, มาริสา รักสุขสมบัติ, และสุภาภรณ์ ขานโบ, 2558) โดยเทคนิค 454/Roche genome sequencer เทคนิค Illumina Genome Analyzer และเทคนิค Ion Torrent/ Life technology ที่ใช้หลักการสังเคราะห์สายดีเอ็นเอเพื่อหาลำดับเบส (sequencing by synthesis; SBS) ขณะที่เทคนิค Applied

Biosystems SOLiDTM Sequencer ใช้หลักการหาลำดับเบสโดยการเชื่อมสายดีเอ็นเอด้วยปฏิกิริยาไลเกชัน (Singh & Singh, 2015)

งานวิจัยก่อนหน้าศึกษาทรานสคริปโตมในอ้อยโดยใช้เทคโนโลยีการวิเคราะห์ลำดับเบสในยุคที่สอง โดย (Huang et al., 2016) ใช้เทคนิค Solexa sequencing (Illumina HiSeqTM 2000) ในการศึกษาทรานสคริปโตมในอ้อยพันธุ์ที่มีความหวานสูง (พันธุ์ GT35) โดยสามารถหาลำดับเบสสายสั้นได้จำนวน 34,105,138 reads ระบุยีนได้ 101,255 ยีน Kasirajan, Hoang, Furtado, Botha, and Henry (2018) ใช้เทคโนโลยี Illumina Hiseq4000 ในการศึกษาทรานสคริปโตมในอ้อยพันธุ์ที่มีเส้นใยสูงและเส้นใยต่ำ เพื่อศึกษาการแสดงออกที่แตกต่างกันของยีนระหว่างเนื้อเยื่อที่เจริญเติบโตเต็มที่และยังเจริญเติบโตไม่เต็มที่ในอ้อยพันธุ์ที่มีองค์ประกอบเส้นใยแตกต่างกัน เช่นเดียวกับ Thirugnanasambandam, Hoang, Furtado, Botha, and Henry (2017) ใช้เทคโนโลยี Illumina Hiseq4000 ในการศึกษาทรานสคริปโตมในอ้อยพันธุ์ที่มีความหวานสูงและความหวานต่ำ เพื่อระบุยีนที่มีความสัมพันธ์และแสดงออกในระหว่างกระบวนการสังเคราะห์และการสะสมน้ำตาลซูโครส ถึงแม้ว่าเทคโนโลยีการวิเคราะห์ลำดับเบสยุคที่สองจะมีความสามารถในการวิเคราะห์ลำดับเบสจำนวนหลายล้านถึงพันล้านเส้นในเวลาอันสั้น แต่ก็ยังมีความเร็วไม่มากพอ อีกทั้งยังอาจให้ข้อมูลลำดับเบสที่มีความลำเอียงเนื่องจากการเพิ่มปริมาณดีเอ็นเอต้นแบบ จึงมีการพัฒนาเทคโนโลยีการวิเคราะห์ลำดับเบสยุคที่สาม (third generation sequencing technology ; TGSTs) ขึ้นมา ซึ่งเป็นเทคโนโลยีที่วิเคราะห์โมเลกุลเดี่ยวของดีเอ็นเอหรืออาร์เอ็นเอซึ่งเป็นโมเลกุลธรรมชาติ ทำให้การวิเคราะห์ลำดับเบสมีศักยภาพมากขึ้น ได้ลำดับเบสที่มีสายยาวมากขึ้น ค่าใช้จ่ายลดลง อีกทั้งลดระยะเวลาในการวิเคราะห์อีกด้วย (กิตติพัฒน์ อุโฆษกิจ และคณะ, 2558)

หนึ่งในเทคโนโลยีการวิเคราะห์ลำดับเบสในยุคที่สามคือ Single-molecule, real-time (SMRT) sequencing เป็นเทคนิคที่ใช้การวิเคราะห์ลำดับเบสดีเอ็นเอธรรมชาติโมเลกุลเดี่ยวแบบตามเวลาจริง สามารถวิเคราะห์ลำดับเบสสายยาว (single molecule long-reads) เฉลี่ยได้มากถึง 10-18 Kb เนื่องจากถูกพัฒนาโดย Pacific Biosciences จึงถูกเรียกว่า PacBio® RS (PacBio® RS High Resolution Genetic Analyzer) SMRT sequencing ซึ่งเป็นการวิเคราะห์ลำดับเบสโดยการตรวจจับกิจกรรมของเอนไซม์ดีเอ็นเอพอลิเมอเรสขณะต่อสายของนิวคลีโอไทด์ กระบวนการทำงานของเทคโนโลยีนี้เกิดบนแผ่นเซลล์ SMRT ในหลุม zero mode waveguides (ZMW) ซึ่งมีเส้นผ่านศูนย์กลางประมาณ 10 นาโนเมตร เอนไซม์ดีเอ็นเอพอลิเมอเรส 1 โมเลกุลถูกตรึงอยู่ที่ก้นหลุมนี้ หน่วยนิวคลีโอไทด์ที่ติดฉลากฟลูออเรสเซนต์เพียงหนึ่งหน่วยเท่านั้นที่สามารถแพร่ผ่านหลุม และถูกนำไปต่อสายดีเอ็นเอในหนึ่งหลุมของ ZMW เมื่อนิวคลีโอไทด์ต่อสายดีเอ็นเอแล้วจะมีการปล่อยแสงฟลูออเรสเซนต์ออกมา ระบบทำการตรวจจับภาพตามเวลาจริง และส่วนของฟลูออเรสเซนต์จะถูกตัด

ออกจากหมู่ฟอสเฟตตามธรรมชาติของการสังเคราะห์ดีเอ็นเอโดยเอนไซม์ดีเอ็นเอพอลิเมอเรส ดังแสดงในภาพที่ 2.2 เนื่องจากเป็นเทคโนโลยีที่วิเคราะห์ลำดับเบสโดยการตรวจจับกิจกรรมของเอนไซม์ดีเอ็นเอพอลิเมอเรสขณะต่อสายของนิวคลีโอไทด์ตามเวลาจริงซึ่งเกิดขึ้นอย่างรวดเร็วมาก อีกทั้งประสิทธิภาพของการตรวจจับภาพยังไม่ดีนัก เพื่อเป็นการเพิ่มความถูกต้องในการระบุชนิดนิวคลีโอไทด์และช่วยลดอัตราการคลาดเคลื่อน จึงสร้างคลัง SMRTbell เพื่อให้ได้สายดีเอ็นเอต้นแบบที่เป็นวงแหวนในการสังเคราะห์ดีเอ็นเอสายใหม่จากการทำงานของ Taq polymerase โดยเอนไซม์จะแยกสาย DNA ต้นแบบออกจากกันแล้วสังเคราะห์ DNA สายใหม่เป็นวงกลมซึ่งสามารถวิเคราะห์ลำดับเบสซ้ำได้ถึง 10–15 รอบของโมเลกุลเดิม (Eid et al., 2009)



ภาพที่ 2.2 เทคโนโลยี PacBio SMRT sequencing (A) โมเลกุลดีเอ็นเอต้นแบบ 1 โมเลกุลเกาะอยู่กับเอนไซม์ดีเอ็นเอพอลิเมอเรสที่ถูกตรึงอยู่ที่ก้นหลุมของ zero mode waveguides (ZMW) นิวคลีโอไทด์ที่ติดฉลากฟลูออเรสเซนต์ถูกนำไปต่อสายดีเอ็นเอจากการทำงานของเอนไซม์ดีเอ็นเอพอลิเมอเรสปล่อยแสงฟลูออเรสเซนต์ออกมา (B) ลำดับเหตุการณ์การตรวจจับแสงฟลูออเรสเซนต์จากก้นหลุม ZMW (1) นิวคลีโอไทด์ที่ติดฉลากฟลูออเรสเซนต์มีปฏิสัมพันธ์กับบริเวณ active site ของเอนไซม์ดีเอ็นเอพอลิเมอเรส (2) นิวคลีโอไทด์ที่ติดฉลากฟลูออเรสเซนต์ต่อกับสายดีเอ็นเอแล้วแปลงแสงฟลูออเรสเซนต์ออกมา ระบบทำการตรวจจับแสงฟลูออเรสเซนต์ (3) ส่วนของฟลูออเรสเซนต์จะถูกตัดออกจากหมู่ฟอสเฟตจากการทำงานของเอนไซม์ดีเอ็นเอพอลิเมอเรส (4) เอนไซม์ดีเอ็นเอพอลิเมอเรสขยับเปลี่ยนตำแหน่ง (5) นิวคลีโอไทด์ที่ติดฉลากฟลูออเรสเซนต์ตัวใหม่เข้ามามีปฏิสัมพันธ์กับบริเวณ active site ของเอนไซม์ดีเอ็นเอพอลิเมอเรสเพื่อเข้าสู่การสังเคราะห์ดีเอ็นเอในรอบใหม่ จาก ” Real-time DNA sequencing from single polymerase molecules,” โดย J. Eid, A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, P. Peluso, D. Rank, P. Baybayan, and B. Bettman, 2009, Science, 323(5910), p. 134.

2.4.1 การวิเคราะห์ลำดับเบสในอ้อยด้วยเทคโนโลยี PacBio SMRT Sequencing

มีการนำเทคโนโลยี PacBio SMRT Sequencing มาใช้ในการวิเคราะห์ลำดับเบสในอ้อย โดย Shearman et al. (2016) ใช้เทคโนโลยี PacBio RSII ในการหาลำดับเบสและประกอบลำดับเบสจีโนมของไมโตคอนเดรียของอ้อยพันธุ์ KK3 พบว่าสามารถวิเคราะห์ลำดับเบสสายยาว (corrected reads) ซึ่งมีขนาดมากกว่า 30 Kb มีค่า read depth เฉลี่ยเท่ากับ 14 โดยประกอบด้วย 20 contigs ซึ่งเป็นคอนทิกของไมโตคอนเดรียจำนวน 4 contigs และคอนทิกของคลอโรพลาสต์จำนวน 2 contigs ซึ่งคอนทิกของไมโตคอนเดรียทั้ง 4 contigs สามารถรวมตัวเป็น circular chromosome ที่แตกต่างกัน 2 แบบ โดยโครโมโซมวงใหญ่ หรือ chromosome 1 มีขนาด 3000778 bp และโครโมโซมวงเล็กหรือ chromosome 2 มีขนาด 144698 bp จากการศึกษาในครั้งนี้ทำให้ทราบว่า mitochondrial genome ของอ้อยนั้น เป็น DNA circle 2 วง ที่แยกออกจากกันอย่างสมบูรณ์ ในปีเดียวกันนั้น Okura et al. (2016) ใช้เทคโนโลยี Illumina HiSeq2000 และ PacBio platforms ในการวิเคราะห์ลำดับเบสของโคลน bacteria artificial chromosome (BAC) ของอ้อยพันธุ์ SP80-3280 จำนวน 178 โคลน พบว่าจากการใช้ PacBio SMRT sequencing platform สามารถสร้างลำดับเบสสายยาว (reads) ได้จำนวน 101,841 reads ที่ความยาวเฉลี่ย 3,637 bp และความยาวรวมเท่ากับ 370.4 Mb Hoang et al. (2017) ศึกษาทรานสคริปโตมในเนื้อเยื่อใบ ลำต้น และรากที่ระยะการเจริญเติบโตแตกต่างกัน (เจริญเติบโตไม่เต็มที่ และเจริญเติบโตเต็มที่) จำนวน 22 พันธุ์ ด้วยเทคนิค PacBio long-read Iso-Seq พบว่าสามารถวิเคราะห์ลำดับเบสได้ทั้งหมด 290,393 reads ประกอบด้วยนิวคลีโอไทด์จำนวน 548,763,750 นิวคลีโอไทด์ จากเซลล์ SMRT จำนวน 6 เซลล์ โดยที่ความยาวของลำดับเบสมีขนาดตั้งแต่ 300 เบสถึง 53,235 เบส มีค่า N50 (ความยาวเฉลี่ยที่ 50% ของความยาวลำดับเบสรวมทั้งหมด) เท่ากับ 2,408 เบส ในส่วนของลำดับเบสที่ได้จาก Quiver polishing จำนวน 107,604 ลำดับเบส (sequences) นั้นมีลำดับเบสที่เป็นลำดับเบสที่มีคุณภาพสูง (high quality) จำนวน 65,715 sequences และลำดับเบสที่มีคุณภาพต่ำ (low quality) จำนวน 41,891 sequences มีขนาดตั้งแต่ 301 ถึง 18,548 เบส มีค่า N50 และ N75 เท่ากับ 1,994 และ 1,271 เบส ตามลำดับ และ Piriyaongsa et al. (2018) สร้างข้อมูลทรานสคริปโตมของเนื้อเยื่อใบจากอ้อยพันธุ์ KK3 โดยใช้เทคโนโลยี PacBio RS II (Iso-Seq method) และสร้างข้อมูลลำดับ RNA สายสั้น จากตัวอย่างเดียวกันโดยใช้เทคโนโลยี Ion Proton platform เพื่อลดความผิดพลาดจากการระบุเบส พบว่าจากเทคโนโลยี PacBio SMRT sequencing สามารถสร้างลำดับเบสได้จำนวน 4,159,174 reads ในจำนวนนี้เป็นลำดับเบสแบบเต็มความยาว (มีลำดับเบสจาก polyA-tail ถึง 5' end) จำนวน 232,328 reads เมื่อผ่านการวิเคราะห์เพื่อลดความผิดพลาดในการระบุเบสเหลือลำดับเบสจำนวน 119,339 reads มีความยาวตั้งแต่ 307 ถึง 26751 bp ที่ N50 มี

ความยาวเท่ากับ 3611 bp ความยาวเฉลี่ยเท่ากับ 3099 bp และ L50 (จำนวนลำดับเบสที่ 50% ของจำนวนลำดับเบสทั้งหมด) มีจำนวน 36456 sequences จากขนาดของ N50 (3611 bp) เป็นความยาวเฉลี่ยที่มากกว่าการศึกษาทรานสคริปโตมก่อนหน้าที่ผ่านมาในอ้อย และในปีเดียวกันนั้น Garsmeur et al. (2018) ใช้เทคโนโลยี PacBio RSII ในการวิเคราะห์หาลำดับเบสของโคลน bacteria artificial chromosome (BAC) ของ minimum tilling path (MTP) ในอ้อยพันธุ์ R570 จำนวน 4083 โคลน พบว่าค่า sequencing depth ต่อ BAC เท่ากับ 100x ได้คอนทิกจำนวน 5896 contigs ที่ความยาวเฉลี่ย 102,858 bp โดยคอนทิกถูกกำหนดให้กับแต่ละ BAC จาก whole-genome profiling (WGP) tag ที่ตรงกัน โดยโคลนของ BAC ส่วนใหญ่ถูกประกอบรวมเป็นลำดับเบสคอนทิกเดียว (85%) หรือไม่ก็ contig จากจำนวนโคลนของ BAC ทั้งหมด 4535 BAC sequence ซึ่งรวมโคลน BAC จำนวน 577 BAC ที่วิเคราะห์ลำดับเบสแล้ว จากจำนวนลำดับเบสของโคลน BAC ทั้งหมดสามารถประกอบเป็นลำดับเบสยาว 531 Mb

2.5 แผนที่ความสัมพันธ์ (association mapping)

ลักษณะทางการเกษตรที่สำคัญส่วนใหญ่ เช่น ปริมาณผลผลิต คุณภาพของผลผลิต และลักษณะการต้านทานโรคล้วนถูกควบคุมด้วยยีนหลายยีน หรือที่เรียกว่าลักษณะปริมาณ (quantitative traits) บริเวณยีนบนจีโนมที่มีความสัมพันธ์กับลักษณะปริมาณนี้ จึงเรียกว่า quantitative trait loci (QTL) เพื่อตรวจสอบบริเวณบนจีโนมที่มีความสัมพันธ์กับฟีโนไทป์ จึงมีการวิเคราะห์แผนที่ยีนหรือเครื่องหมายดีเอ็นเอที่ควบคุมลักษณะปริมาณเพื่อระบุเครื่องหมายดีเอ็นเอที่ลิงก์ (link) กับยีนที่ควบคุมลักษณะทางการเกษตรที่สำคัญ ซึ่งเครื่องหมายดีเอ็นเอที่ระบุได้นี้จะถูกนำมาใช้ในการช่วยคัดเลือกพืชในกระบวนการปรับปรุงพันธุ์ (marker-assisted selection; MAS) วิธีการพัฒนาเครื่องหมายดีเอ็นเอที่ช่วยในการคัดเลือกที่เป็นที่นิยมมากวิธีหนึ่งคือการวิเคราะห์แผนที่ความสัมพันธ์ (association mapping) ซึ่งเป็นการวิเคราะห์หาความสัมพันธ์ระหว่างเครื่องหมายดีเอ็นเอกับความแปรปรวนของฟีโนไทป์ โดยการตรวจสอบความไม่สมดุลลิงเกจ (linkage disequilibrium; LD) อย่างมีนัยสำคัญระหว่างเครื่องหมายดีเอ็นเอกับฟีโนไทป์ โดยใช้ประชากรทดสอบที่มีความหลากหลายทางพันธุกรรมซึ่งอาจรวบรวมหรือเลือกจากประชากรธรรมชาติเชื้อพันธุกรรม, พันธุ์ปรับปรุง, พันธุ์ปลูก หรือพันธุ์ที่เป็น variety การวิเคราะห์แผนที่ความสัมพันธ์แบ่งได้เป็น 2 วิธี วิธีแรกคือ genome wide association mapping (GWAS) เป็นการพิจารณาความแปรปรวนทางพันธุกรรมภายในจีโนมเพื่อหาความสัมพันธ์ระหว่างเครื่องหมายดีเอ็นเอกับลักษณะที่ปรากฏโดยใช้เครื่องหมายดีเอ็นเอที่กระจายทั่วทั้งจีโนม จึงมีการใช้เครื่องหมายโมเลกุลไม่ว่าจะเป็น

simple sequence repeat (SSR) และ/หรือ SNP จำนวนมากเพื่อทดสอบหาความสัมพันธ์กับลักษณะที่ซับซ้อนต่าง ๆ และอีกหนึ่งวิธีคือ candidate gene association mapping ซึ่งเป็นการตรวจสอบพอลิมอร์ฟิซึมแบบเฉพาะเจาะจงในบางบริเวณของจีโนมหรือภายในยีนที่สนใจซึ่งเป็นยีนที่มีหน้าที่อยู่ในวิถีที่เกี่ยวข้องกับการแสดงลักษณะที่สนใจ จำนวนเครื่องหมายดีเอ็นเอที่ใช้จึงมีจำนวนลดลง (Banerjee, Khan, Swapna, Singh, & Kumar, 2020; Zhu, Gore, Buckler, & Yu, 2008)

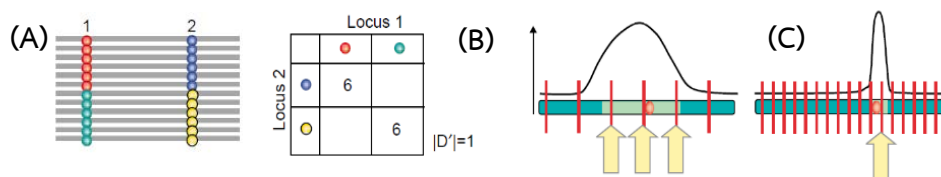
association mapping นอกจากจะสามารถระบุความแปรปรวนภายในยีนที่เป็นสาเหตุของความแตกต่างของฟีโนไทป์ ยังสามารถระบุปฏิสัมพันธ์ระหว่างแอลลีลจากยีนหลาย ๆ ตำแหน่ง และประเมินผลของยีนที่มีต่อฟีโนไทป์แม้ยีนมีผลต่อการแสดงออกของฟีโนไทป์น้อย แต่อย่างไรก็ตามในการวิเคราะห์ association mapping ยังมีข้อจำกัดในกรณีที่ฟีโนไทป์ที่สนใจมีความสัมพันธ์กับโครงสร้างประชากรอย่างมาก ส่งผลให้เกิดความคลาดเคลื่อนในการประเมินอิทธิพลของยีนต่อลักษณะที่ปรากฏและทำให้เกิดความสัมพันธ์ที่ไม่แท้จริง (spurious association) ระหว่างเครื่องหมายดีเอ็นเอกับลักษณะที่สนใจ เพื่อเป็นการลดผลทางบวกที่ไม่ถูกต้อง (false positive) นี้ จึงพิจารณาอิทธิพลทั้งจากการเกิดโครงสร้างประชากรและความสัมพันธ์ทางเครือญาติภายในประชากรที่ศึกษา ร่วมกันในการวิเคราะห์ association mapping ด้วยโมเดลทดสอบแบบผสม (mixed-model) โดยมีตัวแปรร่วมคือโครงสร้างประชากร (Q-matrix) และความเป็นเครือญาติ (kinship) (Banerjee et al., 2020) แต่ในการวิเคราะห์ association mapping เป็นการทดสอบสมมติฐานที่ใช้ข้อมูลวิเคราะห์ซ้ำหลายครั้ง (multiple tests) จึงมีโอกาสสูงที่จะเกิด false positive ดังนั้นจึงไม่สามารถใช้ระดับนัยสำคัญ (p-value) ที่ 0.05 ในการระบุความสัมพันธ์ระหว่างเครื่องหมายดีเอ็นเอกับลักษณะที่สนใจได้ วิธีการแก้ไขปัญหาดังกล่าวที่นิยมใช้มี 2 วิธี ได้แก่ วิธี false discovery rate (FDR) ของ Benjamini, Drai, Elmer, Kafkafi, and Golani (2001) และ Bonferroni correction ของ Storey (2002)

2.5.1 ความไม่สมดุลลิงเกจ (linkage disequilibrium; LD)

ความไม่สมดุลลิงเกจหรือ LD เป็นความสัมพันธ์แบบไม่สุ่มระหว่างแอลลีลที่อยู่ต่างตำแหน่งกันอาจจะอยู่บนโครโมโซมแท่งเดียวกันหรือคนละแท่ง ดังนั้นในการวิเคราะห์ความสัมพันธ์จึงเป็นการตรวจสอบว่าการมี LD กันระหว่างเครื่องหมายดีเอ็นเอกับแอลลีลภายในยีนมีผลต่อการแสดงออกของฟีโนไทป์หรือไม่ รวมถึงการศึกษาขอบเขตของความไม่สมดุลลิงเกจ (extent of LD) และการเสื่อมถอยของความไม่สมดุลลิงเกจ (LD decay) ทำให้ทราบถึงระยะห่างระหว่างเครื่องหมายดีเอ็นเอกับแอลลีลของยีนที่ยังมีความสัมพันธ์กันอยู่ซึ่งมีความสำคัญต่อประสิทธิภาพของการวิเคราะห์แผนที่ความสัมพันธ์ในการระบุความสัมพันธ์ที่แท้จริงระหว่างเครื่องหมายดีเอ็นเอและยีน

ที่ควบคุมลักษณะปริมาณ อีกทั้งยังเป็นตัวกำหนดจำนวนและความหนาแน่นของเครื่องหมายจากภาพที่ 2.3 แสดงความสัมพันธ์ระหว่างขอบเขตของ LD กับความหนาแน่นของเครื่องหมายในแผนที่ความสัมพันธ์ จะเห็นได้ว่าการมีระยะขอบเขต LD ที่กว้าง (ภาพที่ 2.3B) จำนวนเครื่องหมายที่ใช้ในการตรวจสอบและระบุเครื่องหมายที่มีความสัมพันธ์กับยีนที่ควบคุมลักษณะน้อยกว่าการมีขอบเขต LD ระยะที่แคบ (ภาพที่ 2.3C) ก็สามารถระบุเครื่องหมายที่มีความสัมพันธ์ได้ ด้วยเหตุผลนี้จึงมีความสำคัญยิ่งในการศึกษาและทำความเข้าใจเกี่ยวกับ LD (Flint-Garcia, Thornsberry, & Buckler IV, 2003) ทำให้ในปัจจุบันการวิเคราะห์ LD กลายมาเป็นจุดสนใจในการศึกษาพันธุศาสตร์ในพืช เนื่องจากการมี LD ระหว่างเครื่องหมายทำให้สามารถระบุบริเวณบนจีโนมหรือบนยีนที่มีความสัมพันธ์กับลักษณะที่สนใจโดยการศึกษา LD ของเครื่องหมายทั้งจีโนมหรือศึกษาแค่ภายในยีนที่สนใจจากความแปรปรวนของเครื่องหมายดีเอ็นเออย่างเช่นเครื่องหมาย SNP จึงมีงานวิจัยที่ศึกษาขอบเขตของ LD และ LD decay ในพืชหลายชนิด ทั้งในอะราบิดอพซิสที่พบว่าค่าขอบเขตของ LD เฉลี่ยที่ 50 kb และเป็นค่าที่เหมาะสมสำหรับการวิเคราะห์ genome-wide LD mapping ในถั่วเหลืองและผักกาดก้านขาวพบ LD decay ที่ระยะน้อยกว่า 600 kb และ 1 Mb ตามลำดับ ในข้าวและข้าวฟ่างพันธุ์พื้นเมืองมีขอบเขตของ LD อยู่ในช่วง 100 kb ในข้าวโพดซึ่งเป็นพืชผสมข้ามพบว่า LD decay ที่ระยะน้อยกว่า 100 kb เมื่อเปรียบเทียบกับค่า LD ในพืชผสมตัวเองอย่างเช่นข้าวสาลีพบว่ามีค่า LD กว้างกว่าพืชผสมข้ามอย่างข้าวโพด หรือเมื่อเปรียบเทียบกับค่า LD ในข้าวบาร์เลย์ซึ่งเป็นพืชผสมตัวเอง ระหว่างประชากรพันธุ์ป่าและพันธุ์ปลูกพบว่าประชากรที่เป็นพันธุ์ป่านั้นมี LD แคบกว่าพันธุ์ปลูก เนื่องจากรูปแบบของ LD ในประชากรที่เป็นพันธุ์ป่านั้นขึ้นอยู่กับประวัติการเกิดรีคอมบิเนชันที่ยาวนาน โครมโซ่สร้างประชากรและการคัดเลือก (Louis-Marie, Jérôme, Mike, Angélique, & Jean-Christophe, 2008; Sela et al., 2011) ในพืชชนิดอื่น ๆ ความหนาแน่นของเครื่องหมายดีเอ็นเอ ความละเอียดของแผนที่ ประสิทธิภาพของแผนที่ และสถิติที่ใช้ทดสอบล้วนถูกกำหนดจากขอบเขตของ genome-wide LD ความหลากหลายทางพันธุกรรม และความเป็นเครือญาติ (kinship) ภายในประชากร และจากการศึกษา LD ในอ้อยนั้น พบ LD ระหว่างเครื่องหมายที่ลิงก์กันในระยะใกล้เป็นจำนวนมาก และมีการลดลงของ LD เรื่อย ๆ ในระยะประมาณ 5-10 cM ดังนั้นระยะห่างระหว่างเครื่องหมายดีเอ็นเอที่ 10 cM จึงเป็นเกณฑ์ในการกำหนดหรือแยกความสัมพันธ์ของเครื่องหมายว่าเป็นแบบลิงเกจหรือเป็นความสัมพันธ์แบบอื่น ดังนั้นระดับของ LD ที่เพิ่มขึ้นหรือการมี LD block จึงช่วยลดจำนวนเครื่องหมายที่ครอบคลุมทั่วทั้งจีโนมขนาดใหญ่ของอ้อยได้ (Jannoo, Grivet, Dookun, D'Hont, & Glaszmann, 1999; Raboin, Pauquet, Butterfield, D'Hont, & Glaszmann 2008)

จากที่กล่าวมาข้างต้นพบว่ารูปแบบของ LD นั้นขึ้นอยู่กับหลายปัจจัยหนึ่งในปัจจัยที่สำคัญคือการคัดเลือกโดยธรรมชาติหรือโดยนักปรับปรุงพันธุ์ ณ ที่นี้ขอกกล่าวถึงการคัดเลือกจากกระบวนการปรับปรุงพันธุ์โดยนักปรับปรุงพันธุ์ซึ่งจะดำเนินการวางแผนการผสมพันธุ์พ่อแม่และคัดเลือกเก็บไว้เฉพาะลูกผสมที่มีคุณสมบัติตรงตามความต้องการซึ่งมีผลต่อระดับของความหลากหลายทางพันธุกรรมอย่างมีนัยสำคัญ (Lu et al., 2011) การคัดเลือกเก็บเฉพาะลูกผสมที่มีลักษณะตรงตามความต้องการเป็นการคัดเลือกบริเวณบนยีนที่ส่งผลดีต่อลักษณะที่ต้องการ จึงเป็นการตรึงการกลายที่มีผลดี (beneficial mutation) ต่อลักษณะที่ต้องการไว้ (Kim & Nielsen, 2004) เรียกการคัดเลือกแบบนี้ว่า selective sweep เป็นการคัดเลือกแบบกวาดเอาทั้งบล็อกของ beneficial mutation และแอลลีลขนานข้างที่มี LD กับ beneficial mutation (LD block) (Wang et al., 2013) ซึ่ง LD block นี้จะถูกคัดเลือกเก็บไว้ในทุก ๆ รุ่น ทำให้เกิดปรากฏการณ์ทางพันธุกรรมเป็นแบบคอขวด (genetic bottleneck) เนื่องจากการคัดเลือกนี้เกิดจากการคัดเลือกโดยนักปรับปรุงพันธุ์ทำให้แอลลีลใน LD block นี้มีความถี่เพิ่มขึ้นในระยะเวลาอันรวดเร็ว ขณะที่แอลลีลอื่น ๆ ที่ไม่ได้ถูกคัดเลือกมานั้นเกิดการหายไป (Kim & Nielsen, 2004) จึงส่งผลให้ความหลากหลายทางพันธุกรรมลดลง (Stephan, 2019) นอกจากนี้จากการศึกษาที่ผ่านมายังพบว่า selective sweeps นี้มีผลอย่างมากต่อการเปลี่ยนแปลงการกระจายตัวของตำแหน่งความแปรปรวนบนยีนรวมถึงมีผลต่อการเพิ่มระยะ LD ระหว่างตำแหน่ง beneficial mutation อีกด้วย (Kim & Nielsen, 2004; Nimmakayala et al., 2014; Qian, Qian, & Snowdon, 2014)



ภาพที่ 2.3 แผนภาพแสดงหลักการความไม่สมดุลลิงเกจ (linkage disequilibrium; LD) ระหว่างยีน 2 ตำแหน่ง (A) ซึ่งแสดงความสัมพันธ์ (LD) ระหว่างแอลลีลจากยีนทั้ง 2 ตำแหน่ง จะเห็นได้ว่าแอลลีลสีแดงจาก Locus 1 สัมพันธ์กับแอลลีลสีฟ้าจาก Locus 2 เช่นเดียวกับแอลลีลสีเขียวจาก Locus 1 สัมพันธ์กับแอลลีลสีเหลืองจาก Locus 2 และแสดงความสัมพันธ์ระหว่างขอบเขตของ LD กับความหนาแน่นของเครื่องหมาย จะเห็นได้ว่าระยะการเกิด LD ที่กว้าง (B) จำนวนของเครื่องหมาย (แทนด้วยเส้นสีแดง) ที่ใช้ในการวิเคราะห์แผนที่ความสัมพันธ์ มีจำนวนน้อยกว่าการเกิด LD ในระยะแคบ (C) ก็สามารถตรวจสอบและระบุเครื่องหมายที่มีความสัมพันธ์ (ลูกศรสีเหลือง) กับยีนที่ควบคุมลักษณะ (แทนด้วยรูปวงรีสีแดง) ได้ จาก “Applications of single nucleotide polymorphisms in crop genetics,” โดย A. Rafalski, 2002, *Current opinion in plant biology*, 5(2), p. 97.

2.5.2 การวิเคราะห์แผนที่ความสัมพันธ์ในอ้อย

มีงานวิจัยที่วิเคราะห์แผนที่ความสัมพันธ์โดยใช้เครื่องหมายดีเอ็นเอชนิดต่าง ๆ ในอ้อย อย่าง Gouy et al. (2015) วิเคราะห์ความสัมพันธ์ระหว่างเครื่องหมาย AFLP, DArT และ SSR จำนวน 3327 เครื่องหมายกับลักษณะสัณฐานวิทยาและลักษณะการต้านทานโรคจำนวน 13 ลักษณะในตัวอย่างอ้อยจำนวน 183 ตัวอย่างพบว่าสามารถระบุเครื่องหมายดีเอ็นเอจำนวน 11 เครื่องหมายที่มีความสัมพันธ์อย่างมีนัยสำคัญกับลักษณะที่ศึกษา Banerjee et al. (2015) สามารถระบุเครื่องหมาย SSR จำนวน 40 เครื่องหมายที่มีความสัมพันธ์กับลักษณะชูโครสและลักษณะที่เกี่ยวข้องกับองค์ประกอบผลผลิตในอ้อยจำนวน 108 ตัวอย่าง โดยใช้เครื่องหมาย SSR จำนวน 123 เครื่องหมายจากการวิเคราะห์ marker-trait association (MTA) Racedo et al. (2016) ศึกษา genome wide association study (GWAS) เพื่อระบุเครื่องหมายโมเลกุลที่มีความสัมพันธ์กับผลผลิตอ้อยและปริมาณน้ำตาลโดยใช้เครื่องหมาย DArT (diversity array technology) และ TRAP (target region amplified polymorphism) พบว่ามีเครื่องหมายจำนวน 43 เครื่องหมายที่มีความสัมพันธ์อย่างมีนัยสำคัญกับปริมาณผลผลิตและอีก 38 เครื่องหมายที่มีความสัมพันธ์อย่างมีนัยสำคัญกับปริมาณน้ำตาล Siraree et al. (2017) วิเคราะห์ MTA พบว่าสามารถระบุเครื่องหมาย SSR จำนวน 60 เครื่องหมายที่มีความสัมพันธ์กับลักษณะทางสัณฐานวิทยาและองค์ประกอบของผลผลิตในอ้อยจำนวน 22 ลักษณะ จากจำนวนเครื่องหมาย SSR จำนวน 174 เครื่องหมาย และ

ลักษณะทางสัณฐานวิทยาและองค์ประกอบของผลผลิตในอ้อยจำนวน 32 ลักษณะ ในตัวอย่างอ้อยจำนวน 92 ตัวอย่าง Yang et al. (2019) ศึกษา GWAS เพื่อระบุตำแหน่งยีนที่ควบคุมลักษณะองค์ประกอบของเส้นใยในอ้อยจำนวน 299 ตัวอย่าง พบว่าสามารถระบุเครื่องหมายดีเอ็นเอได้จำนวน 107 เครื่องหมายและยีนอีกจำนวน 107 ยีนที่มีความสัมพันธ์กับลักษณะองค์ประกอบของเส้นใยทั้ง 10 ลักษณะ Ukoskit et al. (2019) ตรวจสอบ QTL mapping โดยใช้ association mapping เพื่อระบุ EST-SSRs (expressed sequence tag-simple sequence repeats) ที่มีความสัมพันธ์กับลักษณะน้ำตาล โดยใช้โมเดล MLM ในอ้อยจำนวน 200 ตัวอย่าง พบว่าสามารถระบุ EST-SSR markers ที่มีความสัมพันธ์กับ uridine diphosphate (UDP) glycosyltransferase และ beta-amylase ซึ่งมีความเกี่ยวข้องกับค่าร้อยละของน้ำตาลซูโครส (POL) และผลผลิตน้ำตาล (sugar yield) Fickett et al. (2019) สามารถระบุเครื่องหมาย SNPs และ Indels ที่มีความสัมพันธ์กับลักษณะที่เกี่ยวข้องกับองค์ประกอบผลผลิตอ้อยและน้ำตาล จากการศึกษา genome wide association (GWAS) ในตัวอย่างอ้อยจำนวน 97 ตัวอย่าง เช่นเดียวกับ Yang, Luo, Todd, Sood, and Wang (2020) ที่วิเคราะห์ GWAS ซึ่งสามารถระบุเครื่องหมาย SNP จำนวน 217 เครื่องหมายและยีนจำนวน 225 ยีน ที่มีความสัมพันธ์กับลักษณะองค์ประกอบผลผลิตจำนวน 9 ลักษณะอย่างมีนัยสำคัญในตัวอย่างอ้อยจำนวน 299 ตัวอย่าง

2.5.3 การวิเคราะห์แผนที่ความสัมพันธ์ภายในยีน (candidate gene association mapping)

เป็นวิธีการที่ใช้ในการศึกษาลักษณะที่สนใจศึกษาอย่างละเอียด โดยทำการตรวจสอบพอลิมอร์ฟิซึมภายในยีนที่มีหน้าที่อยู่ในวิถีต่าง ๆ ซึ่งเกี่ยวข้องกับการแสดงลักษณะที่สนใจ และเนื่องจากเป็นการตรวจสอบพอลิมอร์ฟิซึมภายในยีนที่สนใจจึงมักเป็นการศึกษาความหลากหลายของเครื่องหมาย SNP เพราะเป็นเครื่องหมายที่มีจำนวนมาก และมีการกระจายตัวทั่วจีโนม ทำให้แผนที่ QTL มีความละเอียดสูงอีกทั้งการวิเคราะห์ LD ก็มีประสิทธิภาพ ในการระบุทั้ง SNP จึงต้องใช้ตัวอย่างที่มีความหลากหลายทางพันธุกรรม และบริเวณที่ทำการตรวจสอบ SNP ทั้งโปรโมเตอร์ อินตรอน เอกซอน บริเวณที่ไม่มีการแปลรหัส (5'/3'-untranslated region) และบริเวณที่ไม่มีการถอดรหัส (non-coding region) ซึ่งคาดว่าเป็นบริเวณที่มีความหลากหลายของนิวคลีโอไทด์สูงกว่า บริเวณที่มีการถอดรหัส (coding region) และอัตราการเสื่อมถอยของการไม่สมดุลงของลิงเกจของแต่ละยีนล้วนเป็นตัวกำหนดจำนวน SNP ต่อความยาว ซึ่งสำคัญและจำเป็นต่อการวิเคราะห์แผนที่ความสัมพันธ์ (Whitt & Buckler, 2003) ดังนั้นจุดประสงค์หลักของวิธีนี้คือเพื่อระบุ SNP ที่เป็น

สาเหตุของความแปรปรวนของฟีโนไทป์ ซึ่งอาจจะมีผลต่อการเปลี่ยนหน้าที่ของโปรตีนหรือการแสดงออกของยีนนั่นเอง (Tabor, Risch, & Myers, 2002)

มีการศึกษา candidate-gene association mapping ในพืชหลายชนิด โดย Ehrenreich et al. (2009) ศึกษา ยีน 51 ยีนที่เกี่ยวข้องกับเวลาในการออกดอกของต้นอะราบิโดพซิส จำนวน 275 ตัวอย่าง แล้วค้นหาเครื่องหมาย SNP ภายในยีนเหล่านี้ จากการวิเคราะห์ความสัมพันธ์พบว่า มี SNP ภายในยีนจำนวน 27 ยีนที่มีความสัมพันธ์กับลักษณะต่าง ๆ อย่างมีนัยสำคัญ Saidou et al. (2014) ใช้ candidate-gene association mapping ในการยืนยันการมีอยู่ของยีน *PHYTOCHROME C* ที่ได้จากการวิเคราะห์แล้วว่ามีความสัมพันธ์กับช่วงเวลาในการออกดอกและความแปรปรวนของลักษณะสัญญาณวิทยาในลูกเต๋อยจำนวน 90 ตัวอย่าง โดยใช้เครื่องหมาย SNP และ InDel จากบริเวณโดยรอบยีน *PHYTOCHROME C* ทำการหาลำดับเบสแบบเต็มความยาวของยีน *PHYTOCHROME C* (ประมาณ 6 kb) และยีนบริเวณรอบๆ ยีนนี้ด้วยเทคโนโลยี Roche 454 platform (Roche Applied Science) และ PACBIO platform จากการวิเคราะห์ความสัมพันธ์ด้วย mixed model พบว่ามีหลายเครื่องหมายที่มีความสัมพันธ์กับช่วงเวลาในการออกดอกและลักษณะสัญญาณวิทยาต่างๆในลูกเต๋อย Yu et al. (2015) ศึกษา ยีนที่เกี่ยวข้องกับการอยู่รอดในฤดูหนาว และการเจริญเติบโตใหม่ในฤดูใบไม้ผลิของหญ้าไรย์ พบว่าความแปรปรวนของลำดับเบสซึ่งเกิดจาก synonymous substitution ของ SNPs ส่งผลให้มีการเปลี่ยนแปลงของกรดอะมิโนภายในยีนที่เกี่ยวข้องกับการอยู่รอดในฤดูหนาว และการเจริญเติบโตใหม่ในฤดูใบไม้ผลิของหญ้าไรย์ จึงสรุปได้ว่าความแปรปรวนของแอลลีลในยีนที่ศึกษามีความสัมพันธ์กับการอยู่รอดในฤดูหนาว และการเจริญเติบโตใหม่ในฤดูใบไม้ผลิของหญ้าไรย์ Singh et al. (2016) ได้ศึกษาความสัมพันธ์ระหว่างเครื่องหมาย EST-SSR กับลักษณะความต้านทานโรคเหี่ยวเน่าแดงในอ้อย โดยวิธี candidate-gene association mapping ในอ้อยจำนวน 119 ตัวอย่าง โดยใช้เครื่องหมาย SSR จำนวน 944 เครื่องหมาย จากการวิเคราะห์ด้วยวิธี MLM พบว่ามีเครื่องหมายจำนวน 4 เครื่องหมายที่สัมพันธ์กับลักษณะความต้านทานโรคเหี่ยวเน่าแดงในอ้อย และสามารถอธิบายความแปรปรวนของลักษณะฟีโนไทป์ได้ 10-16% Nigro et al. (2019) ศึกษา candidate genes association mapping ของปริมาณโปรตีนในเมล็ดข้าวสาลีดูรัม จากยีนที่สนใจจากกระบวนการต่างๆที่เกี่ยวข้องกับไนโตรเจนเมแทบอลิซึมจำนวน 14 ยีน พบว่ามียีนจำนวน 6 ยีนที่มีความสัมพันธ์อย่างมากกับลักษณะปริมาณโปรตีนในเมล็ด

บทที่ 3

วิธีการวิจัย

3.1 ตัวอย่างอ้อย และการสกัดดีเอ็นเอ

ตัวอย่างอ้อยที่ใช้ในการศึกษามีทั้งหมด 141 พันธุ์ (ภาคผนวก ก) เป็นเชื้อพันธุกรรมนำเข้า (germplasm) จำนวน 68 ตัวอย่างที่รวบรวมมาจากแหล่งในทวีปต่าง ๆ ทั่วโลก (ตัวอย่างที่ 1-68) และพันธุ์ปรับปรุง (breeding line) ในประเทศไทยจำนวน 73 ตัวอย่าง (ตัวอย่างที่ 69-141) ปลูกที่บริษัทมิตรผลวิจัยพัฒนาอ้อยและน้ำตาล จำกัด อำเภอกุฉินารายณ์ จังหวัดชัยภูมิ และมหาวิทยาลัยขอนแก่น จังหวัดขอนแก่น โดยนำไปอ่อนจากแต่ละตัวอย่างมาสกัดดีเอ็นเอด้วยวิธี CTAB (cetyltrimethylammonium bromide) ของ Gawel and Jarret (1991) แล้วตรวจสอบคุณภาพและความเข้มข้นของดีเอ็นเอด้วยวิธีอิเล็กโทรโฟรีซิสในเจลอะกาโรส (agarose gel electrophoresis) และวัดค่าการดูดกลืนแสง (spectrophotometer) ตามลำดับ

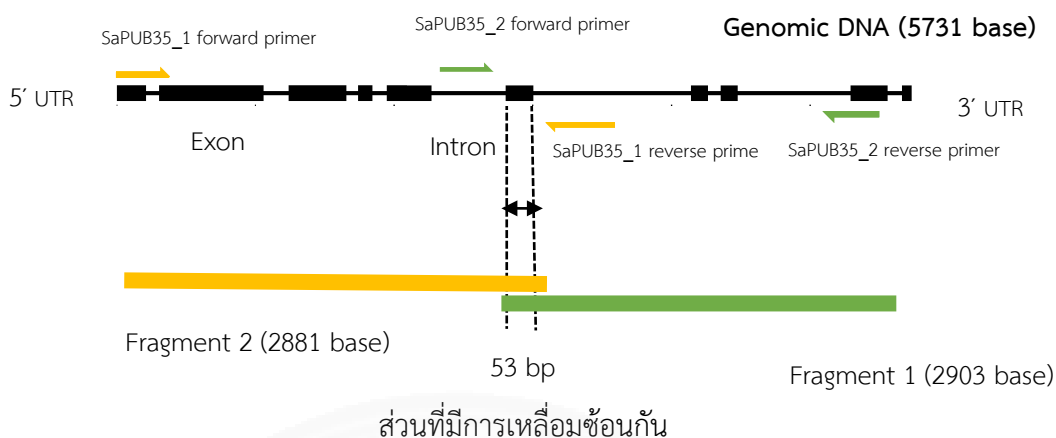
3.2 การเก็บข้อมูลฟีนไทป์ และการวิเคราะห์ค่าทางสถิติ

ดำเนินการปลูกอ้อยทั้ง 141 พันธุ์ จำนวน 2 ปี ปีแรกเป็นอ้อยปลูกและปีที่สองเป็นอ้อยต่อ ด้วยแผนการทดลองแบบ randomized complete block design จำนวน 2 ซ้ำ โดยปลูกอ้อยพันธุ์ละ 3 แถว แถวยาว 5 เมตร ระยะห่างระหว่างแถว 1.65 เมตร โดยปลูกอ้อยปลูกในช่วงเดือนพฤศจิกายนถึงธันวาคม 2560 ณ แปลงทดลองมหาวิทยาลัยขอนแก่น (KKU) จังหวัดขอนแก่น และศูนย์นวัตกรรมและการวิจัยมิตรผล (MPT) อำเภอกุฉินารายณ์ จังหวัดชัยภูมิ และปลูกอ้อยต่อในช่วงเดือนเดือนพฤศจิกายนถึงธันวาคม 2561 แล้วเก็บข้อมูลฟีนไทป์ของลักษณะที่เกี่ยวข้องกับน้ำตาลในอ้อยทั้ง 4 ลักษณะ คือ ค่าความเข้มข้นของของแข็งที่ละลายอยู่ในน้ำอ้อย (total soluble solid; Brix) ปริมาณน้ำตาลที่มีอยู่ในอ้อยที่สามารถหีบสกัดออกมาได้เป็นน้ำตาลทรายขาวบริสุทธิ์ (commercial cane sugar; CCS) ของแข็งที่ไม่ละลายน้ำ (Fiber) และค่าร้อยละของน้ำตาลซูโครส (sucrose content; POL) เมื่ออ้อยมีอายุครบ 12 เดือน ซึ่งดำเนินการโดยบริษัทมิตรผลวิจัย พัฒนาอ้อยและน้ำตาล จำกัด และคัดเลือกอ้อยพันธุ์พ่อแม่ของอ้อยพันธุ์ปรับปรุงซึ่งเป็นพันธุ์อ้อยของประเทศไทยผ่านกระบวนการปรับปรุงพันธุ์มาจำนวน 10 พันธุ์ และเก็บข้อมูลลักษณะดัชนีความเข้มของสีใบ (leaf color index) และประสิทธิภาพการสังเคราะห์ด้วยแสง (photosynthetic efficiency (fv/fm)) ที่อ้อยอายุ 3 เดือน เพื่อใช้เป็นข้อมูลในการทดสอบสมมุติฐานของอิทธิพลของ selective

sweep ทดสอบการแจกแจงแบบปกติของข้อมูลตามข้อตกลงเบื้องต้นในการวิเคราะห์ทางสถิติโดยใช้การทดสอบของ Kolmogorov-Smirnov ที่ระดับนัยสำคัญ 0.01 วิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันระหว่างลักษณะที่เกี่ยวข้องกับน้ำตาลทั้ง 4 ลักษณะด้วย PROC.CORR บน software package SAS 9.3 (SAS Institute Inc., Cary, NC, US) และวิเคราะห์ความแปรปรวน (ANOVA) ด้วย PROC.GLM บน software package SAS 9.3 (SAS Institute Inc., Cary, NC, US) ในการคำนวณค่าเฉลี่ยความแปรผัน (mean square) เพื่อประเมินค่าอัตราพันธุกรรมแบบกว้าง (H^2) แบ่งการคำนวณเป็น 2 แบบคือ แบบแยกแปลงการทดลองจากสมการ $H^2 = \frac{\sigma_g^2}{(\sigma_g^2 + \frac{\sigma_e^2}{r})}$ (Holland, Nyquist, & Cervantes-Martínez, 2003) และแบบรวมแปลงการทดลองจากสมการ $H^2 = \frac{\sigma_g^2}{(\sigma_g^2 + \frac{\sigma_{gn}^2}{r} + \frac{\sigma_e^2}{nr})}$ (Knapp, Stroup, & Ross, 1985) เมื่อ σ_g^2 คือค่าความแปรปรวนทางพันธุกรรม, σ_e^2 คือความแปรปรวนของการคลาดเคลื่อน, σ_{gn}^2 คือค่าความแปรปรวนทางพันธุกรรมที่มีปฏิสัมพันธ์กับสิ่งแวดล้อม, n คือจำนวนแปลงการทดลองหรือสิ่งแวดล้อม, และ r คือจำนวนซ้ำ

3.3 การออกแบบไพรเมอร์

นำลำดับเบส cDNA ของยีน *U-box domain-containing protein 35 (SaPUB35)* มาเปรียบเทียบกับลำดับเบสคอนทิกของอ้อยพันธุ์ขอนแก่น 3 (KK3) เพื่อทำนายตำแหน่งของเอกซอนและอินตรอน แล้วออกแบบไพรเมอร์ด้วยซอฟต์แวร์ Primer3 (Rozen & Skaletsky, 2000) ร่วมกับโปรแกรม Multiple Primer Analyzer (Breslauer, Frank, Blöcker, & Marky, 1986) แบ่งการออกแบบไพรเมอร์เพื่อเพิ่มปริมาณดีเอ็นเอ 2 ชั้น (fragment) คือไพรเมอร์ SaPUB35_1 ประกอบด้วยไพรเมอร์เส้น forward ขนาด 20 เบส (5'- AAACGCCTCGATCCATATCC-3') และไพรเมอร์เส้น reverse ขนาด 19 เบส (5'-CAATCCGAGGACTGAACGA-3') และ SaPUB35_2 ประกอบด้วยไพรเมอร์เส้น forward ขนาด 18 เบส (5'-TTTCCGTTCCGTTACTGC-3') และไพรเมอร์เส้น reverse ขนาด 20 เบส (5'- ACTTCTAGTGCTAGGTGAAG-3') โดยขนาดผลผลิตพีซีอาร์ของไพรเมอร์ SaPUB35_1 และ SaPUB35_2 จากการคาดคะเนมีขนาดประมาณ 2881 เบส และ 2903 เบส ตามลำดับ ซึ่งครอบคลุมความยาวของยีน 5731 bp โดยไพรเมอร์ทั้ง 2 คู่มิตำแหน่งเหลื่อมซ้อนกันระหว่างไพรเมอร์เส้น reverse ของคู่แรกกับไพรเมอร์เส้น forward ของคู่ที่สอง 53 bp อุณหภูมิที่ทำให้ดีเอ็นเอคลายเกลียวออกครึ่งหนึ่ง (melting temperature; Tm) อยู่ระหว่าง 50-65 °C และ Tm ของไพรเมอร์เส้น reverse และเส้น forward ต่างกันไม่เกิน 1-2 °C (ภาพที่ 3.1)



ภาพที่ 3.1 การออกแบบไพรเมอร์จาก cDNA ของยีน *SaPUB35* ขนาด 5731 base โดยแบ่งการออกแบบไพรเมอร์เพื่อเพิ่มปริมาณดีเอ็นเอเป็น 2 ชั้น ชั้นขนาด 2881 และ 2903 base การออกแบบไพรเมอร์ทั้ง 2 คู่ ไพรเมอร์เส้น reverse ของคู่แรก (*SaPUB35_1* reverse primer) มีส่วนที่เหลื่อมซ้อนกันกับไพรเมอร์เส้น forward ของคู่ที่สอง (*SaPUB35_2* forward primer) 53 base

3.4 การเพิ่มปริมาณดีเอ็นเอและการทำให้บริสุทธิ์

3.4.1 การเพิ่มปริมาณดีเอ็นเอ

เพิ่มปริมาณดีเอ็นเอด้วยการทำพีซีอาร์สองรอบ ในการทำพีซีอาร์รอบแรกนั้นใช้ไพรเมอร์จำเพาะที่ออกแบบมาจากยีน *SaPUB35* ซึ่งติดอยู่กับลำดับเบส M13 มีปริมาตร $10\ \mu\text{l}$ ประกอบด้วยดีเอ็นเอต้นแบบ 10 ng, 5x Phusion HF buffer (MgCl_2 7.5 mM), dNTPs 2 mM, คู่ไพรเมอร์อย่างละ 0.005 mM และ 0.2 units Phusion taq ใช้สภาวะในการทำพีซีอาร์ที่อุณหภูมิ $94\ ^\circ\text{C}$ 2 นาที ตามด้วย $94\ ^\circ\text{C}$ 20 วินาที 30 รอบ อุณหภูมิ ($56\text{-}63\ ^\circ\text{C}$) 20 วินาที $72\ ^\circ\text{C}$ 1.30 นาที และสุดท้ายที่อุณหภูมิ $72\ ^\circ\text{C}$ 5 นาทีและพีซีอาร์รอบที่สองเป็นการติดลำดับเบสบาร์โคดเพื่อใช้แยกความแตกต่างระหว่างตัวอย่าง โดยใช้ไพรเมอร์ที่มีลำดับเบสคู่สมกับลำดับเบส M13 ซึ่งมีลำดับเบสบาร์โคดที่ติดอยู่ที่ปลาย M13 มีปริมาตร $50\ \mu\text{l}$ ประกอบด้วยดีเอ็นเอต้นแบบที่เจือจางจากผลผลิตพีซีอาร์รอบแรกในอัตราส่วน 1:50, 5x Phusion HF buffer (MgCl_2 7.5 mM), MgCl_2 75 mM, dNTPs 10 mM, คู่ไพรเมอร์อย่างละ 5 mM และ 1 units Phusion taq สภาวะในการทำพีซีอาร์ที่อุณหภูมิ $98\ ^\circ\text{C}$ 2 นาที ตามด้วย $98\ ^\circ\text{C}$ 30 วินาที จำนวน 4 รอบ อุณหภูมิ $63\ ^\circ\text{C}$ 15 วินาที $72\ ^\circ\text{C}$ 45 วินาที $98\ ^\circ\text{C}$ 30 วินาที อีกจำนวน 6 รอบ ตามด้วยอุณหภูมิ $66\ ^\circ\text{C}$ 15 วินาที $72\ ^\circ\text{C}$ 45 วินาที และสุดท้ายที่อุณหภูมิ $72\ ^\circ\text{C}$ 5 นาที แล้ววิเคราะห์ผลผลิตพีซีอาร์ที่ได้จากการทำปฏิกิริยา

ลูกโซ่พอลิเมอร์ทั้ง 2 รอบเบื้องต้นด้วยวิธีอิเล็กโทรโฟรีซิสในเจลอะกาโรสความเข้มข้น 2% เพื่อตรวจสอบขนาดของชิ้นดีเอ็นเอที่ต้องการตามที่คาดคะเนจากการออกแบบคูไพรเมอร์

3.4.2 การทำผลผลิตพีซีอาร์ให้บริสุทธิ์

นำผลผลิตพีซีอาร์ที่ได้จากการทำพีซีอาร์ครั้งที่ 2 มาทำให้บริสุทธิ์ด้วย AMPure XP purification kit (Beckman Coulter, Brea CA, USA) และวัดความเข้มข้นด้วย Qubit® 2.0 Fluorometer โดยใช้ชุดคิท Qubit® dsDNA BR Assay (Invitrogen, Oregon, USA) จากนั้นผสมรวมผลผลิตพีซีอาร์จากทุกตัวอย่างแยกเป็นชิ้นที่ 1 (คลังดีเอ็นเอ 1) และชิ้นที่ 2 (คลังดีเอ็นเอ 2) ให้ได้ความเข้มข้นสุดท้ายของแต่ละคลังดีเอ็นเอเท่ากับ 50 ng ปริมาตร 30 μ l เพื่อเตรียมสร้างคลัง SMRTbell และวิเคราะห์ลำดับเบสด้วยเครื่อง PacBio® RS High Resolution Genetic Analyzer

3.5 การค้นหาเครื่องหมาย SNP และอินเดล (Insertion/Deletion; InDel)

ค้นหาเครื่องหมาย SNP และ InDel โดยใช้ genome analysis tool kit (GATK) ของ McKenna et al. (2010) จากข้อมูลลำดับเบสของทุกตัวอย่างที่ได้จากเทคโนโลยี PacBio SMRT sequencing มาเทียบกับลำดับจีโนมโมโนพลอยด์ของอ้อยพันธุ์ R570 (Garsmeur et al., 2018) จากนั้นทำการคัดกรองเครื่องหมายจากค่า missing data ที่ระดับน้อยกว่า 20% ค่า และค่า minor allele frequency (MAF) ที่ระดับ 0.05 ของแต่ละเครื่องหมาย

3.6 การวิเคราะห์ความไม่สมดุลของลิงเกจ (linkage disequilibrium; LD) และการเสื่อมถอยของความไม่สมดุลของลิงเกจ (LD decay)

วิเคราะห์ LD โดยใช้ค่ายกกำลังสองของค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างเครื่องหมายสองตำแหน่ง (squared allele frequency correlations; r^2) ของคู่เครื่องหมาย SNP และ InDel ด้วยโปรแกรม Haploview 4.2 (Barrett et al., 2005) จากนั้นวิเคราะห์แผนภาพการกระจายตัวของค่า r^2 ระหว่างคู่เครื่องหมายด้วยสมการแบบไม่เป็นเชิงเส้นตรง ระหว่างแกน X ซึ่งเป็นระยะห่างทางกายภาพ (bp) และแกน Y เป็นค่า r^2 และหาค่าระยะห่างทางกายภาพของ LD decay จากแบบจำลองถดถอยที่ไม่เป็นเชิงเส้น (nonlinear regression) (Zia et al., 2020) ของค่า r^2 บนระยะห่างทางกายภาพโดยกำหนดค่า r^2 ที่ 0.1 เป็นค่าขั้นต่ำของการมีนัยสำคัญของการเกิด LD

ระหว่างคู่เครื่องหมาย (Zhu *et al.*, 2008) แบ่งการวิเคราะห์ LD decay เป็น 2 ประชากรคือ ประชากรตัวอย่างที่เป็นเชื้อพันธุกรรมนำเข้าและประชากรตัวอย่างที่เป็นสายพันธุ์ปรับปรุง

3.7 การวิเคราะห์แผนที่ความสัมพันธ์ภายในยีน (candidate gene association mapping)

วิเคราะห์แผนที่ความสัมพันธ์ระหว่างเครื่องหมาย SNP และ InDel กับลักษณะที่เกี่ยวข้องกับน้ำตาลทั้ง 4 ลักษณะ โดยใช้โมเดลทดสอบทั้งหมด 4 แบบคือ 1) simple general linear model (GLM) 2) GLM ที่ใช้ค่า Q matrix เป็นตัวแปรร่วม (GLM+Q), 3) Mixed linear model (MLM) ที่ใช้ค่า Q matrix และค่าสัมประสิทธิ์ความเป็นเครือญาติ (kinship coefficient, K) เป็นตัวแปรร่วม (MLM+Q+K) ด้วยโปรแกรม TASSEL 5.2.56 และ 4) FarmCPU (Fixed and random model Circulating Probability Unification) ที่ใช้ค่า Q และค่า K เป็นตัวแปรร่วม โดยใช้ R software (R Core Team, 2018) ในการอนุมานค่า Q และค่า K ใช้เครื่องหมาย ILP และ SSR จำนวน 384 เครื่องหมาย ที่พัฒนามาจากงานวิจัยก่อนหน้าของ ธัญธร สุวภาพโพบูลย์กุล (2561) ซึ่งเป็นชุดข้อมูลที่มีความเป็นอิสระจากชุดข้อมูลของเครื่องหมายที่ใช้ในการวิเคราะห์แผนที่ความสัมพันธ์เพื่อลดความสัมพันธ์ที่ไม่แท้จริงจากโครงสร้างประชากรและความเป็นเครือญาติ (Yu & Buckler, 2006) โดยใช้เกณฑ์ p-value 0.01 ในการระบุเครื่องหมายที่มีแนวโน้มที่สัมพันธ์กับลักษณะที่เกี่ยวข้องกับน้ำตาลทั้ง 4 ลักษณะ จากนั้นปรับค่า p-value เป็นค่า false discovery rate (FDR) ตามวิธีของ Benjamini *et al.* (2001) ด้วย QVALUE บนโปรแกรม R (Storey, 2002) เพื่อลดโอกาสการเกิดผลทางบวกที่ไม่ถูกต้องและระบุความสัมพันธ์แบบมีนัยสำคัญระหว่างเครื่องหมาย SNP และ InDel กับลักษณะที่เกี่ยวข้องกับน้ำตาลทั้ง 4 ลักษณะ โดยใช้เกณฑ์ FDR ที่ $q < 0.10$ และเกณฑ์ quantile–quantile (Q-Q) plots จากแต่ละโมเดลทดสอบที่ได้จากการวิเคราะห์ด้วย qqman package บน R software (R Core Team, 2018) เพื่อเลือกโมเดลทดสอบที่เหมาะสมในการวิเคราะห์แผนที่ความสัมพันธ์

บทที่ 4

ผลการวิจัยและอภิปรายผล

4.1 การวิเคราะห์ทางสถิติของข้อมูลฟีโนไทป์

ผลการทดสอบการแจกแจงแบบปกติของข้อมูลฟีโนไทป์ของลักษณะที่เกี่ยวข้องกับลักษณะน้ำตาลในอ้อยทั้ง 4 ลักษณะ พบว่าข้อมูลฟีโนไทป์ของทั้งอ้อยปลูกและอ้อยต่อจากทั้ง 2 แปลงการทดลองเกือบทุกลักษณะมีการกระจายตัวแบบปกติ จากการทดสอบทางสถิติของ Kolmogorov-Smirnov ที่ระดับนัยสำคัญ 0.01 ยกเว้นลักษณะ CCS ของอ้อยต่อ แปลงการทดลอง มหาวิทยาลัยขอนแก่น (KKU) ที่ไม่มีการกระจายตัวแบบปกติ จึงแปลงข้อมูลฟีโนไทป์เดิมของ CCS ด้วยการยกกำลังสอง (square transformation) เพื่อให้มีการกระจายตัวแบบปกติ (ภาคผนวก ข) การวิเคราะห์ความแปรปรวนทางพันธุกรรมของลักษณะที่เกี่ยวข้องกับลักษณะน้ำตาลในอ้อยระหว่างตัวอย่างอ้อยแต่ละพันธุ์ พบว่ามีความแตกต่างกันอย่างมีนัยสำคัญยิ่ง ($p < 0.001$) และจากการวิเคราะห์ค่าความแปรปรวนจีโนไทป์ (genetic variance; σ_g^2) และค่าความแปรปรวนฟีโนไทป์ (phenotypic variance; σ_p^2) พบว่ามีค่าแตกต่างกันไปตามแต่ละลักษณะทั้งในอ้อยปลูกและอ้อยต่อทั้งในแต่ละแปลงการทดลอง และแบบรวมแปลงการทดลอง ซึ่งโดยทั่วไปแล้วค่าความแปรปรวนฟีโนไทป์จะมีค่ามากกว่าค่าความแปรปรวนจีโนไทป์ดังแสดงในตารางที่ 4.1 จากการประเมินค่าอัตราพันธุกรรมแบบกว้าง (H^2) ของลักษณะที่เกี่ยวข้องกับลักษณะน้ำตาลทั้ง 4 ลักษณะจากทั้งอ้อยปลูกและอ้อยต่อแบบแยกแปลงการทดลอง พบว่ามีความผันแปรไปตามแต่ละลักษณะในแต่ละแปลงการทดลอง จากตารางที่ 4.1 แสดงค่า H^2 พบว่ามีค่าอยู่ในระดับสูง (54.13-85.87%) โดยค่า H^2 ของลักษณะ Brix มีค่าตั้งแต่ 76.13-80.98% ลักษณะ CCS มีค่า H^2 อยู่ในช่วง 76.07-82.89% ขณะที่ค่า H^2 ของลักษณะ Fiber มีค่าระหว่าง 54.13-85.87% และค่า H^2 ของลักษณะ POL มีค่าตั้งแต่ 76.55-83.26% ซึ่งค่า H^2 ที่ประเมินได้ของลักษณะ Brix, Fiber และ POL นั้นมีค่าใกล้เคียงกับงานวิจัยของ Fickett et al. (2019) ซึ่งประเมินค่า H^2 ของลักษณะ Brix, Fiber, และ POL ได้เท่ากับ 78, 88, และ 77% ตามลำดับ ขณะที่ Gouy et al. (2015) และ Yang et al. (2020) ประเมินค่า H^2 ของลักษณะ Brix ได้เท่ากับ 61 และ 68% ตามลำดับ ซึ่งมีค่าน้อยกว่าค่าที่ประเมินได้จากงานวิจัยนี้ ซึ่งอาจจะมีสาเหตุมาจากสภาพของสิ่งแวดล้อมที่แตกต่างกัน เมื่อประเมินค่า H^2 แบบรวมแปลงการทดลองพบว่าในอ้อยปลูกมีค่าอยู่ในระดับสูง (77.53-82.43%) แต่พบว่าในอ้อยต่อนั้นมีค่าอยู่ในระดับกลางถึงระดับสูง (49.22-57.22%) จะเห็นได้ว่าค่า H^2 ที่ประเมินได้จากทุกลักษณะจากแต่ละแปลงการทดลอง และจากการรวมแปลงการทดลองนั้นส่วนใหญ่มีค่าอยู่ในระดับสูงบ่งชี้ถึงการจัดการแปลงทดลองที่

เหมาะสมสามารถควบคุมความแปรปรวนจากสิ่งแวดล้อมได้ดี ทำให้ความแปรปรวนของลักษณะที่เกี่ยวข้องกับลักษณะน้ำตาลทั้ง 4 ลักษณะนั้นเกิดจากความแปรปรวนทางพันธุกรรมมากกว่าความแปรปรวนจากสิ่งแวดล้อมหรือกล่าวได้ว่าความแปรปรวนของลักษณะที่เกี่ยวข้องกับลักษณะน้ำตาลทั้ง 4 ลักษณะได้รับอิทธิพลจากยีนมากกว่าสิ่งแวดล้อม ขณะที่ค่า H^2 ของบางลักษณะที่ประเมินจากการรวมแปลงการทดลองมีค่าอยู่ในระดับกลางนั้นอาจจะเกิดจากอิทธิพลของการรวมแปลงการทดลองหรือสิ่งแวดล้อมจึงทำให้มีค่า H^2 ที่ต่ำกว่าการประเมินแบบแยกแปลงการทดลอง

เพื่อทดสอบสมมติฐานของการคัดเลือกแบบ selective sweep เพื่อเก็บอ้อยพันธุ์ดีจากกระบวนการปรับปรุงพันธุ์อ้อย จึงประเมินความแปรปรวนฟีโนไทป์ของลักษณะที่เกี่ยวข้องกับน้ำตาลทั้ง 4 ลักษณะ เปรียบเทียบระหว่างอ้อยพันธุ์เชื้อพันธุกรรมนำเข้าและอ้อยพันธุ์พ่อแม่ของอ้อยพันธุ์ปรับปรุงในประเทศไทยซึ่งเป็นพันธุ์ที่เกิดจากการปรับปรุงพันธุ์โดยนักปรับปรุงพันธุ์เช่นกัน พบว่าความแปรปรวนฟีโนไทป์ของลักษณะที่เกี่ยวข้องกับน้ำตาลส่วนใหญ่ในอ้อยพันธุ์พ่อแม่พันธุ์ปรับปรุงในประเทศไทยมีค่าที่แคบกว่าอ้อยที่เป็นเชื้อพันธุกรรมนำเข้า (ตารางที่ 4.2) เช่นเดียวกับค่าความแปรปรวนฟีโนไทป์ของลักษณะดัชนีความเข้มของสีใบ (ตารางที่ 4.3) ระหว่างตัวอย่างอ้อยเชื้อพันธุกรรมนำเข้าและอ้อยพันธุ์ปรับปรุงในประเทศไทยซึ่งพบว่าค่าความแปรปรวนฟีโนไทป์ของลักษณะดัชนีความเข้มของสีใบของอ้อยพันธุ์ปรับปรุงในประเทศไทยมีค่าที่แคบกว่าตัวอย่างอ้อยเชื้อพันธุกรรมนำเข้ารวมถึงค่าเฉลี่ยส่วนใหญ่ของลักษณะที่เกี่ยวข้องกับน้ำตาลทั้ง 4 ลักษณะและค่าลักษณะดัชนีความเข้มของสีใบของอ้อยพันธุ์พ่อแม่พันธุ์ปรับปรุงในประเทศไทยมีค่าเฉลี่ยที่มากกว่าอ้อยที่เป็นเชื้อพันธุกรรมนำเข้าบ่งชี้ถึงอิทธิพลของ selective sweep ที่มีผลต่อความแปรปรวนและค่าเฉลี่ยของลักษณะทางการเกษตรที่สำคัญอันเนื่องมาจากการผลของการคัดเลือกในกระบวนการปรับปรุงพันธุ์

ตารางที่ 4.1 ค่าเฉลี่ย ช่วงค่า ค่าความแปรปรวนจีโนไทป์ ค่าความแปรปรวนฟีโนไทป์ และค่าอัตราพันธุกรรมแบบกว้าง (broad-sense heritability; H^2) ของลักษณะที่เกี่ยวข้องกับลักษณะน้ำตาลทั้ง 4 ลักษณะ ในอ้อยปลูก (plant cane) และอ้อยตอ (ratoon) ของแปลงการทดลองมหาวิทยาลัยขอนแก่น (KKU) และแปลงการทดลองศูนย์นวัตกรรมและการวิจัยมิตรผล (MPT)

	Statistics	Plant cane KKU	Plant cane MPT	Mean Plant cane	1 st Ratoon KKU	1 st Ratoon MPT	Mean 1 st Ratoon
Brix	Mean±SD	19.75±1.80	19.69±1.91	19.74±1.70	18.9±1.55	19.97±1.58	19.50±1.33
	Range	14.94-23.84	14.57-24.37	14.76-23.74	15.00-22.50	15.71-23.31	15.94-22.77
	σ_g^2	2.52	2.78	5.03	1.67	1.98	2.77
	σ_p^2	3.23	3.66	6.49	2.06	2.45	5.09
	H^2 (%)	78.19	76.13	77.53	80.71	80.98	54.33
CCS	Mean±SD	11.12±2.71	10.8±2.70	10.99±2.50	8.25±2.62	10.68±2.18	9.63±2.10
	Range	1.21-16.66	1.84-15.37	1.99-15.74	1.34-12.87	2.97-15.82	2.21-14.24
	σ_g^2	5.14	4.77	9.68	1147.99	3.28	2077.36
	σ_p^2	6.55	5.93	12.25	1384.96	4.31	4220.27
	H^2 (%)	78.56	80.44	78.99	82.89	76.07	49.22
Fiber	Mean±SD	14.66±3.54	14.31±3.2	14.48±3.28	15.64±3.22	14.16±2.46	14.78±2.56
	Range	10.00-32.00	9.84-28.33	9.92-30.01	9.44-30.24	10.21-24.26	10.21-26.81
	σ_g^2	3.83	2.78	6.57	3.38	1.78	4.85
	σ_p^2	4.46	3.47	7.97	4.07	3.29	8.47
	H^2 (%)	85.87	79.95	82.43	83.21	54.13	57.22
POL	Mean±SD	15.93±2.61	15.61±2.73	15.80±2.47	13.26±2.48	15.61±2.24	14.59±2.07
	Range	7.76-21.84	5.92-20.86	6.84-20.81	5.58-18.33	8.29-20.41	6.94-19.28
	σ_g^2	5.27	6.23	10.86	4.32	3.54	6.47
	σ_p^2	6.83	7.48	13.65	5.32	4.62	12.8
	H^2 (%)	77.18	83.26	79.56	81.25	76.55	50.61

σ_g^2 genetic variance, σ_p^2 phenotypic variance

The square transformed data was applied for 1st ratoon cane CCS of Khon Kaen University location.

ตารางที่ 4.2 ค่าเฉลี่ย ช่วงค่า และค่าความแปรปรวนฟีโนไทป์ ของลักษณะที่เกี่ยวข้องกับลักษณะน้ำตาลทั้ง 4 ลักษณะ ในอ้อยปลูก (plant cane) และอ้อยตอ (ratoon) ระหว่างตัวอย่างอ้อยเชื้อพันธุ์กรรมนำเข้าและอ้อยพันธุ์พ่อแม่ของอ้อยพันธุ์ปรับปรุงในประเทศไทย

Traits	Statistics	Germplasm				Breeding line parents			
		Plant cane KKU	Plant cane MPT	1 st Ratoon KKU	1 st Ratoon MPT	Plant cane KKU	Plant cane MPT	1 st Ratoon KKU	1 st Ratoon MPT
Brix	Mean±SD	20.14±1.64	20.11±1.84	19.16±1.47	20.32±1.55	20.06±1.42	20.40±1.79	18.53±1.39	20.84±0.85
	Range	16.63-23.46	14.84-24.37	16.05-22.50	16.22-23.31	16.86-21.77	17.05-23.62	15.92-20.45	19.51-22.27
	σ_p^2	2.71	3.43	1.77	2.34	2.23	3.54	1.92	0.80
CCS	Mean±SD	11.49±2.39	11.01±2.38	8.85±2.16	11.76±1.27	12.47±1.40	12.39±1.83	7.71±2.99	12.05±1.10
	Range	1.21-15.55	4.40-14.94	4.42-12.87	7.21-12.56	9.35-13.77	9.16-15.31	2.66-11.55	9.92-13.48
	σ_p^2	5.81	5.73	3.82	3.58	1.74	3.70	8.96	1.35
Fiber	Mean±SD	14.68±2.46	14.38±2.30	15.31±2.58	14.14±2.18	13.24±1.33	13.67±1.11	15.08±2.56	13.93±1.46
	Range	10.52-24.37	10.01-22.25	11.26-25.14	10.21-24.26	11.46-15.93	12.10-15.36	12.33-20.77	10.89-16.12
	σ_p^2	4.73	5.35	4.02	4.62	1.97	1.36	6.56	2.38
POL	Mean±SD	16.37±2.27	15.94±2.43	13.80±1.93	16.06±2.06	17.03±1.63	17.25±2.07	12.78±2.88	16.97±1.23
	Range	8.44-21.16	8.94-20.56	10.08-17.94	10.73-20.41	13.79-18.65	13.74-20.86	7.76-15.65	14.67-18.72
	σ_p^2	4.95	6.00	3.07	4.06	1.00	4.77	8.31	1.69

σ_p^2 phenotypic variance

ตารางที่ 4.3 ค่าเฉลี่ย ช่วงค่า และค่าความแปรปรวนฟีโนไทป์ของลักษณะดัชนีความเข้มของสีใบ (leaf color index) และประสิทธิภาพการสังเคราะห์ด้วยแสง (photosynthetic efficiency (fv/fm)) ระหว่างตัวอย่างอ้อยเชื้อพันธุกรรมนำเข้าและอ้อยพันธุ์ปรับปรุงในประเทศไทย

Traits	Germplasm			Breeding line		
	Mean	Range	σ_p^2	Mean	Range	σ_p^2
Leaf color index	39.67±3.31	31.2-47.5	10.96	41.54±3.23	35.9-51.2	10.4
Photosynthetic efficiency (fv/fm)	0.807±0.01	0.781-0.832	0.0001	0.806±0.03	0.614-0.831	0.0006

σ_p^2 phenotypic variance

ตารางที่ 4.4 ค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันระหว่างลักษณะที่เกี่ยวข้องกับน้ำตาลทั้ง 4 ลักษณะ ในอ้อยปลูกและอ้อยต่อของทั้ง 2 แปลงการทดลอง

Crop cycle/Location		CCS	Fiber	POL
Plant KКУ (2017-2018)	Brix	0.752**	0.007 ^{ns}	0.823**
	CCS		-0.359**	0.946**
	Fiber			-0.194 ^{ns}
Plant MPT (2017-2018)	Brix	0.772**	0.162 ^{ns}	0.864**
	CCS		-0.187 ^{ns}	0.968**
	Fiber			-0.023 ^{ns}
1 st Ratoon KКУ (2018-2019)	Brix	0.618**	0.018 ^{ns}	0.742**
	CCS		-0.447**	0.938**
	Fiber			-0.320**
1 st Ratoon MPT (2018-2019)	Brix	0.862**	0.147 ^{ns}	0.936**
	CCS		-0.113 ^{ns}	0.979**
	Fiber			0.039 ^{ns}

**Significant at $P < 0.001$. ^{ns} non-significant.

จากการวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันระหว่างลักษณะที่เกี่ยวข้องกับน้ำตาลทั้ง 4 ลักษณะในอ้อยปลูกและอ้อยต่อของทั้ง 2 แปลงการทดลอง จากตารางที่ 4.4 พบว่าลักษณะ Brix, CCS, และ POL มีความสัมพันธ์เชิงบวกอย่างมีนัยสำคัญยิ่ง ($p < 0.001$) ซึ่งกันและกัน ในทุกชุดข้อมูลพีโนไทป์ โดยค่าสัมประสิทธิ์สหสัมพันธ์น้อยที่สุดมีค่าเท่ากับ 0.618 ระหว่างลักษณะ Brix กับ CCS ของอ้อยต่อแปลงมหาวิทยาลัยขอนแก่น และค่าสัมประสิทธิ์สหสัมพันธ์มากที่สุดระหว่างลักษณะ CCS และ POL ของอ้อยต่อ แปลงการทดลองศูนย์นวัตกรรมและการวิจัยมิตรผล มีค่าเท่ากับ 0.979 แสดงให้เห็นว่าค่า Brix ,CCS, และ POL มีแนวโน้มเพิ่มขึ้นไปด้วยกัน บ่งชี้ได้ว่าค่า Brix และ POL ส่งผลต่อการเพิ่มขึ้นของค่า CCS และลักษณะ Fiber ในอ้อยปลูกของแปลงมหาวิทยาลัยขอนแก่น มีความสัมพันธ์เชิงลบอย่างมีนัยสำคัญยิ่ง ($p < 0.001$) กับ CCS เช่นเดียวกับในอ้อยต่อที่ลักษณะ Fiber มีความสัมพันธ์เชิงลบอย่างมีนัยสำคัญยิ่ง ($p < 0.001$) กับ CCS และ POL แสดงว่าหากค่า Fiber ลดลงมีแนวโน้มว่าค่า CCS และ POL จะเพิ่มมากขึ้น ซึ่งให้ผลสอดคล้องกับงานวิจัยของ Fickett et al. (2019) และ Ukoskit et al. (2019) ซึ่งพบว่าลักษณะ Brix กับ POL มีความสัมพันธ์เชิงบวกต่อกันอย่างมาก และทั้ง 2 ลักษณะนี้ยังมีความสัมพันธ์เชิงลบกับลักษณะ Fiber อีกด้วย

4.2 การวิเคราะห์ลำดับเบสด้วยเทคโนโลยี PacBio SMRT Sequencing

ผลจากการวิเคราะห์ลำดับเบสของยีน *U-box domain-containing protein 35 (SaPUB35)* ด้วยเทคโนโลยี PacBio SMRT Sequencing ในตัวอย่างอ้อยทั้ง 141 พันธุ์พบว่าสามารถวิเคราะห์ลำดับเบสสายยาว (read) ที่มีความยาวตั้งแต่ 2000 เบสขึ้นไปได้จำนวน 17048 reads และทุก read สามารถ map บนจีโนมอ้างอิงได้ (100%) และความยาวของ read เฉลี่ยของชั้นที่ 1 เท่ากับ 2887 และชั้นที่ 2 เท่ากับ 2880 เบส มีค่าจำนวนซ้ำของข้อมูลเบสตำแหน่งเดิมหรือค่า coverage เท่ากับ 26x โดยมีค่า read depth ตั้งแต่ 12-38 เนื่องจากความสามารถในการวิเคราะห์ลำดับเบสสายยาวของเทคโนโลยี PacBio sequencing จึงมีงานวิจัยก่อนหน้านี้ที่ใช้เทคโนโลยีนี้ในการวิเคราะห์ลำดับเบสของอ้อยเพื่อวัตถุประสงค์ในด้านต่าง ๆ ทั้งการประกอบลำดับเบสจีโนมของไมโทคอนเดรียของอ้อยพันธุ์ขอนแก่น 3 โดย Shearman et al. (2016) รวมถึงการวิเคราะห์ลำดับเบสของโคลน bacteria artificial chromosome (BAC) ของ minimum tilling path (MTP) ในอ้อยพันธุ์ R570 ของ Garsmeur et al. (2018) และยังถูกใช้ในการศึกษาทรานสคริปโตมโดย Hoang et al. (2017) และ Piriyaongsa et al. (2018) อีกด้วย จะเห็นได้ว่าการวิเคราะห์ลำดับเบสด้วย

เทคโนโลยี PacBio sequencing มีประสิทธิภาพในการวิเคราะห์ลำดับเบสในอ้อยซึ่งจัดว่าเป็นพืชที่มีจีโนมขนาดใหญ่และมีความซับซ้อนที่สุดชนิดหนึ่ง

4.3 ความแปรปรวนของเครื่องหมาย SNP และ InDel

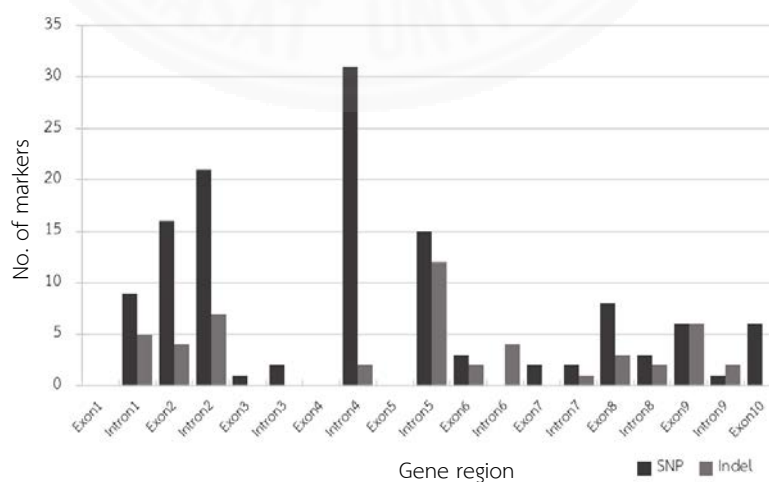
สามารถค้นหาเครื่องหมาย SNP และ InDel บนยีน *SaPUB35* ในประชากรรวม 141 ตัวอย่าง ได้ทั้งหมด 326 และ 192 เครื่องหมาย ตามลำดับ เมื่อคัดกรองเครื่องหมายด้วยค่า missing data ที่ระดับน้อยกว่า 20% และค่า MAF ที่ระดับ 0.05 พบ SNP แบบ bi-allelic ที่ผ่านการคัดกรองจำนวน 126 SNPs และเครื่องหมาย InDel ที่ผ่านการคัดกรองจำนวน 50 InDels (ตารางที่ 4.5) เมื่อเปรียบเทียบความแปรปรวนของดีเอ็นเอระหว่างตัวอย่างที่เป็นประชากรกลุ่มเชื้อพันธุกรรมนำเข้าและอ้อยพันธุ์ปรับปรุง พบว่าประชากรเชื้อพันธุกรรมนำเข้าจำนวน 68 ตัวอย่างมีเครื่องหมาย SNP และ InDel ที่ผ่านการคัดกรองจำนวน 184 เครื่องหมาย และในประชากรอ้อยที่เป็นพันธุ์ปรับปรุงจำนวน 73 ตัวอย่าง มีเครื่องหมายที่ผ่านการคัดกรองเพียงจำนวน 163 เครื่องหมาย จะเห็นได้ว่าความแปรปรวนเครื่องหมายที่ระบุได้ในประชากรอ้อยพันธุ์ปรับปรุงมีจำนวนน้อยกว่าในประชากรที่เป็นเชื้อพันธุกรรมนำเข้า แสดงให้เห็นถึงความหลากหลายทางพันธุกรรมที่ลดลงของยีน *Plant U-box* ในประชากรอ้อยพันธุ์ปรับปรุง อันเป็นผลจากการคัดเลือกโดยนักปรับปรุงพันธุ์เพื่อพัฒนาอ้อยสายพันธุ์ดี

เครื่องหมายที่ผ่านการคัดกรองมีการกระจายตัวอยู่เกือบทั่วทั้งยีน และพบเครื่องหมายทั้งสองชนิดบริเวณอินตรอน (intron) มากกว่าเอกซอน (exon) เนื่องจากเกิดการกลายของลำดับเบสสะสมมากบนบริเวณอินตรอนในระหว่างการเกิดวิวัฒนาการของยีน ส่งผลให้บริเวณอินตรอนมีความแปรปรวนของลำดับเบสมากกว่าบริเวณเอกซอน หรือกล่าวได้ว่าบริเวณอินตรอนมีความอนุรักษ์น้อยกว่าเมื่อเทียบกับบริเวณเอกซอน (Jaikishan, Rajendrakumar, Madhusudhana, Elangovan, & Patil, 2015) เพื่อให้บริเวณเอกซอนซึ่งเป็นบริเวณที่มีรหัสพันธุกรรมยังคงมีความอนุรักษ์ลำดับเบสและเกิดการถอดรหัส แปลรหัสเพื่อสร้างโปรตีนที่ทำหน้าที่ปกติได้ (Mukherjee et al., 2018) ซึ่งสอดคล้องกับการศึกษาก่อนหน้าของ McIntyre et al. (2006) และ Song et al. (2016) ซึ่งจากจำนวน SNP ทั้งหมด 126 SNPs มีตำแหน่งอยู่บน intron จำนวน 84 SNPs (67%) และอยู่บน exon 42 SNPs (33%) และจาก InDel จำนวน 50 InDels มีตำแหน่งอยู่บน intron และ exon จำนวน 35 InDels (70%) และ 15 InDels (30%) ตามลำดับ

ตารางที่ 4.5 จำนวนเครื่องหมาย SNP และ InDel บนยีน *SaPUB35* ที่ได้จากการค้นหาด้วยวิธี GATK

	No. of SNP	No. of InDel
Total raw	326	192
Missing < 20%	306	182
Bi-allelic	300	158
Multi-allelic	6	24
MAF \geq 0.05	126	50

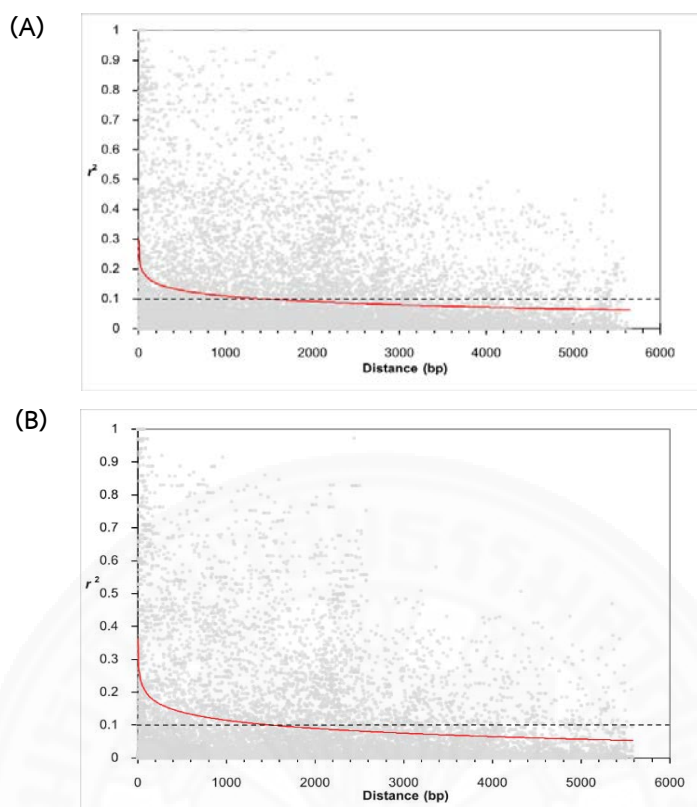
เมื่อพิจารณาการกระจายตัวของเครื่องหมายดีเอ็นเอทั้ง 2 ชนิดบนยีน พบว่าบริเวณ Intron 4 พบ SNP มากที่สุดถึง 31 SNPs ขณะที่บริเวณที่พบ InDels มากที่สุดคือบริเวณ Intron 5 จำนวน 12 InDels และบริเวณที่ไม่พบเครื่องหมายดีเอ็นเอทั้ง 2 ชนิดเลยคือบริเวณ Exon1, Exon4, และ Exon5 (ภาพที่ 4.1) เมื่อวิเคราะห์ความถี่ของการพบเครื่องหมายดีเอ็นเอทั้ง 2 ชนิดบนยีนนี้ซึ่งมีขนาดความยาวของยีนเท่ากับ 5731 เบสพบว่า สามารถพบ InDels ทุก ๆ 115 เบส และ พบ SNP โดยเฉลี่ยทุก ๆ 46 เบส ซึ่งใกล้เคียงกับการศึกษาของ McIntyre et al. (2006) ศึกษาและระบุแอลลีลใน sucrose phosphate synthase gene family III ในอ้อยพบว่าสามารถระบุ SNP ภายในยีนได้ทุก ๆ 40 เบส ขณะที่การศึกษาและระบุ SNP ภายใน ESTs (expressed sequence tags) ในอ้อยของ Cordeiro et al. (2006) สามารถระบุ SNP ได้ทุก ๆ 50 เบสของลำดับเบสที่มีการถอดรหัส



ภาพที่ 4.1 การกระจายตัวของเครื่องหมาย SNP และ InDel บนยีน *SaPUB35*

4.4 การไม่สมดุลของลิงเกจ (linkage disequilibrium; LD) และการเสื่อมถอยของความไม่สมดุลของลิงเกจ (LD decay)

จากการวิเคราะห์ LD decay เปรียบเทียบระหว่างตัวอย่างที่เป็นประชากรกลุ่มเชื้อพันธุกรรมนำเข้าและอ้อยพันธุ์ปรับปรุง การกระจายตัวของค่า r^2 ระหว่างคู่เครื่องหมายกับระยะห่างทางกายภาพด้วยสมการแบบไม่เป็นเชิงเส้นตรง (ภาพที่ 4.2) พบว่าในประชากรเชื้อพันธุกรรมอ้อย ค่า r^2 มีค่าลดลงมาจนถึง 0.1 ที่ระยะห่างทางกายภาพประมาณ 1400 bp (ภาพที่ 4.2A) ขณะที่ค่า r^2 ของประชากรพันธุ์ปรับปรุงลดลงถึง 0.1 ที่ระยะห่างทางกายภาพประมาณ 1500 bp (ภาพที่ 4.2B) จะเห็นได้ว่าระยะของความไม่สมดุลลิงเกจของยีน *SaPUB35* ที่มีความเกี่ยวข้องกับลักษณะความหวานในอ้อยพันธุ์ปรับปรุงมีระยะที่กว้างกว่าเชื้อพันธุกรรมนำเข้า เนื่องจากประชากรอ้อยที่เป็นพันธุ์ปรับปรุงนั้นเกิดจากควบคุมการผสมพันธุ์ระหว่างอ้อยพันธุ์ดีที่ผ่านการคัดเลือกแล้วและคัดเลือกลูกผสมเพื่อให้ได้อ้อยที่มีลักษณะตรงตามความต้องการ การคัดเลือกนี้จึงเป็นการลดความแปรปรวนทางพันธุกรรมและตรึงการกลายที่มีผลดี (beneficial mutation) ต่อลักษณะที่ต้องการไว้ (Kim & Nielsen, 2004) ผลของการคัดเลือกทำให้เกิด selective sweep เป็นการคัดเลือกแบบกวาดเอาทั้ง beneficial mutation และแอลลีลที่ขนาบข้างตำแหน่ง beneficial mutation เกิดเป็นบล็อกของ LD (LD block) (Wang et al., 2013) ดังนั้นบล็อกนี้จึงถูกคัดเลือกเก็บไว้ในทุก ๆ รุ่นโดยนักปรับปรุงพันธุ์ ทำให้มีพันธุกรรมเป็นแบบคอขวด (genetic bottleneck) ส่งผลต่อการลดลงของความหลากหลายทางพันธุกรรม (ดังกล่าวข้างต้น) และยังมีผลต่อการเพิ่มระยะ LD ระหว่างตำแหน่ง beneficial mutation (Nimmakayala et al., 2014; Qian et al., 2014) จึงส่งผลให้ระยะของ LD ของยีน *SaPUB35* ในอ้อยพันธุ์ปรับปรุงมีระยะที่กว้างกว่าเชื้อพันธุกรรมนำเข้า เช่นเดียวกับการศึกษาในข้าวบาร์เลย์ระหว่างประชากรพันธุ์ป่าและพันธุ์ปลูกพบว่าประชากรที่เป็นพันธุ์ป่านั้นมี LD แคบกว่าพันธุ์ปลูก เนื่องจากรูปแบบของ LD ในประชากรที่เป็นพันธุ์ป่านั้นขึ้นอยู่กับประวัติการเกิดรีคอมบิเนชันที่ยาวนานกว่าพันธุ์ปลูก (Caldwell, Russell, Langridge, & Powell, 2006) จากการศึกษาในข้าวโพดพบว่าพันธุ์ที่เป็นเชื้อพันธุกรรมซึ่งมีความหลากหลายทางพันธุกรรมมีระยะ LD ที่แคบกว่าข้าวโพดสายพันธุ์ปรับปรุงเพื่อการค้า (Inghelandt, Reif, Dhillon, Flament, & Melchinger, 2011)



ภาพที่ 4.2 แผนภาพการกระจายตัวของค่า r^2 ระหว่างคู่เครื่องหมายด้วยสมการแบบไม่เป็นเชิงเส้นตรง ระหว่างแกน X ซึ่งเป็นระยะห่างทางกายภาพ (bp) และแกน Y เป็นค่า r^2 เส้นโค้งสีแดงแสดง nonlinear logarithmic regression ของค่า r^2 บนระยะห่างทางกายภาพโดยกำหนดค่า r^2 ที่ 0.1 เป็นค่าขั้นต่ำของการมีนัยสำคัญของการเกิด LD ระหว่างคู่เครื่องหมาย (A) ประชากรตัวอย่างที่เป็นเชื้อพันธุกรรมนำเข้า (B) ประชากรตัวอย่างที่เป็นพันธุ์ปรับปรุง

4.5 การวิเคราะห์แผนที่ความสัมพันธ์

วิเคราะห์แผนที่ความสัมพันธ์ระหว่างลักษณะที่เกี่ยวข้องกับลักษณะน้ำตาลทั้ง 4 ลักษณะได้แก่ Brix, CCS, Fiber และ POL ของอ้อยปลูกและอ้อยต่อจากแปลงการทดลองมหาวิทยาลัยขอนแก่น (KKU) และแปลงการทดลองศูนย์นวัตกรรมและการวิจัยมิตรผล (MPT) กับเครื่องหมาย SNP และ InDel ที่ผ่านการคัดกรองด้วยค่า missing ที่ระดับน้อยกว่า 20% MAF ที่ระดับ 0.05 และ LD ที่ระดับนัยสำคัญทางสถิติ Bonferroni correction ที่ $0.05/n$ ($n = 15400$) หรือ $p < 3.24 \times 10^{-6}$ จำนวน 36 เครื่องหมาย เป็นเครื่องหมาย SNP และ InDel อย่างละ 18 เครื่องหมาย โดยใช้โมเดลทดสอบจำนวน 4 แบบคือโมเดล GLM, GLM+Q, MLM+K+Q, และ FarmCPU เพื่อระบุโมเดลทดสอบที่เหมาะสมในการวิเคราะห์ความสัมพันธ์ระหว่างเครื่องหมาย SNP

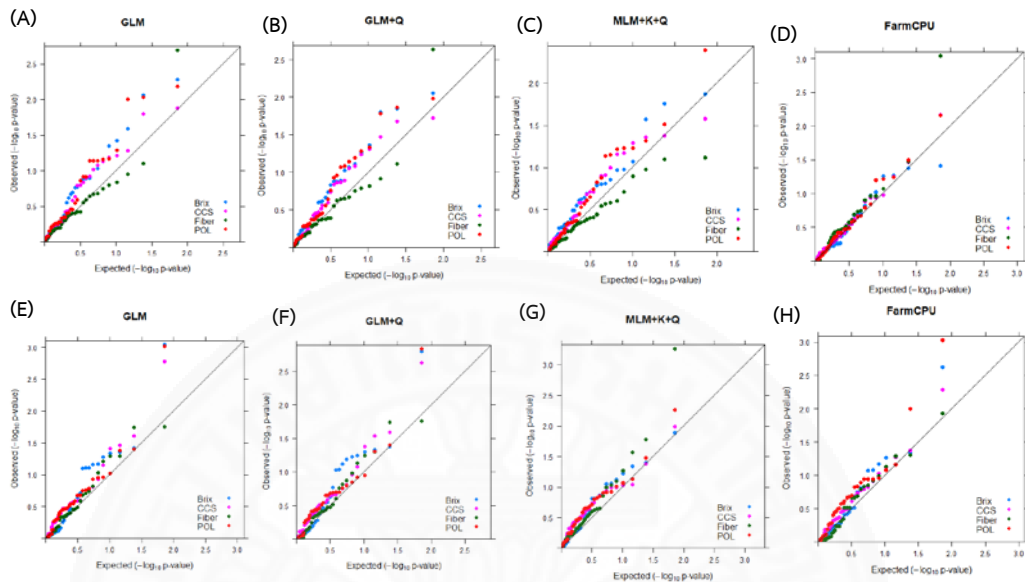
และ InDel กับลักษณะที่เกี่ยวข้องกับน้ำตาลทั้ง 4 ลักษณะโดยพิจารณาการเบี่ยงของ quantile–quantile (Q-Q) probability plot พบว่าโมเดล GLM ที่ไม่นำโครงสร้างประชากรและความเป็นเครือญาติมาเป็นตัวแปรร่วมในการวิเคราะห์หาความสัมพันธ์ Q-Q plot มีการเบี่ยงออกจากเส้นตรง 1:1 มากที่สุด (ภาพที่ 4.3(A,E), ภาพที่ 4.4(A,E), ภาพที่ 4.5(A,E)) แสดงถึงการเกิดผลบวกวง (false positive) หรือ type I error จึงนำโครงสร้างประชากรมาวิเคราะห์ร่วมด้วยโมเดล GLM+Q พบว่า Q-Q plot มีการเบี่ยงออกจากเส้นตรง 1:1 น้อยลง (ภาพที่ 4.3(B,F), ภาพที่ 4.4(B,F), ภาพที่ 4.5(B,F)) และเมื่อนำทั้งความเป็นเครือญาติและโครงสร้างประชากรมาเป็นตัวแปรร่วมในการวิเคราะห์ด้วยโมเดล MLM+K+Q และ FarmCPU พบว่าโมเดล MLM ทั้งสองนี้มีการเบี่ยงเบนออกจากเส้นตรงน้อยที่สุด แต่เมื่อเปรียบเทียบระหว่างโมเดล MLM ระหว่างโมเดล MLM+K+Q และ FarmCPU พบว่าโมเดล FarmCPU เข้าใกล้เส้นตรง 1:1 มากที่สุด แสดงถึงประสิทธิภาพของโมเดลในการควบคุม false positive และ false negative ได้ดีที่สุดในที่สุด ซึ่งให้ผลเช่นเดียวกับงานวิจัยก่อนหน้า (Bollinedi et al., 2020; Jaiswal et al., 2019; Kaler, Gillman, Beissinger, & Purcell, 2020) ดังนั้นในงานวิจัยนี้จึงเลือกใช้โมเดล FarmCPU ในการวิเคราะห์แผนที่ความสัมพันธ์ระหว่างเครื่องหมาย SNP และ InDel กับลักษณะที่เกี่ยวข้องกับลักษณะน้ำตาล

ผลการวิเคราะห์ความสัมพันธ์ระหว่างเครื่องหมาย SNP และ InDel กับลักษณะที่เกี่ยวข้องกับน้ำตาลทั้ง 4 ลักษณะในอ้อยปลูก (ตารางที่ 4.6) โดยใช้เกณฑ์โมเดล FarmCPU ที่เกณฑ์ นัยสำคัญที่ p -value 0.01 และเกณฑ์ FDR ที่ $q < 0.10$ พบว่าแปลงการทดลองมหาวิทยาลัยขอนแก่นสามารถระบุความสัมพันธ์ระหว่างเครื่องหมายกับลักษณะที่เกี่ยวข้องกับน้ำตาลได้ 2 เครื่องหมาย คือ SNP61 ซึ่งมีแนวโน้มมีความสัมพันธ์ที่ p -value 0.01 กับ POL และสามารถอธิบายความแปรปรวนของ POL ได้ 4.88% และอีกหนึ่งเครื่องหมายคือ SNP89 มีความสัมพันธ์แบบมีนัยสำคัญที่ FDR ที่ $q < 0.10$ กับ Fiber และสามารถอธิบายความแปรปรวนของฟิโนไทป์ได้ 7.09% โดยกลุ่มตัวอย่างที่พบเครื่องหมายและไม่พบเครื่องหมาย SNP61 มีค่าเฉลี่ยของ CCS เท่ากับ 14.82 และ 16.27 ตามลำดับ (ภาพที่ 4.6A) ซึ่งมีความแตกต่างกันอย่างมีนัยสำคัญ (p -value < 0.01) และค่าเฉลี่ยของ Fiber ของกลุ่มที่พบ SNP89 มีค่าสูงกว่ากลุ่มที่ไม่พบ SNP89 เท่ากับ 1.31 อย่างมีนัยสำคัญ (p -value < 0.01) (ภาพที่ 4.6B) ผลจากแปลงการทดลองศูนย์นวัตกรรมและการวิจัยมิตรผล พบว่าสามารถระบุเครื่องหมายที่มีความสัมพันธ์ได้ 3 เครื่องหมายเป็นเครื่องหมายที่มีความสัมพันธ์แบบมีนัยสำคัญที่ FDR ที่ $q < 0.10$ จำนวน 2 เครื่องหมายคือเครื่องหมาย SNP61 มีความสัมพันธ์แบบมีนัยสำคัญกับ CCS ซึ่งอธิบายความแปรปรวนของ CCS ได้ 8.46% ซึ่งกลุ่มตัวอย่างที่พบเครื่องหมาย SNP61 มีค่าเฉลี่ยของ CCS เท่ากับ 9.61 และกลุ่มตัวอย่างไม่พบเครื่องหมาย SNP61 มีค่าเฉลี่ยของ CCS เท่ากับ 11.38 (ภาพที่ 4.6D) ซึ่งมีความแตกต่างกันอย่างมีนัยสำคัญ

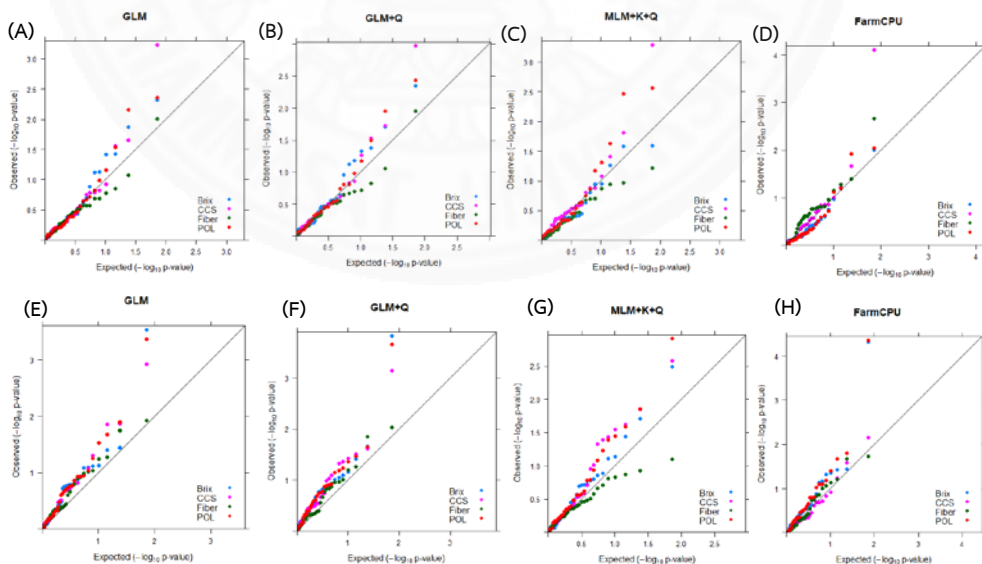
(p -value < 0.01) และเครื่องหมาย SNP89 มีความสัมพันธ์กับ Fiber แบบมีนัยสำคัญ สามารถอธิบายความแปรปรวนของ Fiber ได้ 4.96% โดยกลุ่มที่พบเครื่องหมาย SNP89 มีค่าเฉลี่ยของ Fiber สูงกว่ากลุ่มที่ไม่พบเครื่องหมายอย่างมีนัยสำคัญ (p -value < 0.01) (ภาพที่ 4.6E) และอีกหนึ่งเครื่องหมายคือ ID43 ซึ่งมีแนวโน้มมีความสัมพันธ์ที่ p -value 0.01 กับ POL สามารถอธิบายความแปรปรวนของ POL ได้ 5.74% โดยกลุ่มตัวอย่างที่พบเครื่องหมายและไม่พบเครื่องหมาย ID43 มีค่าเฉลี่ยของ POL เท่ากับ 13.76 และ 15.87 ตามลำดับ ซึ่งมีความแตกต่างกันอย่างมีนัยสำคัญ (p -value < 0.01) (ภาพที่ 4.6C), และเมื่อวิเคราะห์ความสัมพันธ์แบบรวมแปลงการทดลองสามารถระบุเครื่องหมายที่มีความสัมพันธ์แบบมีนัยสำคัญที่ FDR ที่ $q < 0.10$ ได้จำนวน 2 เครื่องหมายคือ SNP61 มีความสัมพันธ์แบบมีนัยสำคัญกับ CCS และ SNP 89 มีความสัมพันธ์แบบมีนัยสำคัญกับ Fiber โดยเครื่องหมาย SNP61 และ SNP89 สามารถอธิบายความแปรปรวนของ CCS และ Fiber ได้ 8.19 และ 6.70% ตามลำดับ โดยกลุ่มตัวอย่างที่พบเครื่องหมาย SNP61 มีค่าเฉลี่ยต่ำกว่ากลุ่มที่ไม่พบเครื่องหมาย SNP61 อย่างมีนัยสำคัญ (p -value < 0.01) (ภาพที่ 4.6F) ขณะที่กลุ่มตัวอย่างที่พบเครื่องหมาย SNP89 มีค่าเฉลี่ยสูงกว่ากลุ่มที่ไม่พบเครื่องหมาย SNP89 อย่างมีนัยสำคัญ (p -value < 0.01) (ภาพที่ 4.6G)

ผลการวิเคราะห์ความสัมพันธ์ระหว่างเครื่องหมาย SNP และ InDel กับลักษณะที่เกี่ยวข้องกับน้ำตาลทั้ง 4 ลักษณะในอ้อยตอ (ตารางที่ 4.6) เมื่อพิจารณาผลการวิเคราะห์แผนที่ความสัมพันธ์โดยใช้เกณฑ์โมเดล FarmCPU เกณฑ์นัยสำคัญที่ p -value 0.01 และเกณฑ์ FDR ที่ $q < 0.10$ พบว่าแปลงการทดลองมหาวิทยาลัยขอนแก่น สามารถระบุเครื่องหมายที่มีความสัมพันธ์กับลักษณะที่เกี่ยวข้องกับน้ำตาลได้เพียงเครื่องหมายเดียวคือ ID83 ซึ่งมีความสัมพันธ์แบบมีนัยสำคัญที่ FDR ที่ $q < 0.10$ กับลักษณะ Brix และ POL และมีแนวโน้มมีความสัมพันธ์ที่ p -value 0.01 กับ CCS โดยสามารถอธิบายความแปรปรวนของ Brix, CCS, และ POL ได้เท่ากับ 8.85, 8.01, และ 8.86% ตามลำดับ โดยพบว่ากลุ่มตัวอย่างที่พบเครื่องหมาย ID83 มีค่าเฉลี่ยของ Brix, CCS, และ POL สูงกว่ากลุ่มตัวอย่างที่ไม่พบเครื่องหมาย ID83 อย่างมีนัยสำคัญ (p -value < 0.01) (ภาพที่ 4.7A-C) และในแปลงการทดลองศูนย์นวัตกรรมและการวิจัยมิตรผลที่ระบุเครื่องหมายที่มีความสัมพันธ์ได้เพียงเครื่องหมายเดียวคือ ID43 โดยมีความสัมพันธ์แบบมีนัยสำคัญที่ FDR ที่ $q < 0.10$ กับลักษณะ Brix และ POL และมีแนวโน้มมีความสัมพันธ์ที่ p -value 0.01 กับ CCS และสามารถอธิบายความแปรปรวนของ Brix, CCS, และ POL ได้เท่ากับ 9.10, 7.34, และ 8.63% ตามลำดับ โดยพบว่ากลุ่มตัวอย่างที่พบเครื่องหมาย ID43 มีค่าเฉลี่ยของ Brix, CCS, และ POL ต่ำกว่ากลุ่มตัวอย่างที่ไม่พบเครื่องหมาย ID43 อย่างมีนัยสำคัญ (p -value < 0.01) (ภาพที่ 4.7D-F) และเมื่อวิเคราะห์แบบรวมแปลงการทดลองทั้งสองโดยใช้เกณฑ์โมเดล FarmCPU เกณฑ์นัยสำคัญที่

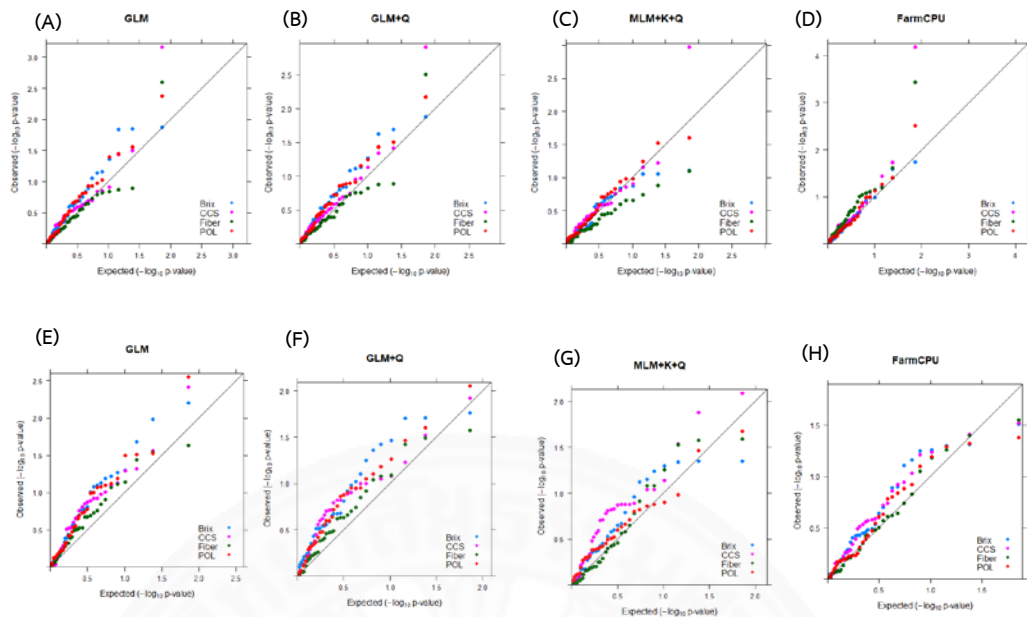
p -value 0.01 และเกณฑ์ FDR ที่ $q < 0.10$ พบว่าไม่สามารถระบุความสัมพันธ์ระหว่างเครื่องหมายกับลักษณะที่เกี่ยวข้องกับน้ำตาลทั้ง 4 ลักษณะได้



ภาพที่ 4.3 การกระจายตัวของค่า p -value ที่ได้จากโมเดลทดสอบ 4 แบบที่ใช้วิเคราะห์ความสัมพันธ์ระหว่างเครื่องหมาย SNP และ InDel กับลักษณะที่เกี่ยวข้องกับน้ำตาลทั้ง 4 ลักษณะของอ้อยปลูก (A-D) และอ้อยโต (E-H) แปลงการทดลองมหาวิทยาลัยขอนแก่น



ภาพที่ 4.4 การกระจายตัวของค่า p -value ที่ได้จากโมเดลทดสอบ 4 แบบที่ใช้วิเคราะห์ความสัมพันธ์ระหว่างเครื่องหมาย SNP และ InDel กับลักษณะที่เกี่ยวข้องกับน้ำตาลทั้ง 4 ลักษณะของอ้อยปลูก (A-D) และอ้อยโต (E-H) แปลงการทดลองศูนย์นวัตกรรมและการวิจัยมิตรผล



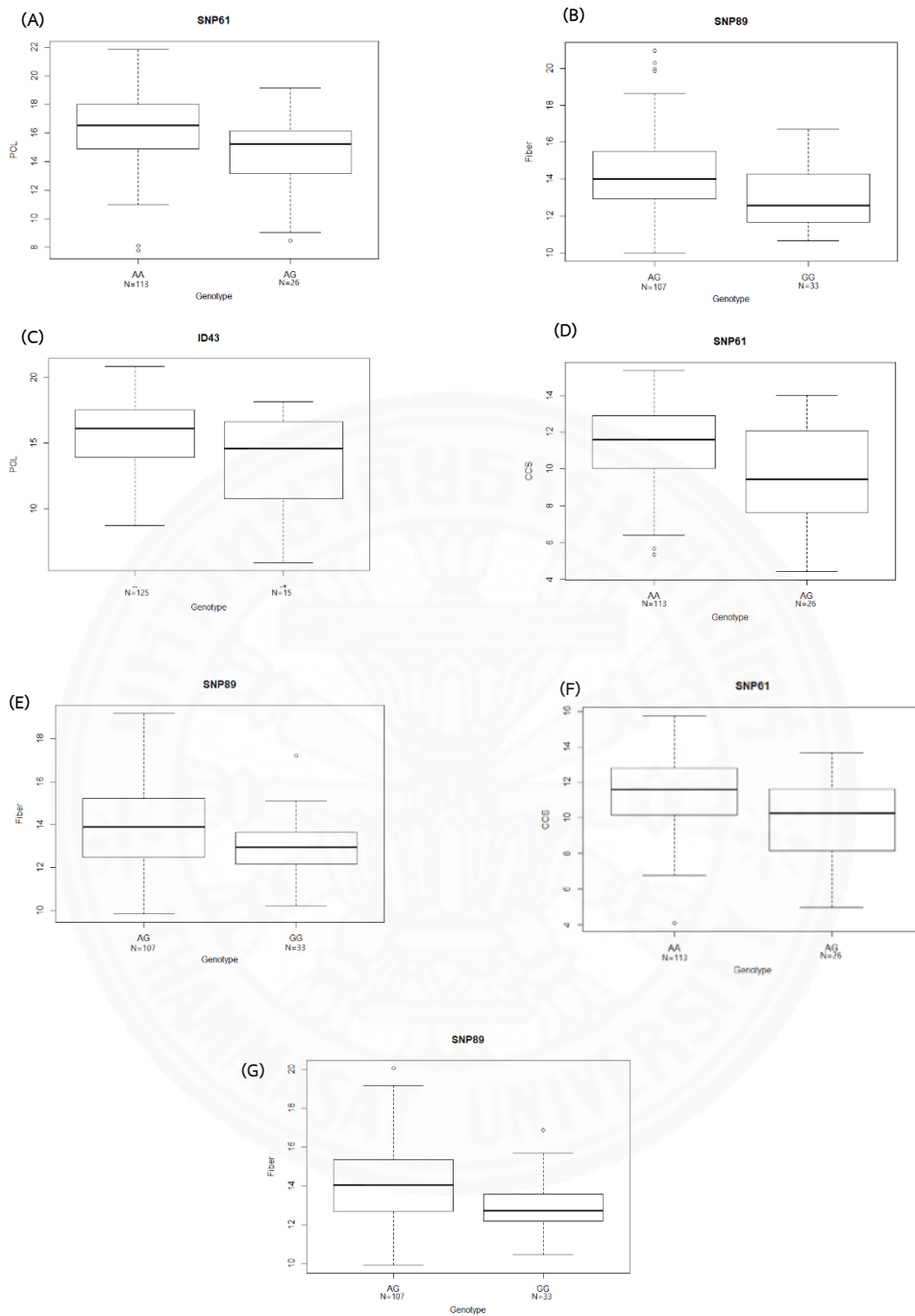
ภาพที่ 4.5 การกระจายตัวของค่า p-value ที่ได้จากโมเดลทดสอบ 4 แบบที่ใช้วิเคราะห์ความสัมพันธ์ระหว่างเครื่องหมาย SNP และ InDel กับลักษณะที่เกี่ยวข้องกับน้ำตาลทั้ง 4 ลักษณะของอ้อยปลูก (A-D) และอ้อยต่อ (E-H)

จะเห็นได้ว่าทั้งเครื่องหมาย SNP และ InDel ที่ถูกระบุว่ามีความสัมพันธ์กับลักษณะที่เกี่ยวข้องกับน้ำตาลอย่างมีนัยสำคัญนั้นมีตำแหน่งอยู่บนอินตรอนของยีนดังแสดงในตารางที่ 4.6 ซึ่งเป็นส่วนของลำดับเบสที่ไม่ได้กำหนดรหัสพันธุกรรม (non-coding) ถึงแม้ว่าเครื่องหมาย SNP บน non-coding ไม่สามารถเปลี่ยนกรดอะมิโนอย่างเครื่องหมาย SNP ที่มีตำแหน่งบนบริเวณลำดับเบสที่กำหนดรหัสพันธุกรรมหรือเอกซอน แต่อาจส่งผลต่อตำแหน่งของการจับของ transcription factor (transcription factor binding sites), ตำแหน่งของการตัดอินตรอน (splice sites) เพื่อเชื่อมต่อเอกซอน และตำแหน่งที่มีหน้าที่อื่น ๆ ในระดับการถอดรหัส (transcriptional level) (Xu et al., 2014) และการมี SNP บนบริเวณ intron นั้นทำให้เกิดความแปรปรวนของลำดับเบสซึ่งอาจจะมีผลต่อการแสดงออกของยีนหรือมีผลต่อลำดับเบสอื่น ๆ ที่มีลิงก์กระชับใกล้เคียง (tightly link) ซึ่งเครื่องหมายที่วิเคราะห์ว่ามีความสัมพันธ์กับลักษณะที่เกี่ยวข้องกับน้ำตาลภายในยีน *SaPUB35* อาจจะมี LD กับแอลลีลบริเวณที่ควบคุมการแสดงออกหรือบริเวณที่มีความสำคัญต่อการแสดงออกของยีนที่เกี่ยวข้องกับลักษณะน้ำตาล ส่งผลให้เกิดความแปรปรวนของลักษณะที่เกี่ยวข้องกับลักษณะน้ำตาลในอ้อย (Singh, Singh, Mishra, Singh, & Singh, 2019)

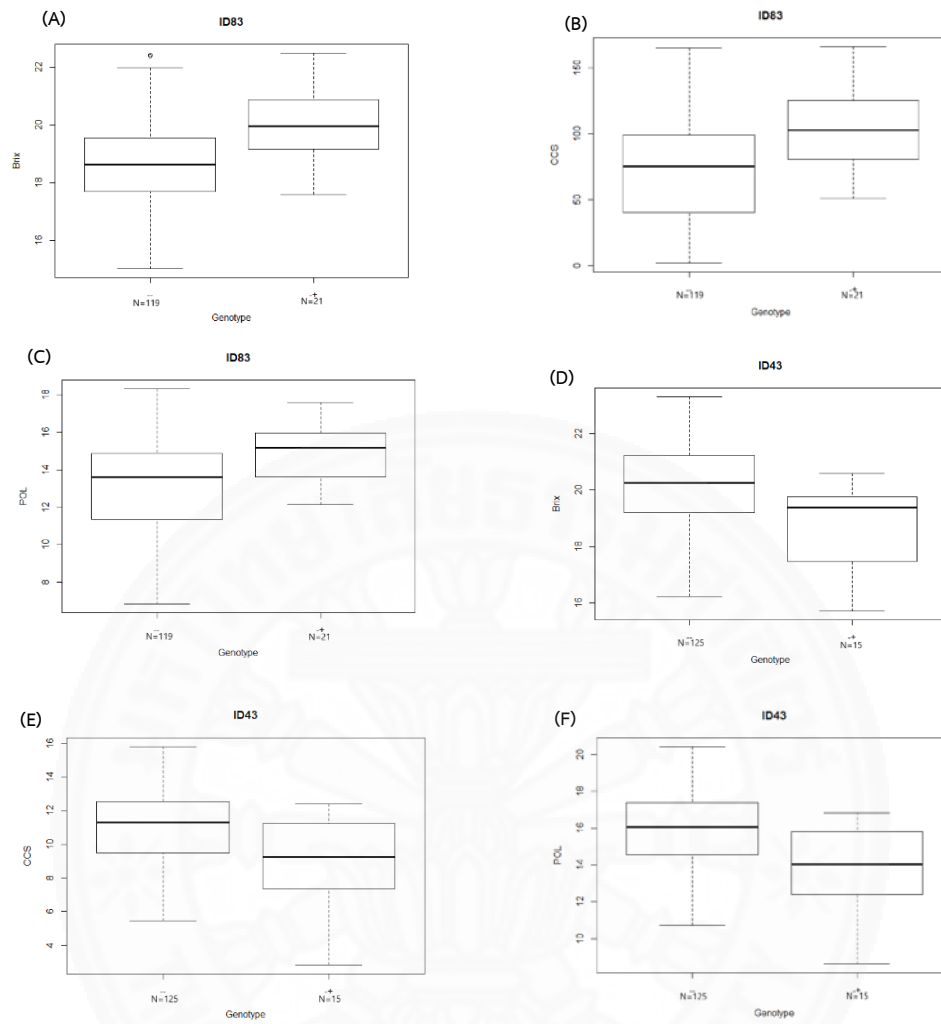
ตารางที่ 4.6 เครื่องหมาย SNP และ InDel ที่มีความสัมพันธ์กับลักษณะที่เกี่ยวข้องกับลักษณะน้ำตาลของอ้อยปลูก (plant) และอ้อยตอ (1st ratoon) แปลงการทดลองมหาวิทยาลัยขอนแก่น (KKU) แปลงการทดลองศูนย์นวัตกรรมและการวิจัยมิตรผล (MPT) และรวมแปลงการทดลอง (combined)

Crop cycle	Location	Traits	Marker	Allele	Position on gene region	GLM		GLM+Q		MLM+Q+K		FarmCPU		
						p-value	r ² (%)	p-value	r ² (%)	p-value	r ² (%)	p-value	r ² (%)	
Plant cane	KKU	POL	SNP61	A/G	Intron 7	<u>9.24x10⁻³</u>	4.88	1.39x10 ⁻²	4.3	3.08x10 ⁻²	3.49	6.98x10 ⁻³	4.88	
		Fiber	SNP89	A/G	Intron 5	<u>2.04x10⁻³</u>	7.09	<u>2.32x10⁻³</u>	6.96	7.92x10 ⁻²	2.52	<u>9.10x10⁻⁴</u>	7.09	
	MPT	POL	ID43	G/GA	Intron 8	<u>4.37x10⁻³</u>	5.74	3.75x10 ⁻³	5.7	<u>3.39x10⁻³</u>	6.21	9.15x10 ⁻³	5.74	
		CCS	SNP61	A/G	Intron 7	<u>5.93x10⁻⁴</u>	8.46	<u>1.06x10⁻³</u>	7.47	<u>5.11x10⁻⁴</u>	9.32	<u>7.86x10⁻⁵</u>	8.46	
	Combined	Fiber	SNP89	A/G	Intron 5	9.72x10 ⁻³	4.96	1.13x10 ⁻²	4.79	2.05x10 ⁻¹	1.26	<u>2.17x10⁻³</u>	4.96	
		CCS	SNP61	A/G	Intron 7	<u>6.99x10⁻⁴</u>	8.19	<u>1.22x10⁻³</u>	7.21	<u>1.07x10⁻³</u>	8.3	<u>6.66x10⁻⁵</u>	8.19	
	1 st Ratoon	KKU	Fiber	SNP89	A/G	Intron 5	<u>2.54x10⁻³</u>	6.7	<u>3.09x10⁻³</u>	6.47	2.17x10 ⁻¹	1.22	<u>3.63x10⁻⁴</u>	6.7
			Brix	ID83	T/TG	Intron 6	<u>9.18x10⁻⁴</u>	8.85	<u>1.63x10⁻³</u>	7.96	1.30x10 ⁻²	5.23	<u>2.36x10⁻³</u>	8.85
CCS			ID83	T/TG	Intron 6	<u>1.66x10⁻³</u>	8.01	<u>2.32x10⁻³</u>	7.55	1.03x10 ⁻²	5.82	5.14x10 ⁻³	8.01	
MPT		POL	ID83	T/TG	Intron 6	<u>9.63x10⁻⁴</u>	8.86	<u>1.48x10⁻³</u>	8.22	5.53x10 ⁻³	6.77	<u>9.30x10⁻⁴</u>	8.86	
		Brix	ID43	G/GA	Intron 8	<u>2.93x10⁻⁴</u>	9.1	<u>1.50x10⁻⁴</u>	9.03	3.21x10 ⁻³	6.22	<u>4.74x10⁻⁵</u>	9.1	
		CCS	ID43	G/GA	Intron 8	<u>1.21x10⁻³</u>	7.34	<u>7.20x10⁻⁴</u>	7.28	2.65x10 ⁻³	6.6	6.89x10 ⁻³	7.34	
Combined	POL	ID43	G/GA	Intron 8	<u>4.28x10⁻⁴</u>	8.63	<u>2.22x10⁻⁴</u>	8.56	1.22x10 ⁻³	7.61	<u>4.45x10⁻⁵</u>	8.63		

Underline type indicates marker-trait associations having a false discovery rate (FDR) at $q < 0.10$.



ภาพที่ 4.6 แผนภาพ box plot แสดงอิทธิพลของเครื่องหมาย SNP และ InDel ที่สัมพันธ์กับลักษณะที่เกี่ยวข้องกับลักษณะน้ำตาลในอ้อยปลูก



ภาพที่ 4.7 แผนภาพ box plot แสดงอิทธิพลของเครื่องหมาย SNP และ InDel ที่สัมพันธ์กับลักษณะที่เกี่ยวข้องกับลักษณะน้ำตาลในอ้อยต่อ

บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

งานวิจัยนี้สามารถระบุเครื่องหมาย SNP และ InDel ภายในยีน *SaPUB35* ในอ้อยโดยใช้เทคนิค PacBio SMRT sequencing ได้จำนวน 126 และ 50 เครื่องหมายหรือโดยเฉลี่ยแล้วสามารถพบ SNP และ InDel ได้ทุก ๆ 46 และ 115 เบส ตามลำดับ ความเสื่อมถอยของความไม่สมดุลลิงเกจของตัวอย่างที่เป็นเชื้อพันธุกรรมนำเข้ากับพันธุ์ปรับปรุงมีระยะที่ 1400 และ 1500 คู่เบส ตามลำดับ โดยระยะการเสื่อมถอยของความไม่สมดุลลิงเกจของอ้อยพันธุ์ปรับปรุงมีระยะที่กว้างกว่าอ้อยที่เป็นเชื้อพันธุกรรมนำเข้าเนื่องจากอิทธิพลของ selective sweep ซึ่งเป็นการคัดเลือกในการกระบวนการปรับปรุงพันธุ์อ้อย นอกจากนี้ยังพบว่า selective sweep มีผลต่อค่าความแปรปรวนของพีโนไทป์ และค่าเฉลี่ยของลักษณะที่เกี่ยวข้องกับน้ำตาลและค่าดัชนีความเข้มของสีใบ โดยในอ้อยพันธุ์ปรับปรุงนี้มีค่าความแปรปรวนของลักษณะที่เกี่ยวข้องกับลักษณะน้ำตาลส่วนใหญ่และค่าดัชนีความเข้มของสีใบมีค่าที่ต่ำกว่าในอ้อยที่เป็นเชื้อพันธุกรรมนำเข้า และค่าเฉลี่ยของลักษณะที่เกี่ยวข้องกับลักษณะน้ำตาลส่วนใหญ่และค่าดัชนีความเข้มของสีใบในอ้อยที่เป็นพันธุ์ปรับปรุงมีค่าที่มากกว่าในอ้อยที่เป็นเชื้อพันธุกรรมนำเข้า ผลการวิเคราะห์แผนความสัมพันธ์ภายในยีน *SaPUB35* พบว่า สามารถระบุเครื่องหมาย SNP และ InDel ได้จำนวน 4 เครื่องหมายที่มีความสัมพันธ์กับลักษณะที่เกี่ยวข้องกับน้ำตาลทั้ง 4 ลักษณะอย่างมีนัยสำคัญ ผลการศึกษาในครั้งนี้อาจจะเป็นประโยชน์ในการนำเครื่องหมายดีเอ็นเอมาช่วยคัดเลือกพันธุ์ในโครงการการปรับปรุงพันธุ์อ้อยต่อไป

รายการอ้างอิง

บทความวารสาร

- กิตติพัฒน์ อุโฆษกิจ, มาริสา รักสุขสมบัติ, และสุภาภรณ์ ขานโบ. (2558). เทคโนโลยีการวิเคราะห์ลำดับเบสยุคที่ สองและสาม. วารสารวิทยาศาสตร์และเทคโนโลยี, 4, 633-650. doi:10.14456/tstj.2015.22
- สำนักพัฒนาอุตสาหกรรมอ้อย น้ำตาลทรายและอุตสาหกรรมต่อเนื่อง. (2560). คนรักอ้อย. กรุงเทพมหานคร. สำนักงานคณะกรรมการอ้อยและน้ำตาลทราย กระทรวงอุตสาหกรรม.
- สำนักพัฒนาอุตสาหกรรมอ้อยและน้ำตาลทราย. (2563). รายงานสถานการณ์การปลูกอ้อยปีการผลิต 2562/63. กรุงเทพมหานคร. สำนักงานคณะกรรมการอ้อยและน้ำตาลทราย.
- Agarwal, M., Shrivastava, N., & Padh, H. (2008). Advances in molecular marker techniques and their applications in plant sciences. *Plant cell reports*, 27(4), 617-631.
- Alqudah, A. M., Haile, J. K., Alomari, D. Z., Pozniak, C. J., Kobiljski, B., & Börner, A. (2020). Genome-wide and SNP network analyses reveal genetic control of spikelet sterility and yield-related traits in wheat. *Scientific reports*, 10(1), 1-12.
- Anagbogu, C. F., Bhattacharjee, R., Ilori, C., Tongyoo, P., Dada, K. E., Muyiwa, A. A., Gepts, P., & Beckles, D. M. (2019). Genetic diversity and re-classification of coffee (*Coffea canephora* Pierre ex A. Froehner) from South Western Nigeria through genotyping-by-sequencing-single nucleotide polymorphism analysis. *Genetic Resources and Crop Evolution*, 1-12.
- Atwell, S., Huang, Y. S., Vilhjálmsson, B. J., Willems, G., Horton, M., Li, Y., Meng, D., Platt, A., Tarone, A. M., & Hu, T. T. (2010). Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature*, 465(7298), 627.
- Bae, H., & Kim, W. T. (2014). Classification and interaction modes of 40 rice E2 ubiquitin-conjugating enzymes with 17 rice ARM-U-box E3 ubiquitin ligases. *Biochemical biophysical research communications*, 444(4), 575-580.
- Balsalobre, T. W. A., da Silva Pereira, G., Margarido, G. R. A., Gazaffi, R., Barreto, F. Z., Anoni, C. O., Cardoso-Silva, C. B., Costa, E. A., Mancini, M. C., & Hoffmann, H. P.

- (2017). GBS-based single dosage markers for linkage and QTL mapping allow gene mining for yield-related traits in sugarcane. *BMC genomics*, 18(1), 72.
- Banerjee, N., Khan, M. S., Swapna, M., Singh, R., & Kumar, S. (2020). Progress and Prospects of Association Mapping in Sugarcane (*Saccharum* Species Hybrid), a Complex Polyploid Crop. *Sugar Tech*, 1-15.
- Banerjee, N., Siraree, A., Yadav, S., Kumar, S., Singh, J., Pandey, D. K., & Singh, R. K. (2015). Marker-trait association study for sucrose and yield contributing traits in sugarcane (*Saccharum* spp. hybrid). *Euphytica*, 205(1), 185-201.
- Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N., & Golani, I. (2001). Controlling the false discovery rate in behavior genetics research. *Behavioural brain research*, 125(1-2), 279-284.
- Bhatramakki, D., Dolan, M., Hanafey, M., Wineland, R., Vaske, D., Register, J. C., Tingey, S. V., & Rafalski, A. (2002). Insertion-deletion polymorphisms in 3' regions of maize genes occur frequently and can be used as highly informative genetic markers. *Plant molecular biology*, 48(5-6), 539-547.
- Bollinedi, H., Yadav, A. K., Vinod, K., Krishnan, S. G., Bhowmick, P. K., Nagarajan, M., Neeraja, C., Ellur, R. K., & Singh, A. K. (2020). Genome-Wide Association Study Reveals Novel Marker-Trait Associations (MTAs) Governing the Localization of Fe and Zn in the Rice Grain. *Frontiers in genetics*, 11.
- Breslauer, K. J., Frank, R., Blöcker, H., & Marky, L. A. (1986). Predicting DNA duplex stability from the base sequence. *Proceedings of the National Academy of Sciences*, 83(11), 3746-3750.
- Byun, M. Y., Cui, L. H., Oh, T. K., Jung, Y. J., Lee, A., Park, K. Y., Kang, B. G., & Kim, W. T. (2017). Homologous U-box E3 ubiquitin ligases OsPUB2 and OsPUB3 are involved in the positive regulation of low temperature stress response in rice (*Oryza sativa* L.). *Frontiers in plant science*, 8(16), 1-19.
- Caldwell, K. S., Russell, J., Langridge, P., & Powell, W. (2006). Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*. *Genetics*, 172(1), 557-567.

- Cheavegatti-Gianotto, A., de Abreu, H. M., Arruda, P., Bespalhok Filho, J. C., Burnquist, W. L., Creste, S., di Ciero, L., Ferro, J. A., de Oliveira Figueira, A. V., de Sousa Filgueiras, T., et al. (2011). Sugarcane (*Saccharum X officinarum*): A Reference Study for the Regulation of Genetically Modified Cultivars in Brazil. *Tropical Plant Biology*, *4*(1), 62-89. doi:10.1007/s12042-011-9068-3
- Chen, L., Deng, R., Liu, G., Jin, J., Wu, J., & Liu, X. (2019). Cytological and transcriptome analyses reveal OsPUB73 defect affects the gene expression associated with tapetum or pollen exine abnormality in rice. *BMC plant biology*, *19*(1), 546.
- Choi, I.-Y., Hyten, D. L., Matukumalli, L. K., Song, Q., Chaky, J. M., Quigley, C. V., Chase, K., Lark, K. G., Reiter, R. S., & Yoon, M.-S. (2007). A soybean transcript map: gene distribution, haplotype and single-nucleotide polymorphism analysis. *Genetics*, *176*(1), 685-696.
- Clevenger, J., Chavarro, C., Pearl, S. A., Ozias-Akins, P., & Jackson, S. A. (2015). Single nucleotide polymorphism identification in polyploids: a review, example, and recommendations. *Molecular plant*, *8*(6), 831-846.
- Cordeiro, G. M., Elliott, F., McIntyre, C. L., Casu, R. E., & Henry, R. J. (2006). Characterisation of single nucleotide polymorphisms in sugarcane ESTs. *Theoretical and Applied Genetics*, *113*(2), 331-343.
- D'Hont, A. (2005). Unraveling the genome structure of polyploids using FISH and GISH; examples of sugarcane and banana. *Cytogenetic and Genome Research*, *109*(1-3), 27-33. doi:10.1159/000082378
- D'Hont, A., & Glaszmann, J. C. (2001). Sugarcane Genome Analysis with Molecular Markers - A First Decade of Research. *Proc. Int. Soc. Sugarcane Technol*, *24*.
- D'Hont et al. (1996). Characterisation of the double genome structure of modern sugarcane (*Saccharum* spp.) by molecular cytogenetics. *Molecular Genetics and Genomics*, *250*, 405-413.
- D'Hont. (2005). Unraveling the genome structure of polyploids using FISH and GISH; examples of sugarcane and banana. *Cytogenetic and Genome Research*, *109*(1-3), 27-33.

- Ehrenreich, I. M., Hanzawa, Y., Chou, L., Roe, J. L., Kover, P. X., & Purugganan, M. D. (2009). Candidate gene association mapping of *Arabidopsis* flowering time. *Genetics*, *183*(1), 325-335.
- Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., & Bettman, B. (2009). Real-time DNA sequencing from single polymerase molecules. *Science*, *323*(5910), 133-138.
- Fickett, N., Gutierrez, A., Verma, M., Pontif, M., Hale, A., Kimbeng, C., & Baisakh, N. (2019). Genome-wide association mapping identifies markers associated with cane yield components and sucrose traits in the Louisiana sugarcane core collection. *Genomics*, *111*(6), 1794-1801.
- Flint-Garcia, S. A., Thornsberry, J. M., & Buckler IV, E. S. (2003). Structure of linkage disequilibrium in plants. *Annual review of plant biology*, *54*(1), 357-374.
- Garsmeur, O., Droc, G., Antonise, R., Grimwood, J., Potier, B., Aitken, K., Jenkins, J., Martin, G., Charron, C., Catherine, H., et al. (2018). A mosaic monoploid reference sequence for the highly complex genome of sugarcane. *Nature communications*, *9*(1), 1-10.
- Gawel, N., & Jarret, R. (1991). A modified CTAB DNA extraction procedure for *Musa* and *Ipomoea*. *Plant molecular biology reporter*, *9*(3), 262-266.
- Gore, M. A., Wright, M. H., Ersoz, E. S., Bouffard, P., Szekeres, E. S., Jarvie, T. P., Hurwitz, B. L., Narechania, A., Harkins, T. T., & Grills, G. S. (2009). Large-scale discovery of gene-enriched SNPs. *The plant genome*, *2*(2), 121-133.
- Gouy, M., Rousselle, Y., Chane, A. T., Anglade, A., Royaert, S., Nibouche, S., & Costet, L. (2015). Genome wide association mapping of agro-morphological and disease resistance traits in sugarcane. *Euphytica*, *202*(2), 269-284.
- Hoang, N. V., Furtado, A., J Mason, P., Marquardt, A., Kasirajan, L., Thirugnanasambandam, P. P., Botha, F. C., & Henry, R. J. (2017). A survey of the complex transcriptome from the highly polyploid sugarcane genome using full-length isoform sequencing and de novo assembly from short read sequencing. *BMC genomics*, *18*(395), 1-22.

- Holland, J. B., Nyquist, W. E., & Cervantes-Martínez, C. T. (2003). Estimating and interpreting heritability for plant breeding: an update. *Plant breeding reviews*, 22.
- Huang , D. L., Gao, Y. J., Gui, Y. Y., Chen, Z. L., Qin, C. X., Wang, M., Liao, Q., Yang, L. T., & Li, Y. R. (2016). Transcriptome of high-sucrose sugarcane variety GT35. *Sugar Tech*, 18(5), 520-528.
- Huang, X., Feng, Q., Qian, Q., Zhao, Q., Wang, L., Wang, A., Guan, J., Fan, D., Weng, Q., & Huang, T. (2009). High-throughput genotyping by whole-genome resequencing. *Genome research*, 19(6), 1068-1076.
- Inghelandt, D. V., Reif, J. C., Dhillon, B. S., Flament, P., & Melchinger, A. E. (2011). Extent and genome-wide distribution of linkage disequilibrium in commercial maize germplasm. *Theoretical and Applied Genetics*, 123(1), 11-20.
- Jaikishan, I., Rajendrakumar, P., Madhusudhana, R., Elangovan, M., & Patil, J. V. (2015). Development and utility of PCR-based intron polymorphism markers in sorghum [*Sorghum bicolor* (L.) Moench]. *Journal of crop science and biotechnology*, 18(5), 309-318.
- Jaiswal, V., Bandyopadhyay, T., Gahlaut, V., Gupta, S., Dhaka, A., Ramchiary, N., & Prasad, M. (2019). Genome-wide association study (GWAS) delineates genomic loci for ten nutritional elements in foxtail millet (*Setaria italica* L.). *Journal of cereal science*, 85, 48-55.
- Jannoo , N., Grivet , L., Chantret , N., Garsmeur , O., Glaszmann , J. C., Arruda , P., & D'Hont , A. (2007). Orthologous comparison in a gene-rich region among grasses reveals stability in the sugarcane polyploid genome. *The Plant Journal*, 50(4), 574-585.
- Jannoo, N., Grivet, L., Dookun, A., D'Hont, A., & Glaszmann, J.-C. (1999). Linkage disequilibrium among modern sugarcane cultivars. *Theoretical Applied Genetics*, 99(6), 1053-1060.
- Jehan, T., & Lakhanpaul, S. (2006). Single nucleotide polymorphism (SNP)-methods and applications in plant genetics: a review.

- Kaler, A. S., Gillman, J. D., Beissinger, T., & Purcell, L. C. (2020). Comparing different statistical models and multiple testing corrections for association mapping in soybean and maize. *Frontiers in plant science*, *10*, 1794.
- Kasirajan, L., Hoang, N. V., Furtado, A., Botha, F. C., & Henry, R. J. (2018). Transcriptome analysis highlights key differentially expressed genes involved in cellulose and lignin biosynthesis of sugarcane genotypes varying in fiber content. *Scientific reports*, *8*(1), 1-16.
- Kim, J., Lee, S. Y., Kim, D., Lee, E.-S., Lee, H.-E., Han, K., & Kang, B.-C. (2019). Genotyping of octoploid strawberry inbred lines by SNP discovery using genotyping-by-sequencing. *Horticulture, Environment, and Biotechnology*, *60*(1), 69-80.
- Kim, Y., & Nielsen, R. (2004). Linkage disequilibrium as a signature of selective sweeps. *Genetics*, *167*(3), 1513-1524.
- Knapp, S., Stroup, W., & Ross, W. (1985). Exact confidence intervals for heritability on a progeny mean basis 1. *Crop Science*, *25*(1), 192-194.
- Li, W., Ahn, I.-P., Ning, Y., Park, C.-H., Zeng, L., Whitehill, J. G., Lu, H., Zhao, Q., Ding, B., & Xie, Q. (2012). The U-Box/ARM E3 ligase PUB13 regulates cell death, defense, and flowering time in *Arabidopsis*. *Plant Physiology*, *159*(1), 239-250.
- Lian, X., Zeng, J., Zhang, H., Converse, R., Wang, Y., Bai, X., & Zhu, L. (2019). *PUB7*, a pollen expression gene induced by self-pollination, negatively regulates pollen germination. *Acta biochimica et biophysica Sinica*, *51*(5), 548-551.
- Louis-Marie, R., Jérôme, P., Mike, B., Angélique, D. H., & Jean-Christophe, G. (2008). Analysis of genome-wide linkage disequilibrium in the highly polyploid sugarcane. *Theoretical Applied Genetics*, *116*(5), 701-714.
- Lu, X., Shu, N., Wang, D., Wang, J., Chen, X., Zhang, B., Wang, S., Guo, L., Chen, C., & Ye, W. (2020). Genome-wide identification and expression analysis of *PUB* genes in cotton. *BMC genomics*, *21*(1), 1-12.
- Lu, Y., Shah, T., Hao, Z., Taba, S., Zhang, S., Gao, S., Liu, J., Cao, M., Wang, J., & Prakash, A. B. (2011). Comparative SNP and haplotype analysis reveals a higher genetic diversity and rapider LD decay in tropical than temperate germplasm in maize. *PLoS One*, *6*(9), e24861.

- McIntyre, C. L., Goode, M. L., Cordeiro, G., Bundock, P., Elliott, F., Henry, R. J., Casu, R. E., Bonnett, G. D., & Aitken, K. S. (2015). Characterisation of alleles of the sucrose phosphate synthase gene family in sugarcane and their association with sugar-related traits. *Molecular breeding*, *35*(3), 98.
- McIntyre, C. L., Jackson, M., Cordeiro, G. M., Amouyal, O., Hermann, S., Aitken, F. S., Elliott, F., Henry, R. J., Casu, R. E., & Bonnett, G. D. (2006). The identification and characterisation of alleles of sucrose phosphate synthase gene family III in sugarcane. *Molecular breeding*, *18*(1), 39-50.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., & Daly, M. (2010). The Genome Analysis Toolkit: a Map Reduce framework for analyzing next-generation DNA sequencing data. *Genome research*, *20*(9), 1297-1303.
- Mukherjee, D., Saha, D., Acharya, D., Mukherjee, A., Chakraborty, S., & Ghosh, T. C. (2018). The role of introns in the conservation of the metabolic genes of *Arabidopsis thaliana*. *Genomics*, *110*(5), 310-317.
- Nelson, J. C., Wang, S., Wu, Y., Li, X., Antony, G., White, F. F., & Yu, J. (2011). Single-nucleotide polymorphism discovery by high-throughput sequencing in sorghum. *BMC genomics*, *12*(1), 352.
- Nigro, D., Gadaleta, A., Mangini, G., Colasuonno, P., Marcotuli, I., Giancaspro, A., Giove, S., Simeone, R., & Blanco, A. (2019). Candidate genes and genome-wide association study of grain protein content and protein deviation in durum wheat. *Planta*, *249*(4), 1157-1175.
- Nimmakayala, P., Levi, A., Abburi, L., Abburi, V. L., Tomason, Y. R., Saminathan, T., Vajja, V. G., Malkaram, S., Reddy, R., & Wehner, T. C. (2014). Single nucleotide polymorphisms generated by genotyping by sequencing to characterize genome-wide diversity, linkage disequilibrium, and selective sweeps in cultivated watermelon. *BMC genomics*, *15*(1), 767.
- Okura, V. K., de Souza, R. S., de Siqueira Tada, S. F., & Arruda, P. (2016). BAC-pool sequencing and assembly of 19 Mb of the complex sugarcane genome. *Frontiers in plant science*, *7*, 342.

- Oraguzie , N. C., Gardiner , S. E., Rikkerink , E. H., & Silva , H. N. (2007). *Association mapping in plants*: Springer.
- Parida, S. K., Kalia, S., Pandit, A., Nayak, P., Singh, R. K., Gaikwad, K., Srivastava, P. S., Singh, N. K., & Mohapatra, T. (2016). Single nucleotide polymorphism in sugar pathway and disease resistance genes in sugarcane. *Plant cell reports*, *35*(8), 1629-1653.
- Piriyapongsa, J., Kaewprommal, P., Vaiwsri, S., Anuntakarun, S., Wirojsirasak, W., Punpee, P., Klomsa-ard, P., Shaw, P. J., Pootakham, W., & Yoocha, T. (2018). Uncovering full-length transcript isoforms of sugarcane cultivar Khon Kaen 3 using single-molecule long-read sequencing. *PeerJ*, *6*, e5818.
- Qian, L., Qian, W., & Snowdon, R. J. (2014). Sub-genomic selection patterns as a signature of breeding in the allopolyploid *Brassica napus* genome. *BMC genomics*, *15*(1), 1170.
- Qin, Q., Wang, Y., Huang, L., Du, F., Zhao, X., Li, Z., Wang, W., & Fu, B. (2020). A U-box E3 ubiquitin ligase OsPUB67 is positively involved in drought tolerance in rice. *Plant molecular biology*, *102*(1-2), 89-107.
- Raboin, L. M., Oliveira, K. M., Lecunff, L., Telismart, H., Roques, D., Butterfield, M., Hoarau, J. Y., & D'Hont, A. (2006). Genetic mapping in sugarcane, a high polyploid, using bi-parental progeny: identification of a gene controlling stalk colour and a new rust resistance gene. *Theoretical Applied Genetics*, *112*(7), 1382-1391.
- Raboin , L. M., Pauquet , J., Butterfield , M., D'Hont , A., & Glaszmann , J. C. (2008). Analysis of genome-wide linkage disequilibrium in the highly polyploid sugarcane. *Theoretical Applied Genetics*, *116*(5), 701-714. doi:10.1007/s00122-007-0703-1
- Racedo, J., Gutiérrez, L., Perera, M. F., Ostengo, S., Pardo, E. M., Cuenya, M. I., Welin, B., & Castagnaro, A. P. (2016). Genome-wide association mapping of quantitative traits in a breeding population of sugarcane. *BMC plant biology*, *16*(1), 142.
- Rafalski, A. (2002). Applications of single nucleotide polymorphisms in crop genetics. *Current opinion in plant biology*, *5*(2), 94-100.

- Rozen, S., & Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. In *Bioinformatics methods and protocols* (pp. 365-386): Springer.
- Ryu, M. Y., Cho, S. K., Hong, Y., Kim, J., Kim, J. H., Kim, G. M., Chen, Y.-J., Knoch, E., Møller, B. L., & Kim, W. T. (2019). Classification of barley U-box E3 ligases and their expression patterns in response to drought and pathogen stresses. *BMC genomics*, *20*(326), 1-15.
- Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, & DL, W. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, *409*(6822), 928.
- Saïdou, A.-A., Clotault, J., Couderc, M., Mariac, C., Devos, K. M., Thuillet, A.-C., Amoukou, I. A., & Vigouroux, Y. (2014). Association mapping, patterns of linkage disequilibrium and selection in the vicinity of the PHYTOCHROME C gene in pearl millet. *Theoretical Applied Genetics*, *127*(1), 19-32.
- Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the national academy of sciences*, *74*(12), 5463-5467.
- Sela, H., Loutre, C., Keller, B., Schulman, A., Nevo, E., Korol, A., & Fahima, T. (2011). Rapid linkage disequilibrium decay in the *Lr10* gene in wild emmer wheat (*Triticum dicoccoides*) populations. *Theoretical and Applied Genetics*, *122*(1), 175-187.
- Shearman, J. R., Sonthirod, C., Naktang, C., Pootakham, W., Yoocha, T., Sangsrakru, D., Jomchai, N., Tragoonrung, S., & Tangphatsornruang, S. (2016). The two chromosomes of the mitochondrial genome of a sugarcane cultivar: assembly and recombination analysis using long PacBio reads. *Scientific reports*, *6*(1), 1-7.
- Shendure, J., & Ji, H. (2008). Next-generation DNA sequencing. *Nature biotechnology*, *26*(10), 1135.
- Singh, B., & Singh, A. K. (2015). *Marker-assisted plant breeding: principles and practices*: Springer.

- Singh, D., Singh, B., Mishra, S., Singh, A. K., & Singh, N. K. (2019). Candidate gene based association analysis of salt tolerance in traditional and improved varieties of rice (*Oryza sativa* L.). *Journal of Plant Biochemistry and Biotechnology*, *28*(1), 76-83.
- Singh , R. K., Banerjee , N., Khan , M. S., Yadav , S., Kumar , S., Duttamajumder , S. K., Lal , R. J., Patel , J. D., Guo , H., & Zhang , D. (2016). Identification of putative candidate genes for red rot resistance in sugarcane (*Saccharum* species hybrid) using LD-based association mapping. *Molecular Genetics Genomics*, *291*(3), 1363-1377.
- Siraree, A., Banerjee, N., Kumar, S., Khan, M., Singh, P., Sharma, S., Singh, R., & Singh, J. (2017). Identification of marker-trait associations for morphological descriptors and yield component traits in sugarcane. *Physiology molecular biology of plants*, *23*(1), 185-196.
- Song , J., Yang , X., Resende Jr , M. F., Neves , L. G., Todd , J., Zhang , J., Comstock , J. C., & Wang , J. (2016). Natural allelic variations in highly polyploidy *Saccharum* complex. *Frontiers in plant science*, *7*(804), 1-18.
- Stephan, W. (2019). Selective sweeps. *Genetics*, *211*(1), 5-13.
- Storey, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B*, *64*(3), 479-498.
- Sunyaev, S., Hanke, J., Aydin, A., Wirkner, U., Zastrow, I., Reich, J., & Bork, P. (1999). Prediction of nonsynonymous single nucleotide polymorphisms in human disease-associated genes. *Journal of molecular medicine*, *77*(11), 754-760.
- Tabor, H. K., Risch, N. J., & Myers, R. M. (2002). Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nature Reviews Genetics*, *3*(5), 391-397.
- Thirugnanasambandam, P. P., Hoang, N. V., Furtado, A., Botha, F. C., & Henry, R. J. (2017). Association of variation in the sugarcane transcriptome with sugar content. *BMC genomics*, *18*(1), 909.
- Trujillo, M. (2018). News from the PUB: plant U-box type E3 ubiquitin ligases. *Journal of experimental botany*, *69*(3), 371-384.

- Ukoskit , K., Posudsavang , G., Pongsiripat , N., Chatwachirawong , P., Klomsa-ard , P., Poomipant , P., & Tragoonrung , S. (2019). Detection and validation of EST-SSR markers associated with sugar-related traits in sugarcane using linkage and association mapping. *Genomics*, *111*(1), 1-9.
- Vilela, M. d. M., Del Bem, L. E., Van Sluys, M.-A., de Setta, N., Kitajima, J. P., Cruz, G. M. Q., Sforça, D. A., de Souza, A. P., Ferreira, P. C. G., & Grativol, C. (2017). Analysis of three sugarcane homo/homeologous regions suggests independent polyploidization events of *Saccharum officinarum* and *Saccharum spontaneum*. *Genome biology and evolution*, *9*(2), 266-278.
- Wang , J., Qu, B., Dou, S., Li, L., Yin, D., Pang, Z., Zhou, Z., Tian, M., Liu, G., & Xie, Q. (2015). The E3 ligase OsPUB15 interacts with the receptor-like kinase PID2 and regulates plant cell death and innate immunity. *BMC plant biology*, *15*(1), 1-15.
- Wang , W., Wang , W., Wu , Y., Li , Q., Zhang , G., Shi , R., Yang , J., Wang , Y., & Wang , W. (2020). The involvement of wheat U-box E3 ubiquitin ligase TaPUB1 in salt stress tolerance. *Journal of Integrative Plant Biology*, *62*(5), 631-651.
- Wang, Y.-H., Upadhyaya, H. D., Burrell, A. M., Sahraeian, S. M. E., Klein, R. R., & Klein, P. E. (2013). Genetic structure and linkage disequilibrium in a diverse, representative collection of the C4 model plant, *Sorghum bicolor*. *G3: Genes, Genomes, Genetics*, *3*(5), 783-793.
- Whitt, S. R., & Buckler, E. S. (2003). Using natural allelic diversity to evaluate gene function. In *Plant functional genomics* (pp. 123-139): Springer.
- Wu, Y., Zhou, Q., Huang, S., Wang, G., & Xu, L. (2019). SNP development and diversity analysis for *Ginkgo biloba* based on transcriptome sequencing. *Trees*, *33*(2), 587-597.
- Xu, J., Yuan, Y., Xu, Y., Zhang, G., Guo, X., Wu, F., Wang, Q., Rong, T., Pan, G., & Cao, M. (2014). Identification of candidate genes for drought tolerance by whole-genome resequencing in maize. *BMC plant biology*, *14*(1), 83.
- Yang, X., Luo, Z., Todd, J., Sood, S., & Wang, J. (2020). Genome-wide association study of multiple yield traits in a diversity panel of polyploid sugarcane (*Saccharum* spp.). *The Plant Genome*, e20006.

- Yang, X., Sood, S., Glynn, N., Islam, M. S., Comstock, J., & Wang, J. (2017). Constructing high-density genetic maps for polyploid sugarcane (*Saccharum* spp.) and identifying quantitative trait loci controlling brown rust resistance. *Molecular breeding*, 37(10), 116.
- Yang, X., Todd, J., Arundale, R., Binder, J. B., Luo, Z., Islam, M. S., Sood, S., & Wang, J. (2019). Identifying loci controlling fiber composition in polyploid sugarcane (*Saccharum* spp.) through genome-wide association study. *Industrial Crops and Products*, 130, 598-605.
- Yu, J., & Buckler, E. S. (2006). Genetic association mapping and genome organization of maize. *Current opinion in biotechnology*, 17(2), 155-160.
- Yu, X., Pijut, P. M., Byrne, S., Asp, T., Bai, G., & Jiang, Y. (2015). Candidate gene association mapping for winter survival and spring regrowth in perennial ryegrass. *Plant science*, 235, 37-45.
- Zhang, J., Zhang, X., Tang, H., Zhang, Q., Hua, X., Ma, X., Zhu, F., Jones, T., Zhu, X., & Bowers, J. (2018). Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. *Nature genetics*, 50(11), 1565.
- Zhang, W., Wang, L., Wang, Y., Wang, Y., & Gao, Q. (2020). Cloning and characterization of *TaPUB4*, a U-box gene from common wheat (*Triticum aestivum* L.) involved in regulation of pollen development by influencing sucrose-starch pathway in anther. *Molecular breeding*, 40(6), 1-18.
- Zhao, J., Zhao, L., Zhang, M., Zafar, S. A., Fang, J., Li, M., Zhang, W., & Li, X. (2017). Arabidopsis E3 ubiquitin ligases PUB22 and PUB23 negatively regulate drought tolerance by targeting ABA receptor PYL9 for degradation. *International journal of molecular sciences*, 18(9), 1841.
- Zhu, C., Gore, M., Buckler, E. S., & Yu, J. (2008). Status and prospects of association mapping in plants. *The Plant Genome*, 1(1), 5-20.

วิทยานิพนธ์

ฉัตร สุภาพไพบุลย์กุล.(2561). การทวนสอบเครื่องหมายดีเอ็นเอที่มีความสัมพันธ์กับลักษณะความหวานในอ้อย. (โครงการปัญหาพิเศษ). มหาวิทยาลัยธรรมศาสตร์, คณะวิทยาศาสตร์และเทคโนโลยี, สาขาวิชาเทคโนโลยีชีวภาพ.

Khanbo, S. (2020). *Association Mapping of Sugar Related Traits in Sugarcane (Saccharum spp.)*. (Doctoral Dissertation), Thammasat University, Pathum Thani.

อื่น ๆ

R Core Team. (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.

S.A.S. Institute Inc. (2011). *Statistical Software Analysis for Windows (9.3 ed.)*. Cary, NC, USA

ภาคผนวก



ภาคผนวก ก
ตัวอย่างอ้อยที่ใช้ในการศึกษา

ตัวอย่างอ้อยจำนวน 141 พันธุ์ พันธุ์ประวัติ และแหล่งที่มา

No.	Varieties	Female	Male	Sources	No.	Varieties	Female	Male	Sources
1	B41721			Barbados	26	F177	F108	Co290	Taiwan
2	B4744	Co421	B4098	Barbados	27	FJI105			Fiji
3	B76718			Barbados	28	GALOA	LF 46/1902	CO290	Fiji
4	BL4	POJ2878	Polycross	Barbados	29	GT29			China
5	BO14	Co331	POJ2878	India	30	IAC49_131	CP27-108		Brazil
6	BO24	Co313	Co356	India	31	IAC51_205	POJ2878		Brazil
7	CAC57_23			Philippines	32	IAC52_156			Brazil
8	CO1287	Co603	Co453	India	33	IRK67_1			USA
9	CO290	Co221	Kansar	India	34	เคียวจีน			China
10	CO331	Co213	Co214	India	35	KWT3			China
11	CP45_150	Co281	US1694	USA	36	LF48_8077			Fiji
12	CP61_39	CP55-38	CP48-103	USA	37	LF76_2300			Fiji
13	CP63_306	CP50-28	CP52-15	USA	38	LF78_960			Fiji
14	CP72_120			USA	39	LF79_2964			Fiji
15	CP75_330			USA	40	M124/59			Mauritius
16	CP76_340			USA	41	M13/58			Mauritius
17	D158	Co421	D166/34	Guyana	42	M147/158			Mauritius
18	DB7160	Saipan17	B63118	Barbados	43	MACKAY			Australia
19	Ebene1_37	Ebene1/34	M27/16	Mauritius	44	MANA			Fiji
20	Eros	POJ2878	31MQ228	Australia	45	MOSSMAN			Australia
21	F140	PT43-52	CP34-79	Taiwan	46	MY55_14	CP34-79	B45181	Cuba
22	F148	F138	PT48-21	Taiwan	47	N50_211	POJ2725	Co285	South Africa
23	F152	H32-8560	PT43-52	Taiwan	48	NCO310	Co421	Co312	South Africa
24	F153	NCo310	34-136	Taiwan	49	NCO382	POJ2725	Co301	South Africa
25	F170	CoL9	54-182	Taiwan	50	PHIL66_07	Phil5460	Co440	Philippines

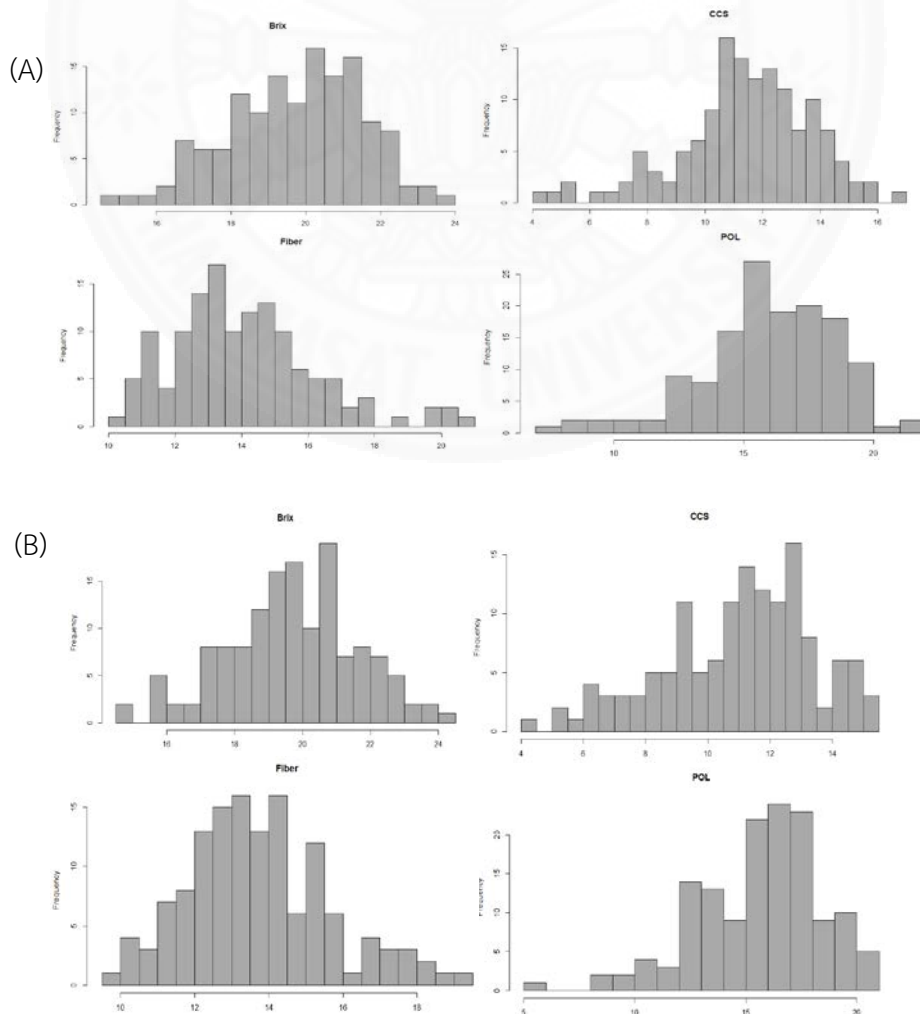
No.	Varieties	Female	Male	Sources	No.	Varieties	Female	Male	Sources
51	PINDAR	Co270	33MQ157	Australia	76	CN1	F160	Co775	Thailand
52	PL310				77	K84_200	ROC1	CP63-588	Thailand
53	PR3067			USA	78	K92_80	K84-200	K76-4	Thailand
54	Q100	Trojan	Co475	Australia	79	K93_219	UT1	Ehaew	Thailand
55	Q115	NCo310	54N7096	Australia	80	K95_84	K90-79	K84-200	Thailand
56	Q117	Q77	58N829	Australia	81	K99_72	K84-200	Ehaew	Thailand
57	Q146	62N1659	66C807	Australia	82	KK3	K84-200	85-2-352	Thailand
58	Q208	Q135	QN61-1232	Australia	83	KKU99_01			Thailand
59	Q47	Co290	POJ2878	Australia	84	KKU99_03			Thailand
60	Q77	H59-198	H39-3633	Australia	85	KKU99_06			Thailand
61	Q81	Trojan	M336	Australia	86	KPK98_40			Thailand
62	Q83	Trojan	M336	Australia	87	KPS01_12	Kps 94-13	UT3	Thailand
63	R397	Co281	V114	France	88	KPS01_25	Kps 94-13	UT3	Thailand
64	ROC1	F146	CP58-48	Taiwan	89	LK92_11	K84-200	Ehaew	Thailand
65	ROC22			Taiwan	90	MP1	CoC88-011	Co87-025	Thailand
66	ROC7	F160	60-2321	Taiwan	91	MP2	CoC67-1	Co88-005	Thailand
67	TUC74_6	L60-25	CP44-155	Argentina	92	MP3	PT52-227	Q142	Thailand
68	YASAWA	LF49-3863	36MQ2717	Fiji	93	MPT02_458			Thailand
69	BIOTECH1	CN1	<i>S. spontaneum</i>	Thailand	94	MPT02_612			Thailand
70	BIOTECH2	CN1	<i>S. spontaneum</i>	Thailand	95	MPT02_665			Thailand
71	BIOTECH3	CN1	<i>S. spontaneum</i>	Thailand	96	MPT02_669			Thailand
72	BIOTECH4			Thailand	97	MPT03_166			Thailand
73	BIOTECH5			Thailand	98	MPT04_169			Thailand
74	BIOTECH6			Thailand	99	MPT04_467			Thailand
75	SP50	SP074		Thailand	100	MPT05_187			Thailand

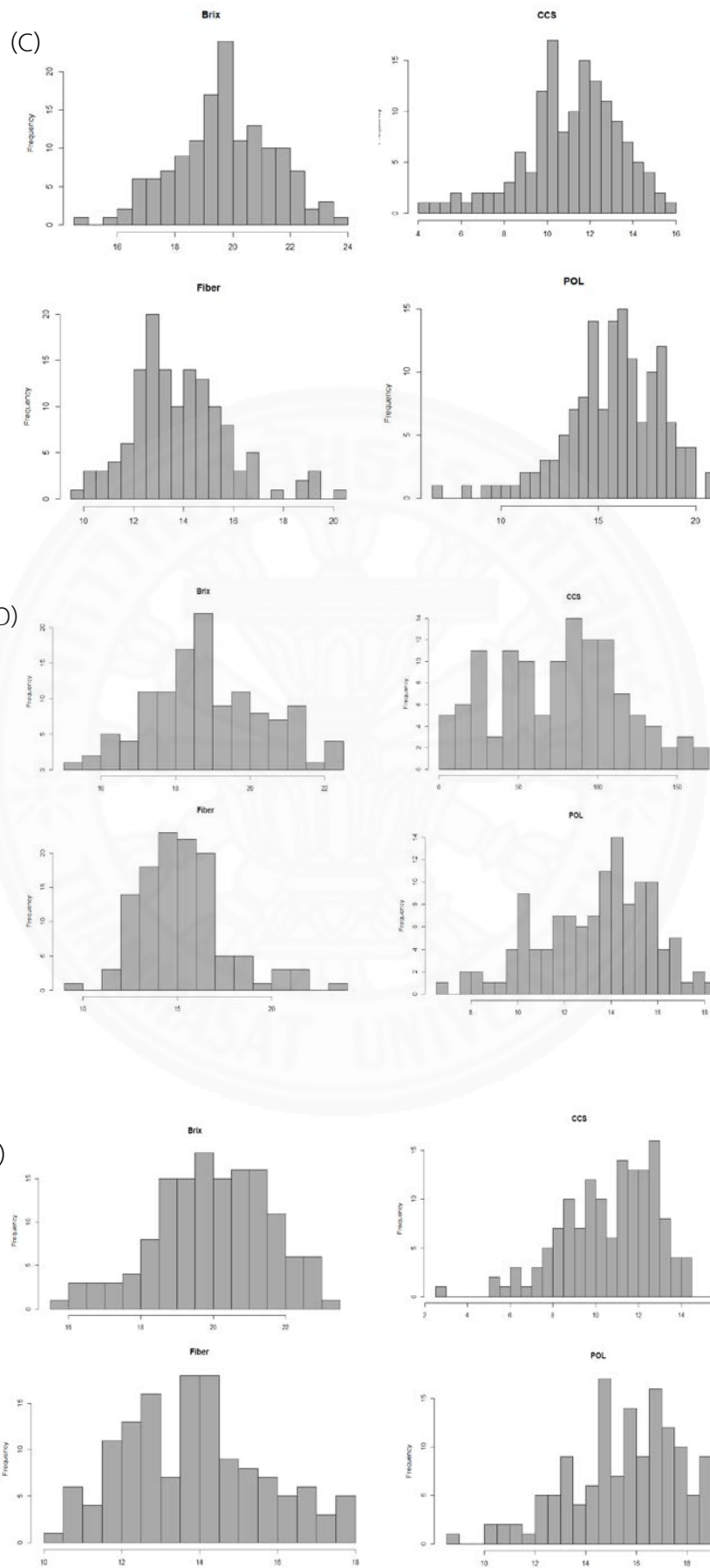
No.	Varieties	Female	Male	Sources	No.	Varieties	Female	Male	Sources
101	MPT06_144			Thailand	122	MPT14_1_468	K93-219	KpK98-40	Thailand
102	MPT06_166			Thailand	123	MPT14_1_504	K93-219	KpK98-40	Thailand
103	MPT06_171			Thailand	124	MPT14_1_546	K93-219	KpK98-40	Thailand
104	MPT06_344			Thailand	125	MPT14_1_857	K93-219	KpK98-40	Thailand
105	MPT06_367			Thailand	126	MPT14_1_892	K93-219	KpK98-40	Thailand
106	MPT07_1			Thailand	127	MPT14_1_976	K93-219	KpK98-40	Thailand
107	MPT07_152			Thailand	128	MPT14_2_662	K93-219	TBy20- 2248	Thailand
108	MPT07_71			Thailand	129	MPT14_5_216	UT1	KpK98-40	Thailand
109	MPT08_191			Thailand	130	MPT14_5_239	UT1	KpK98-40	Thailand
110	MPT08_3			Thailand	131	MPT14_6_63	KK3	TBy20- 2248	Thailand
111	MPT08_50			Thailand	132	MPT14_9_133	LK92-11	LK95-124	Thailand
112	MPT09_118			Thailand	133	MPT99_1444			Thailand
113	MPT09_204			Thailand	134	SP80	85-2- 352	K84-200	Thailand
114	MPT09_296			Thailand	135	TBY20_2248	K83-74	K84-200	Thailand
115	MPT10_54			Thailand	136	UT1	F172		Thailand
116	MPT11_181			Thailand	137	UT12	SP80	UT3	Thailand
117	MPT12_141			Thailand	138	UT3	UT1	UT2	Thailand
118	MPT12_1782			Thailand	139	UT5	87-2- 1033		Thailand
119	MPT13_118			Thailand	140	UT6	87-2- 973	83-2-888	Thailand
120	MPT14_1_172	K93-219	KpK98-40	Thailand	141	UT8	K84-200	UT3	Thailand
121	MPT14_1_180	K93-219	KpK98-40	Thailand					

ภาคผนวก ข

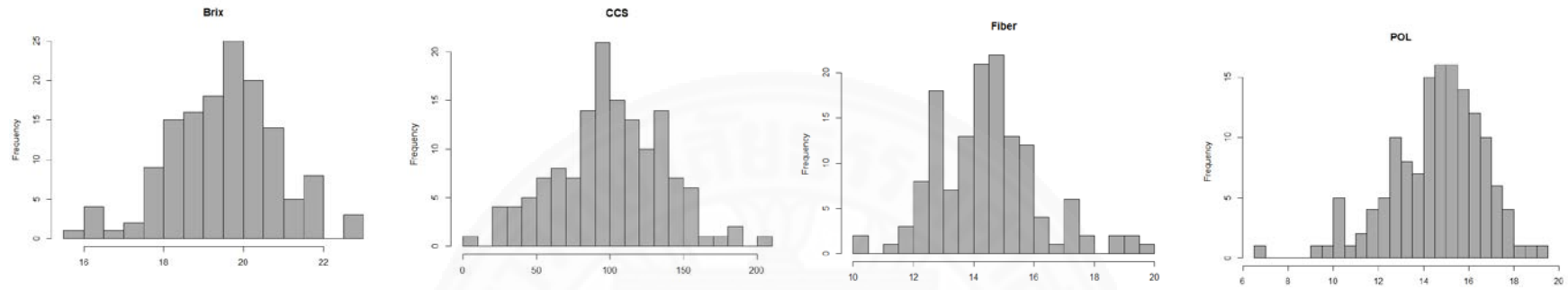
การแจกแจงความถี่ของข้อมูลลักษณะที่เกี่ยวข้องกับน้ำตาลทั้ง 4 ลักษณะ

ฮิสโทแกรมแสดงการแจกแจงความถี่ของข้อมูลลักษณะน้ำตาลทั้ง 4 ลักษณะ ได้แก่ ค่าความเข้มข้นของของแข็งที่ละลายอยู่ในน้ำอ้อย (Brix) ปริมาณน้ำตาลที่มีอยู่ในอ้อยที่สามารถหีบสกัดออกมาได้เป็นน้ำตาลทรายขาวบริสุทธิ์ (commercial cane sugar; CCS) ของแข็งที่ไม่ละลายน้ำ (Fiber) และค่าร้อยละของน้ำตาลซูโครส (POL) ของอ้อยปลูก แปลงการทดลองมหาวิทยาลัยขอนแก่น จังหวัดขอนแก่น (A) แปลงศูนย์นวัตกรรมและการวิจัยมิตรผล อำเภอกุฉีชุม จังหวัดชัยภูมิ (B) และอ้อยปลูกแบบรวมทั้งสองแปลงการทดลอง (C) อ้อยต่อ แปลงการทดลองมหาวิทยาลัยขอนแก่น จังหวัดขอนแก่น (D) แปลงศูนย์นวัตกรรมและการวิจัยมิตรผล อำเภอกุฉีชุม จังหวัดชัยภูมิ (E) และอ้อยต่อแบบรวมแปลงการทดลองทั้งสองแปลง (F) และตารางแสดงการทดสอบการแจกแจงแบบปกติของข้อมูลด้วยสถิติ Kolmogorov-Smirnov ที่ระดับนัยสำคัญ 0.01 (G)





(F)



(G)

Tests of Normality																		
Traits	KK PLANTED			MPT PLANTED			PLANTED			KK 1st RATOON			MPT 1st RATOON			1 st RATOON		
	Kolmogorov-Smirnova			Kolmogorov-Smirnova			Kolmogorov-Smirnova			Kolmogorov-Smirnova			Kolmogorov-Smirnova			Kolmogorov-Smirnova		
	Statistic	df	Sig.	Statistic	df	Sig.	Statistic	df	Sig.	Statistic	df	Sig.	Statistic	df	Sig.	Statistic	df	Sig.
Brix	0.07	140	0.093	0.037	141	0.200	0.046	141	0.200	0.084	122	0.34	0.047	141	0.200	0.48	141	0.200
CCS	0.082	137	0.024	0.083	138	0.020	0.064	139	0.200	0.059	122	0.2	0.086	141	0.013	0.54	141	0.200
Fiber	0.077	133	0.050	0.072	135	0.085	0.063	135	0.200	0.086	119	0.029	0.085	137	0.018	0.71	138	0.082
POL	0.065	140	0.200	0.073	141	0.066	0.064	141	0.200	0.090	121	0.018	0.074	141	0.058	0.071	141	0.080

ประวัติผู้เขียน

ชื่อ	นางสาวกฤติยา นิลแก้ว
วันเดือนปีเกิด	16 เมษายน พ.ศ.2538
วุฒิการศึกษา	ปีการศึกษา 2559: วิทยาศาสตร์บัณฑิต (เทคโนโลยีชีวภาพ) มหาวิทยาลัยธรรมศาสตร์
ทุนการศึกษา (ถ้ามี)	ปีงบประมาณ 2560: ทุนบัณฑิตเรียนดี คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ ปีงบประมาณ 2561: ทุนผู้ช่วยสอน มหาวิทยาลัยธรรมศาสตร์
ประสบการณ์ทำงาน	พ.ศ.2561: อาจารย์พิเศษวิชา ทช.319 เทคโนโลยีชีวภาพ สำหรับการบำบัดของเสียและน้ำเสีย พ.ศ.2561-2562: อาจารย์พิเศษวิชา ทช.284 ปฏิบัติการ ชีวเคมี

ผลงานทางวิชาการ

กฤติยา นิลแก้ว, สิทธิโชค ตั้งภัสสรเรือง, ชุติมา สนธิรอด, และกิตติพัฒน์ อุโฆษกิจ. การวิเคราะห์ความไม่สมดุลของสิ่งเจือปนภายในยีน *Plant U-box* ในอ้อย. วารสาร Thai Journal of Science and Technology. กำลังจัดพิมพ์ 2020.