

การวิเคราะห์ความไม่สมดุลของลิงเกจภายในยีน *Plant U-box* ในอ้อย

Linkage Disequilibrium in *Plant U-box* Gene in Sugarcane (*Saccharum* spp.)

กฤติยา นิลแก้ว และ กิตติพัฒน์ อุโฆษกิจ*

สาขาวิชาเทคโนโลยีชีวภาพ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ ศูนย์รังสิต

สิทธิโชค ตั้งภัสสรเรือง และ ชุตินา สนธิรอด

ศูนย์โอมิกส์แห่งชาติ สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ (สวทช.)

Krittiya Ninkaeo and Kittipat Ukoskit*

Department of Biotechnology, Faculty of Science and Technology, Thammasat University, Rangsit Centre

Sithichoke Tangphatsornruang and Chutima Sonthirod

National Omics Center, National Science and Technology Development Agency (NSTDA)

Received: March 13, 2020 ; Accepted: January 18, 2021

บทคัดย่อ

ความเข้าใจเกี่ยวกับความไม่สมดุลของลิงเกจ (linkage disequilibrium; LD) มีความสำคัญต่อการวางแผนกลยุทธ์การปรับปรุงพันธุ์ เพื่อกำหนดแอลลีลเป้าหมายที่เป็นประโยชน์ต่อลักษณะทางการเกษตร งานวิจัยนี้วิเคราะห์รูปแบบ LD ของยีน U-box domain-containing protein 35 (SaPUB35) ในประชากรเชื้อพันธุ์กรรมนำเข้ามา 68 ตัวอย่าง และอ้อยพันธุ์ปรับปรุง 73 ตัวอย่าง สามารถระบุเครื่องหมายสลับได้ 126 เครื่องหมาย และอินเดล 50 เครื่องหมายจากยีน SaPUB35 ขนาด 5731 คู่เบสด้วยเทคโนโลยี PacBio single molecule, real-time (SMRT) sequencing ความแปรปรวนลำดับเบสในประชากรอ้อยเชื้อพันธุ์กรรมนำเข้ามา มากกว่าประชากรอ้อยพันธุ์ปรับปรุง การวิเคราะห์ LD ระหว่างคู่เครื่องหมายโดยใช้เกณฑ์ค่ายกกำลังสองของค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างเครื่องหมายสองตำแหน่งที่ 0.1 พบว่าในประชากรพันธุ์ปรับปรุงมีขอบเขต LD 1500 คู่เบส แต่มีความเสื่อมลงไปที่ 1400 คู่เบสในประชากรเชื้อพันธุ์กรรมนำเข้ามา แสดงเป็นสัญญาณถึงอิทธิพลของ selective sweep ซึ่งคัดเลือกการกลายที่มีผลดีในระหว่างกระบวนการปรับปรุงพันธุ์อ้อย การค้นพบการลดความหลากหลายของแอลลีลและเกิดความหลากหลายทางพันธุกรรมแบบคอขวดในระหว่างกระบวนการปรับปรุงพันธุ์อ้อยสมัยใหม่จะเป็นประโยชน์ในการทำความเข้าใจเป้าหมายของการปรับปรุงพันธุ์อ้อยและวางแผนกลยุทธ์การปรับปรุงพันธุ์ในอนาคต เพื่อกำหนดเป้าหมายแอลลีลที่เป็นประโยชน์

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

Abstract

Understanding linkage disequilibrium (LD) is important for planning breeding strategies to target beneficial alleles for agronomic traits. In the present study, we analyzed the LD pattern of *U-box domain-containing protein35* gene (*SaPUB35*) in 68 exotic sugarcane collections and 73 sugarcane breeding lines. 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 sugarcane population has more variants than the breeding line population. 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 population but is eroded to 1400 bp in exotic sugarcane population. This implies a signature of a selective sweep for a beneficial mutation during the selection processes of sugarcane breeding. Uncovering allelic erosion and a diversity bottleneck due to modern sugarcane breeding will be useful for understanding sugarcane breeding targets and in devising future breeding strategies to target beneficial alleles currently not pursued.

Keyword: sugarcane; single nucleotide polymorphism (SNP); linkage disequilibrium (LD); PacBio single molecule, real-time (SMRT) sequencing technology

1. บทนำ

อ้อยเป็นพืชหลักที่ใช้ในการผลิตน้ำตาลทราย และเป็นพืชเศรษฐกิจที่สำคัญของไทย เพื่อเป็นการเพิ่มมูลค่าทางเศรษฐกิจและปริมาณผลผลิต จึงมีการปรับปรุงพันธุ์อ้อยเพื่อให้ได้พันธุ์ที่มีปริมาณผลผลิตและความหวานสูง ในปัจจุบันมีการประยุกต์ใช้เครื่องหมายดีเอ็นเอในการคัดเลือกลูกผสมที่ได้จากการปรับปรุงพันธุ์ เนื่องจากช่วยลดทั้งระยะเวลาและค่าใช้จ่าย แต่มีประสิทธิภาพและความแม่นยำในการคัดเลือกพันธุ์อ้อยที่มีลักษณะตามที่ต้องการได้มากขึ้น โดยเครื่องหมายดีเอ็นเอที่กำลังเป็นที่นิยมศึกษาและพัฒนาเพื่อใช้เป็นเครื่องหมายคัดเลือกในอนาคตคือเครื่องหมายสแน็ป (single nucleotide polymorphism; SNP) เป็นความแปรปรวนของพันธุกรรมที่เกิดจากการกลายและรีคอมบิเนชัน จึงมีความหนาแน่นและกระจายตัวทั่วจีโนมสามารถวิเคราะห์ความสัมพันธ์ของยีน (linkage disequilibrium; LD) ภายในยีนได้ อีกทั้ง

การศึกษาเครื่องหมาย SNP ภายในดีเอ็นเอเป้าหมายในปัจจุบันเป็นเรื่องที่ง่ายขึ้น เนื่องจากการพัฒนาของเทคโนโลยีการวิเคราะห์ลำดับเบสยุคใหม่ หนึ่งในนั้นคือเทคโนโลยี PacBio single molecule, real-time (SMRT) sequencing ซึ่งเป็นเทคโนโลยีที่วิเคราะห์ดีเอ็นเอธรรมชาติโมเลกุลเดี่ยวแบบตามเวลาจริง โดยตรวจจับการเข้าต่อสายของนิวคลีโอไทด์ที่ติดฉลากฟลูออเรสเซนต์ขณะที่เอนไซม์ดีเอ็นเอพอลิเมอเรสกำลังสังเคราะห์ดีเอ็นเอในหลุมขนาดเล็กบนแผ่นเซลล์ SMRT ข้อดีของเทคโนโลยีนี้คือสามารถวิเคราะห์ลำดับเบสสายยาวเฉลี่ยได้มากถึง 10-18 Kb และสามารถวิเคราะห์ลำดับเบสวนซ้ำได้ 10-15 รอบของโมเลกุลเดิมเนื่องจากดีเอ็นเอต้นแบบเป็นแบบวงแหวน ช่วยลดอัตราการคลาดเคลื่อนของการวิเคราะห์ลำดับเบสได้ (Eid et al., 2009; Sela et al., 2011) จึงมีงานวิจัยที่ใช้เทคโนโลยีนี้ในการวิเคราะห์ลำดับเบสในพืชชนิดต่าง ๆ ทั้งในข้าวโพด ข้าวฟ่าง และอ้อย (Abdel-

Ghany et al., 2016; Piriyaopongsa et al., 2018; Shearman et al., 2016; Wang et al., 2016) หนึ่งในวิธีการพัฒนาเครื่องหมายดีเอ็นเอที่ช่วยในการคัดเลือกที่เป็นที่นิยมมากคือการทำวิเคราะห์แผนของความสัมพันธ์ (association mapping) เนื่องจากสามารถระบุตำแหน่งยีนควบคุมลักษณะปริมาณซึ่งส่วนใหญ่เป็นลักษณะทางการเกษตรที่สำคัญทั้งปริมาณผลผลิต และคุณภาพของผลผลิต จากการตรวจสอบความสัมพันธ์ระหว่างเครื่องหมายดีเอ็นเอกับลักษณะที่สนใจ โดยอาศัยพื้นฐานของ LD ซึ่งเป็นความสัมพันธ์แบบไม่สุ่มระหว่างแอลลีลที่อยู่ต่างตำแหน่งกันอาจจะอยู่บนโครโมโซมโซมแห่งเดียวกันหรือคนละแห่ง ดังนั้นในการวิเคราะห์ความสัมพันธ์จึงเป็นตรวจสอบการมี LD กันระหว่างเครื่องหมายดีเอ็นเอกับแอลลีลภายในยีนที่ควบคุมลักษณะปริมาณว่ามีผลต่อการแสดงออกของลักษณะฟีโนไทป์หรือไม่ รวมถึงการศึกษาขอบเขตของความไม่สมดุลลิงเกจ และการเสื่อมถอยของความไม่สมดุลลิงเกจ (LD decay) ทำให้ทราบถึงระยะห่างระหว่างเครื่องหมายดีเอ็นเอกับแอลลีลของยีนที่ยังมีความสัมพันธ์กันอยู่ซึ่งมีความสำคัญต่อประสิทธิภาพของการวิเคราะห์แผนความสัมพันธ์ในการระบุความสัมพันธ์ที่แท้จริงระหว่างเครื่องหมายดีเอ็นเอและยีนที่ควบคุมลักษณะปริมาณ จึงมีการศึกษาขอบเขตของ LD ในพืชหลายชนิด เช่น อะราบิโดพซิสที่พบว่าค่าขอบเขตของ LD เฉลี่ยที่ 50 kb และเป็นค่าที่เหมาะสมสำหรับการวิเคราะห์ genome-wide LD mapping (Sela et al., 2011) ในข้าวและข้าวฟ่างพันธุ์พื้นเมืองมีขอบเขตของ LD อยู่ในช่วง 100 kb และในอ้อยมี LD decay ที่ระยะห่างทางพันธุกรรมประมาณ 5 cM เป็นต้น (Raboin et al., 2008)

จากการศึกษาก่อนหน้าของ Khanbo (2020) ในการพัฒนาเครื่องหมาย intron length polymorphism (ILP) จากยีนในกลุ่ม Plant U-box

(PUB) พบว่าสามารถระบุความสัมพันธ์กับค่าความเข้มข้นของของแข็งที่ละลายอยู่ในน้ำอ้อย (Brix) ซึ่งเป็นลักษณะที่เกี่ยวข้องกับลักษณะน้ำตาลในอ้อย (กำลังเตรียมต้นฉบับตีพิมพ์) จากการศึกษาเชิงเปรียบเทียบยีนนี้ในจีโนมของข้าวฟ่างซึ่งเป็นพืชที่มีบรรพบุรุษร่วมกันกับอ้อย (Jannoo et al., 2007) พบว่ามีความคล้ายคลึงกับยีน U-box domain-containing protein 35 ในข้าวฟ่าง โดยจากการศึกษาโปรตีน PUB family พบว่าประกอบด้วยโดเมน U-box ซึ่งเป็นลำดับกรดอะมิโนประมาณ 70 กรดอะมิโนที่มีความอนุรักษ์สูง สามารถจัดกลุ่มตามโดเมนที่พบได้ 9 กลุ่ม (Ryu et al., 2019) โปรตีน U-box domain-containing protein 35 ประกอบด้วยโดเมน U-box ที่ด้านปลายคาร์บอกซิลและโดเมน serine/threonine kinase ที่ด้านปลายอะมิโน จึงจัดอยู่ในกลุ่ม IV ของโปรตีน PUB family ในปัจจุบันมีการศึกษาและทำนายโปรตีน PUB ในพืชหลายชนิดทั้งในอะราบิโดพซิส, ข้าว, ข้าวบาร์เลย์, และข้าวสาลี (Chen et al., 2019; Ryu et al., 2019; Wang et al., 2020; Zhao et al., 2017) และจากการศึกษาหน้าที่ของโปรตีน PUB ในพืชชั้นสูง พบว่ามีความเกี่ยวข้องกับหลายกระบวนการ เช่นการตอบสนองต่อความเครียดทั้งแบบ abiotic (ความแล้ง อุณหภูมิและความเค็ม) และ biotic (เชื้อก่อโรค และเชื้อรา) การแบ่งและขยายขนาดของเซลล์ การผสมตัวเองไม่ติด การพัฒนาละอองเรณู จากการศึกษาในที่ตอบสนองต่อความแล้งในข้าวของ Qin et al. (2020) โดยการดัดแปลงพันธุกรรมในข้าวโดยทำให้ยีน OsPUB67 ไม่มีการแสดงออก (knock out) ด้วยระบบ CRISPR/Cas9 และทำให้ยีนนี้มีการแสดงออกที่มากขึ้น (overexpression) พบว่ายีน OsPUB67 ถูกชักนำให้มีการแสดงออกโดยความเครียดแบบ abiotic ทั้งความแล้ง อุณหภูมิต่ำ ความเค็ม รวมทั้ง abscisic acid และยังพบว่าภายใต้สภาวะแล้งนั้นข้าวที่ได้รับการดัดแปลงให้ยีน

OsPUB67 มีการแสดงออกมากขึ้นมีความทนแล้งมากกว่าข้าวที่ยีน OsPUB67 ไม่มีการแสดงออก ผลจากการศึกษานี้จึงคาดว่ายีน OsPUB67 อาจจะมี ความเกี่ยวข้องกับทนต่อความแล้ง การศึกษา ของ Zhang et al. (2020) พบว่ายีนที่ควบคุม กระบวนการเมแทบอลิซึมของน้ำตาลและแป้งมี ระดับการถอดรหัสลดลงอย่างมีนัยสำคัญในอะรา บิดอพซิสที่มีการกลายของยีน atpub4 แต่ในอะรา บิดอพซิสที่มีการกลายของยีน atpub4 ที่ได้รับการ ส่งถ่ายยีน TaPUB ที่ได้จาก cDNA ของอัลลลของ เรณูที่กำลังพัฒนาในข้าวสาลีพบว่าอะราบิดอพซิส สามารถกลับมาอีกระดับการถอดรหัสของยีนที่ ควบคุมกระบวนการเมแทบอลิซึมของน้ำตาลและ แป้งที่ใกล้เคียงกับอะราบิดอพซิสที่มีแอลลิลปกติ บ่งชี้ได้ว่ายีน TaPUB4 อาจควบคุมการพัฒนา ของละอองเรณูโดยควบคุมการสะสมแป้งในละออง เรณู และจากการศึกษายีน OsPUB73 ในข้าวของ Chen et al. (2019) โดยการ knock out ยีน OsPUB73 ด้วย CRISPR/Cas9 แล้วศึกษาเซลล์ วิทยาและทรานสคริปโตมของยีนที่เกี่ยวข้องกับการ พัฒนาของละอองเรณูและการเป็นหมันในข้าวเพื่อ ระบุยีนที่แสดงออกแตกต่างกันระหว่างตัวอย่างที่ไม่ มีการแสดงออกของยีน OsPUB73 และตัวอย่างที่มี แอลลิลปกติในระหว่างกระบวนการไมโอซิส พบว่า ยีนที่เกี่ยวข้องกับหลายกระบวนการมีการแสดงออก น้อยลง เช่นกระบวนการเมแทบอลิซึมของกลีเซอโร ลิพิด (glycerolipid) กระบวนการสังเคราะห์กรด ไขมันไม่อิ่มตัวและฟลาโวนอยด์ (flavonoid) กระบวนการย่อยสลายกรดไขมัน รวมถึง กระบวนการเมแทบอลิซึมของน้ำตาลและแป้ง จาก การศึกษาเหล่านี้แสดงให้เห็นว่าโปรตีน PUB นั้น อาจมีความเกี่ยวข้องกับการเมแทบอลิ ซึมของน้ำตาลและแป้งในพืช

ในการปรับปรุงพันธุ์โดยนักปรับปรุงพันธุ์นั้น จะดำเนินการวางแผนการผสมพันธุ์พ่อแม่และ

คัดเลือกเก็บไว้เฉพาะลูกผสมที่มีคุณสมบัติตรงตาม ความต้องการซึ่งมีผลต่อระดับของความหลากหลาย ทางพันธุกรรมอย่างมีนัยสำคัญ (Lu et al., 2011) การคัดเลือกเก็บเฉพาะลูกผสมที่มีลักษณะตรงตาม ความต้องการเป็นการคัดเลือกบริเวณบนยีนที่ส่งผล ต่อลักษณะที่ต้องการ จึงเป็นการตรึงการกลายที่มี ผลดี (beneficial mutation) ต่อลักษณะที่ต้องการไว้ (Kim and Nielsen, 2004) เรียกการคัดเลือกแบบนี้ ว่า selective sweep เป็นการคัดเลือกแบบกวาดเอา ทั้งบล็อกของ beneficial mutation และแอลลิล ขนาบข้างที่มี LD กับ beneficial mutation (LD block) (Wang et al., 2013) ซึ่ง LD block นี้จะถูก คัดเลือกเก็บไว้ในทุก ๆ รุ่น ทำให้เกิดปรากฏการณ์ ทางพันธุกรรมเป็นแบบคอขวด (genetic bottleneck) เนื่องจากการคัดเลือกนี้เกิดจากการ คัดเลือกโดยนักปรับปรุงพันธุ์ทำให้แอลลิลใน LD block นี้มีความถี่เพิ่มขึ้นในระยะเวลานับรวดเร็ว ขณะที่แอลลิลอื่น ๆ ที่ไม่ได้ถูกคัดเลือกมานั้นเกิด การหายไป (Kim and Nielsen, 2004) จึงส่งผลให้ ความหลากหลายทางพันธุกรรมลดลง (Stephan, 2019) นอกจากนี้จากการศึกษาที่ผ่านมายังพบว่า selective sweeps นี้มีผลอย่างมากต่อการ เปลี่ยนแปลงการกระจายตัวของตำแหน่งความ แปรปรวนบนยีนรวมถึงมีผลต่อการเพิ่มระยะ LD ระหว่างตำแหน่ง beneficial mutation อีกด้วย (Kim and Nielsen, 2004; Nimmakayala et al., 2014; Qian et al., 2014)

วัตถุประสงค์งานวิจัยนี้เป็นการวิเคราะห์ความ ไม่สมดุลของลิงเกจจากเครื่องหมาย SNP และ InDel ภายในยีน U-box domain-containing protein 35 (SaPUB35) ของอ้อยด้วยเทคโนโลยี PacBio SMRT sequencing เพื่อทำความเข้าใจในความไม่ สมดุลลิงเกจภายในยีนที่เกี่ยวข้องกับลักษณะ น้ำตาลในอ้อย อันเกิดจากการคัดเลือกในโครงการ ปรับปรุงพันธุ์อ้อย

2. วิธีการ

2.1 ตัวอย่างอ้อย และการสกัดดีเอ็นเอ

ตัวอย่างอ้อยที่ใช้ในการศึกษามีทั้งหมด 141 ตัวอย่าง (ตารางที่ 1) เป็นเชื้อพันธุกรรมนำเข้า (germplasm) จากแหล่งในทวีปต่าง ๆ ทั่วโลก จำนวน 68 ตัวอย่าง (ตารางที่ 1 ลำดับตัวอย่างที่ 1-68) และอ้อยพันธุ์ปรับปรุง (breeding line) ของประเทศไทยจำนวน 73 ตัวอย่าง (ตารางที่ 1 ลำดับตัวอย่างที่ 69-141) ปลูก ณ แปลงทดลองศูนย์นวัตกรรมและการวิจัยมิตรผล อำเภอภูเขียว จังหวัดชัยภูมิ สกัดดีเอ็นเอจากใบอ่อนของอ้อยด้วยวิธี CTAB (cetyltrimethylammonium bromide) ของ Gawel and Jarret (1991) ตรวจสอบคุณภาพของดีเอ็นเอด้วยวิธีอิเล็กโทรโฟรีซิสในเจลอะกาโรส และวัดค่าความเข้มข้นของดีเอ็นเอด้วยการวัดค่าการดูดกลืนแสงด้วย spectrophotometer

2.2 การออกแบบไพรเมอร์จากยีน SaPUB35

นำลำดับเบส cDNA ของยีน SaPUB35 มาเปรียบเทียบกับลำดับเบสคอนทิกของอ้อยพันธุ์ KK3 เพื่อทำนายตำแหน่งของเอกซอนและอินตรอน แล้วออกแบบไพรเมอร์ด้วยซอฟต์แวร์ Primer3 (Rozen and Skaletsky, 2000) ร่วมกับโปรแกรม Multiple Primer Analyzer (Breslauer *et al.*, 1986) แบ่งการออกแบบไพรเมอร์เพื่อเพิ่มปริมาณดีเอ็นเอ 2 ชิ้น (fragment) คือไพรเมอร์ SaPUB35_1 ประกอบด้วย ไพรเมอร์เส้น forward ขนาด 18 เบส (5'-TTTCCGTTCCGTTACTGC-3') และไพรเมอร์เส้น reverse ขนาด 20 เบส (5'-ACTTCTAGT GCTAGGTGAAG- 3') และ SaPUB35_2 ประกอบด้วยไพรเมอร์เส้น forward ขนาด 20 เบส (5'-AAACGCCTCGA TCCATATCC-3') และไพร

เมอร์เส้น reverse ขนาด 19 เบส (5'-CAATCCGAGGACTGAACGA-3') โดยขนาดผลผลิตพีซีอาร์ของไพรเมอร์ SaPUB35_1 และ SaPUB35_2 จากการคาดคะเนมีขนาดประมาณ 2903 เบส และ 2881 เบส ตามลำดับ ซึ่งครอบคลุมความยาวของยีน 5731 คู่เบส โดยไพรเมอร์ทั้ง 2 คู่มีตำแหน่งเหลื่อมซ้อนกันระหว่างไพรเมอร์เส้น reverse ของคู่แรกกับไพรเมอร์เส้น forward ของคู่ที่สอง 53 คู่เบส อุณหภูมิ Tm อยู่ระหว่าง 50-65 °C และ Tm ของไพรเมอร์เส้น reverse และเส้น forward ต่างกันไม่เกิน 1-2 °C

2.3 การเพิ่มปริมาณดีเอ็นเอและการทำให้บริสุทธิ์

เพิ่มปริมาณดีเอ็นเอด้วยการทำพีซีอาร์สองรอบ ในการทำพีซีอาร์รอบแรกนั้นใช้ไพรเมอร์จำเพาะที่ออกแบบมาจากยีน SaPUB35 ซึ่งติดอยู่กับลำดับเบส M13 มีปริมาตร 10 μ l ประกอบด้วยดีเอ็นเอต้นแบบ 10 ng, 5x Phusion HF buffer (MgCl₂ 7.5 mM), dNTPs 2 mM, คู่ไพรเมอร์อย่างละ 0.005 mM และ 0.2 units Phusion taq ใช้สถานะในการทำพีซีอาร์ที่อุณหภูมิ 94 °C 2 นาที ตามด้วย 94 °C 20 วินาที 30 รอบ อุณหภูมิ (56-63 °C) 20 วินาที 72 °C 1.30 นาที และสุดท้ายที่อุณหภูมิ 72 °C 5 นาทีและพีซีอาร์รอบที่สองเป็นการติดลำดับเบสบาร์โคดเพื่อใช้แยกความแตกต่างระหว่างตัวอย่างโดยใช้ไพรเมอร์ที่มีลำดับเบสคู่สมกับลำดับเบส M13 ซึ่งมีลำดับเบสบาร์โคดที่ติดอยู่ที่ปลาย M13 มีปริมาตร 50 μ l ประกอบด้วยดีเอ็นเอต้นแบบที่เจือจางจากผลผลิตพีซีอาร์รอบแรกในอัตราส่วน 1:50, 5x Phusion HF buffer (MgCl₂ 7.5 mM), MgCl₂ 75 mM, dNTPs 10 mM, คู่ไพรเมอร์อย่างละ 5 mM และ 1 units Phusion taq สถานะในการทำพีซีอาร์ที่อุณหภูมิ 98 °C 2 นาที ตามด้วย 98 °C 30 วินาที จำนวน 4 รอบ อุณหภูมิ 63 °C 15

วินาที 72 °C 45 วินาที 98 °C 30 วินาที อีกจำนวน 6 รอบ ตามด้วยอุณหภูมิ 66 °C 15 วินาที 72 °C 45 วินาที และสุดท้ายที่อุณหภูมิ 72 °C 5 นาที แล้ววิเคราะห์ผลผลิตพีซีอาร์ที่ได้จากการทำปฏิกิริยา ลูกโซ่พอลิเมอไรเซชัน 2 รอบเบื้องต้นด้วยวิธีอิเล็กโทรโฟเรซิสในเจลอะกาโรสความเข้มข้น 2% เพื่อตรวจสอบขนาดของชิ้นดีเอ็นเอที่ต้องการตามที่ได้คาดคะเนจากการออกแบบคู่ไพรเมอร์ จากนั้นนำผลผลิตพีซีอาร์ที่ได้จากการทำพีซีอาร์ครั้งที่ 2 มาทำให้บริสุทธิ์ด้วย AMPure XP purification kit (Beckman Coulter, Brea CA, USA) และวัดความเข้มข้นด้วย Qubit® 2.0 Fluorometer โดยใช้ชุดคิท Qubit® dsDNA BR Assay (Invitrogen, Oregon, USA) จากนั้นผสมรวมผลผลิตพีซีอาร์จากทุกตัวอย่างแยกเป็นชั้นที่ 1 (คลั่งดีเอ็นเอ 1) และชั้นที่ 2 (คลั่งดีเอ็นเอ 2) ให้ได้ความเข้มข้นสุดท้ายของแต่ละคลั่งดีเอ็นเอเท่ากับ 50 ng ปริมาตร 30 µl เพื่อเตรียมสร้างคลัง SMRTbell และวิเคราะห์ลำดับเบสด้วยเครื่อง PacBio® RS High Resolution Genetic Analyzer

2.4 การค้นหาเครื่องหมาย SNP และ InDel

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

2.5 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 ประชากรคือประชากรตัวอย่างที่เป็นเชื้อพันธุ์กรรมนำเข้าและประชากรตัวอย่างที่เป็นสายพันธุ์ปรับปรุง

3. ผลการวิจัยและวิจารณ์ผล

3.1 การวิเคราะห์หาลำดับเบสด้วยเทคโนโลยี PacBio SMRT sequencing

ผลจากการวิเคราะห์หาลำดับเบสของยีน SaPUB35 ด้วยเทคโนโลยี PacBio SMRT sequencing ในอ้อยทั้ง 141 ตัวอย่าง พบว่าสามารถวิเคราะห์ลำดับเบสสายยาว (read) ที่มีความยาวตั้งแต่ 2000 เบสขึ้นไปได้จำนวน 17048 reads และทุก read สามารถ map บนจีโนมอ้างอิงได้

Table 1 Lists of the 141 sugarcane varieties.

| No. | Varieties |
|-----|-----------|-----|-----------|-----|-----------|-----|-----------|-----|-----------|
|-----|-----------|-----|-----------|-----|-----------|-----|-----------|-----|-----------|

| | | | | | | | | | |
|----|-----------|----|-----------|----|----------|-----|-----------|-----|-------------|
| 1 | B41721 | 29 | GT29 | 57 | Q146 | 85 | KPK98-40 | 113 | MPT09-296 |
| 2 | B4744 | 30 | IAC49_131 | 58 | Q208 | 86 | KPS01-12 | 114 | MPT10-54 |
| 3 | B76718 | 31 | IAC51_205 | 59 | Q47 | 87 | KPS01-25 | 115 | MPT11-181 |
| 4 | BL4 | 32 | IAC52_156 | 60 | Q77 | 88 | LK92-11 | 116 | MPT12-141 |
| 5 | BO14 | 33 | IRK67_1 | 61 | Q81 | 89 | MP1 | 117 | MPT12-1782 |
| 6 | BO24 | 34 | Keawgean | 62 | Q83 | 90 | MP2 | 118 | MPT13-118 |
| 7 | CAC57_23 | 35 | KWT3 | 63 | R397 | 91 | MP3 | 119 | MPT14-1-172 |
| 8 | CO1287 | 36 | LF48_8077 | 64 | ROC1 | 92 | MPT02-458 | 120 | MPT14-1-180 |
| 9 | CO290 | 37 | LF76_2300 | 65 | ROC22 | 93 | MPT02-612 | 121 | MPT14-1-468 |
| 10 | CO331 | 38 | LF78_960 | 66 | ROC7 | 94 | MPT02-665 | 122 | MPT14-1-504 |
| 11 | CP45_150 | 39 | LF79_2964 | 67 | TUC74_6 | 95 | MPT02-669 | 123 | MPT14-1-546 |
| 12 | CP61_39 | 40 | M124/59 | 68 | YASAWA | 96 | MPT03-166 | 124 | MPT14-1-857 |
| 13 | CP63_306 | 41 | M13/58 | 69 | BIOTECH1 | 97 | MPT04-169 | 125 | MPT14-1-892 |
| 14 | CP72_120 | 42 | M147/158 | 70 | BIOTECH2 | 98 | MPT04-467 | 126 | MPT14-1-976 |
| 15 | CP75_330 | 43 | MACKAY | 71 | BIOTECH3 | 99 | MPT05-187 | 127 | MPT14-2-662 |
| 16 | CP76_340 | 44 | MANA | 72 | BIOTECH4 | 100 | MPT06-144 | 128 | MPT14-5-216 |
| 17 | D158 | 45 | MOSSMAN | 73 | BIOTECH5 | 101 | MPT06-166 | 129 | MPT14-5-239 |
| 18 | DB7160 | 46 | MY55_14 | 74 | BIOTECH6 | 102 | MPT06-171 | 130 | MPT14-6-63 |
| 19 | Ebene1/37 | 47 | N50_211 | 75 | CN1 | 103 | MPT06-344 | 131 | MPT14-9-133 |
| 20 | Eros | 48 | NCO310 | 76 | K84-200 | 104 | MPT06-367 | 132 | MPT99-1444 |
| 21 | F140 | 49 | NCO382 | 77 | K92-80 | 105 | MPT07-1 | 133 | SP50 |
| 22 | F148 | 50 | PHIL66_07 | 78 | K93-219 | 106 | MPT07-152 | 134 | SP80 |
| 23 | F152 | 51 | PINDAR | 79 | K95-84 | 107 | MPT07-71 | 135 | TBY20-2248 |
| 24 | F153 | 52 | PL310 | 80 | K99-72 | 108 | MPT08-191 | 136 | UT1 |
| 25 | F170 | 53 | PR3067 | 81 | KK3 | 109 | MPT08-3 | 137 | UT12 |
| 26 | F177 | 54 | Q100 | 82 | KKU99-01 | 110 | MPT08-50 | 138 | UT3 |
| 27 | FIJI105 | 55 | Q115 | 83 | KKU99-03 | 111 | MPT09-118 | 139 | UT5 |
| 28 | GALOA | 56 | Q117 | 84 | KKU99-06 | 112 | MPT09-204 | 140 | UT6 |
| | | | | | | | | 141 | UT8 |

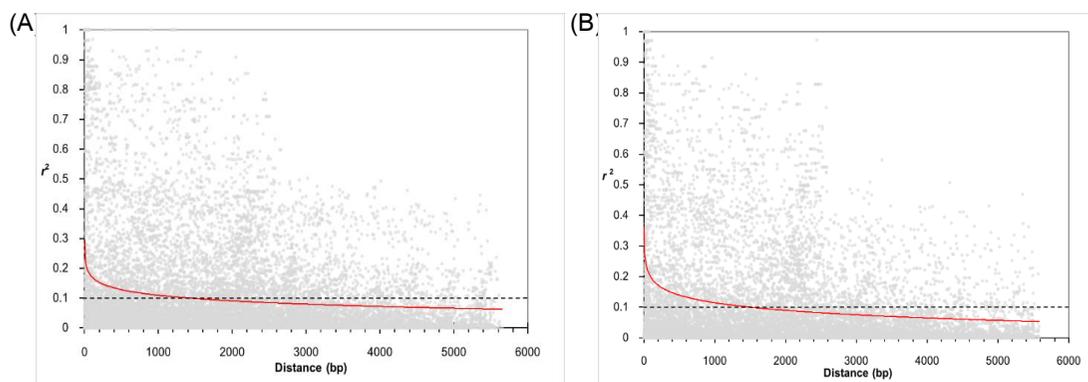


Figure 1 r^2 values of pairwise comparisons between polymorphic sites plotted against their physical distances (bp) of germplasm population (A) and breeding lines population (B). Logarithmic regression curves are in red. LD Decay is considered below $r^2 = 0.1$ threshold.

และค่าความยาวเฉลี่ยของ read ที่ได้จากไพร์เมอร์ SaPUB35_1 และ SaPUB35_2 มีขนาดเท่ากับ 2887 และ 2880 เบส ตามลำดับ ค่าจำนวนซ้ำของข้อมูลเบสตำแหน่งเดิมหรือค่า coverage เท่ากับ 26x โดยมีค่า read depth ตั้งแต่ 12-38 เนื่องจากความสามารถในการวิเคราะห์ลำดับเบสสายยาวของเทคโนโลยี PacBio SMRT sequencing จึงมีงานวิจัยก่อนหน้าใช้เทคโนโลยีนี้วิเคราะห์ลำดับเบสในอ้อยเพื่อวัตถุประสงค์ด้านต่าง ๆ ทั้งการประกอบลำดับเบสจีโนมไมโทคอนเดรียของอ้อยพันธุ์ขอนแก่น 3 โดย Shearman *et al.* (2016) รวมถึงการวิเคราะห์หาลำดับเบสของโคลน bacteria artificial chromosome ของ minimum tilling path ในอ้อยพันธุ์ R570 โดย Garsmeur *et al.* (2018) และยังถูกใช้ในการศึกษาทรานสคริปโตมโดย Hoang *et al.* (2017) และ Piriyaongsa *et al.* (2018) อีกด้วย จะเห็นได้ว่าการวิเคราะห์ลำดับเบสด้วยเทคโนโลยี PacBio SMRT sequencing มีประสิทธิภาพในการวิเคราะห์ลำดับเบสในอ้อยซึ่งจัดเป็นพืชที่มีจีโนมขนาดใหญ่และมีความซับซ้อนมากชนิดหนึ่ง

3.2 ความแปรปรวนของเครื่องหมาย SNP และ InDel

จากการค้นหาเครื่องหมาย SNP และ InDel บนยีน SaPUB35 ที่ผ่านการคัดกรอง พบว่ามีเครื่องหมาย SNP ที่ผ่านการคัดกรองจำนวน 126 SNPs มีตำแหน่งอยู่บนเอกซอนจำนวน 42 SNPs และอยู่บนอินตรอนจำนวน 84 SNPs ขณะที่เครื่องหมาย InDel นั้น ผ่านการคัดกรองจำนวน 50 InDels มีตำแหน่งอยู่บนบริเวณเอกซอนและอินตรอนจำนวน 15 และ 35 InDels ตามลำดับ จะเห็นได้ว่าเครื่องหมายทั้งสองชนิดมีตำแหน่งบนบริเวณอินตรอนมากกว่าเอกซอน สอดคล้องกับการศึกษาก่อนหน้าของ McIntyre *et al.* (2006) และ Song *et al.* (2016) ในการระบุเครื่องหมาย SNP บนบริเวณอินตรอนได้มากกว่าบริเวณเอกซอน เนื่องจากเกิดการกลายของลำดับเบสสะสมมากบนบริเวณอินตรอนในระหว่างการเกิดวิวัฒนาการของยีน ส่งผลให้บริเวณอินตรอนมีความแปรปรวนของลำดับเบสมากกว่าบริเวณเอกซอน หรือกล่าวได้ว่าบริเวณอินตรอนมีความอนุรักษ์น้อยกว่าเมื่อเทียบกับบริเวณเอกซอน (Jaikishan *et al.*, 2015) เมื่อพิจารณาความถี่ของการพบเครื่องหมาย InDel ได้

ทุก ๆ 115 เบส และพบ SNP โดยเฉลี่ยทุก ๆ 46 เบส ซึ่งใกล้เคียงกับการศึกษาของ McIntyre *et al.* (2006) ที่ศึกษาและระบุแอลลีลใน 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 เบส ของลำดับเบสที่มีการถอดรหัส

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

จากการวิเคราะห์ LD decay เปรียบเทียบระหว่างตัวอย่างที่เป็นประชากรกลุ่มเชื้อพันธุกรรมนำเข้าและอ้อยพันธุ์ปรับปรุง พบว่าประชากรเชื้อพันธุกรรมนำเข้าจำนวน 68 ตัวอย่างมีเครื่องหมาย SNP และ InDel ที่ผ่านการคัดกรองจำนวน 184 เครื่องหมาย และในประชากรอ้อยที่เป็นพันธุ์ปรับปรุงจำนวน 73 ตัวอย่าง มีเครื่องหมายที่ผ่านการคัดกรองจำนวน 163 เครื่องหมาย จะเห็นได้ว่าจำนวนเครื่องหมาย SNP และ InDel ที่ระบุได้ในประชากรอ้อยพันธุ์ปรับปรุงมีจำนวนน้อยกว่าในประชากรที่เป็นเชื้อพันธุกรรมนำเข้า แสดงให้เห็นถึงความหลากหลายทางพันธุกรรมที่ลดลงในประชากรอ้อยพันธุ์ปรับปรุง (Stephan, 2019) อันเป็นผลจากการคัดเลือกโดยนักปรับปรุงพันธุ์เพื่อพัฒนาอ้อยสายพันธุ์ดี จากแผนภาพการกระจายตัวของค่า r^2 ระหว่างคู่เครื่องหมายกับระยะห่างทางกายภาพด้วยสมการแบบไม่เป็นเชิงเส้นตรง (ภาพที่ 2) พบว่าในประชากรเชื้อพันธุกรรมนำเข้า ค่า r^2 มีค่าลดลงมาจนถึง 0.1 ที่ระยะห่างทางกายภาพประมาณ 1400 bp (ภาพที่ 2A) ขณะที่ค่า r^2 ของประชากรพันธุ์ปรับปรุงลดลงถึง 0.1 ที่ระยะห่างทางกายภาพประมาณ 1500 bp (ภาพที่ 2B) จะเห็นได้ว่าระยะของความไม่สมดุลลิงเกจของยีน SaPUB35

ที่มีความเกี่ยวข้องกับลักษณะความหวานในอ้อยพันธุ์ปรับปรุงมีระยะที่กว้างกว่าเชื้อพันธุกรรมนำเข้า เนื่องจากประชากรอ้อยที่เป็นพันธุ์ปรับปรุงนั้นเกิดจากควบคุมการผสมพันธุ์ระหว่างอ้อยพันธุ์ดีที่ผ่านการคัดเลือกแล้วและคัดเลือกลูกผสมเพื่อให้ได้อ้อยที่มีลักษณะตรงตามความต้องการ การคัดเลือกนี้จึงเป็นการลดความแปรปรวนทางพันธุกรรมและตรึงการกลายที่มีผลดี (beneficial mutation) ต่อลักษณะที่ต้องการไว้ (Kim and Nielsen, 2004) กระบวนการคัดเลือกนี้เรียกว่า selective sweep เป็นการคัดเลือกแบบกวาดเอาทั้ง beneficial mutation และ แอลลีลที่ชนาบบ้างตำแหน่ง 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 เช่นเดียวกับการศึกษาในข้าวบาร์เลย์ระหว่างประชากรพันธุ์ป่าและพันธุ์ปลูก พบว่าประชากรที่เป็นพันธุ์ป่านั้นมี LD แคบกว่าพันธุ์ปลูก เนื่องจากรูปแบบของ LD ในประชากรที่เป็นพันธุ์ป่านั้นขึ้นอยู่กับประวัติการเกิดรีคอมบิเนชันที่ยาวนานกว่าพันธุ์ปลูก (Caldwell *et al.*, 2006) จากการศึกษาในข้าวโพดพบว่าพันธุ์ที่เป็นเชื้อพันธุกรรมซึ่งมีความหลากหลายทางพันธุกรรมมีระยะ LD ที่แคบกว่าข้าวโพดสายพันธุ์ปรับปรุงเพื่อการค้า (Inghelandt *et al.*, 2011)

4. สรุป

งานวิจัยนี้ได้แสดงให้เห็นว่ายีน *SaPUB35* ซึ่งเป็นยีนที่เกี่ยวข้องกับหลายลักษณะที่สำคัญของอ้อยมีความหลากหลายทางพันธุกรรมที่ลดลงในประชากรอ้อยพันธุ์ปรับปรุง และมีระยะของความไม่สมดุลลิงเกจ กว้างกว่าประชากรอ้อยนำเข้า อาจเป็นผลเนื่องจากอิทธิพลของ selective sweep จากกระบวนการเลือกจับคู่ผสมระหว่างอ้อยพันธุ์ที่ได้รับการคัดเลือกแล้ว และคัดเลือกลูกผสมที่มีลักษณะที่ต้องการ ความรู้นี้จะมีประโยชน์สำหรับการวางแผนกลยุทธ์ในการปรับปรุงพันธุ์อ้อยเพื่อป้องกันการเสื่อมลงของความหลากหลายทางพันธุกรรมในโครงการปรับปรุงพันธุ์อ้อย

5. กิตติกรรมประกาศ

ขอขอบคุณทุนบัณฑิตศึกษาดีเพื่อศึกษาต่อระดับบัณฑิตศึกษา คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ ตามบันทึกข้อตกลงเลขที่ ทบ 09/2560

6. References

Abdel-Ghany, S. E., Hamilton, M., Jacobi, J. L., Ngam, P., Devitt, N., Schilkey, F., Ben-Hur, A., and Reddy, A. S., 2016, A survey of the sorghum transcriptome using single-molecule long reads, *Nature communications*. 7(1). 1-11.

Barrett, J. C., Fry, B., Maller, J., and Daly, M. J., 2005, Haploview: analysis and visualization of LD and haplotype maps, *Bioinformatics*. 21(2). 263-265.

Breslauer, K. J., Frank, R., Blöcker, H., and Marky, L. A., 1986, Predicting DNA duplex stability from the base sequence,

Proceedings of the National Academy of Sciences. 83(11). 3746-3750.

Caldwell, K. S., Russell, J., Langridge, P., and Powell, W. , 2006, Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*, *Genetics*. 172(1). 557-567.

Chen, L., Deng, R., Liu, G., Jin, J., Wu, J., and 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(546). 1-19.

Cordeiro, G. M., Elliott, F., McIntyre, C. L., Casu, R. E. , and Henry, R. J. , 2006, Characterisation of single nucleotide polymorphisms in sugarcane ESTs, *Theoretical and Applied Genetics*. 113(2). 331-343.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., and Bettman, B., 2009, Real-time DNA sequencing from single polymerase molecules, *Science*. 323(5910). 133-138.

Garsmeur, O. , Droc, G. , Antonise, R. , Grimwood, J. , Potier, B. , Aitken, K. , Jenkins, J. , Martin, G. , Charron, C. , Catherine, H. , Laurent, C. , Nabila, Y. , Adam, H., David, S., Yesesri, C., Avinash, S., Andrzej, K., Agnes, C., Marie-Anne Van, S., Kankshita, S., Christopher, T., Hélène, B., Blake, S., Jean Christophe, G., Edwin van der, V., Robert, H., Jeremy, S., and Angélique, D. H. , 2018, A mosaic

- monoploid reference sequence for the highly complex genome of sugarcane, *Nature communications*. 9(1). 1-10.
- Gawel, N. , and Jarret, R. , 1991, A modified CTAB DNA extraction procedure for *Musa* and *Ipomoea*, *Plant molecular biology reporter*. 9(3). 262-266.
- Hoang, N. V., Furtado, A., O’Keeffe, A. J., Botha, F. C., and Henry, R. J., 2017, Association of gene expression with biomass content and composition in sugarcane, *PLoS One*. 12(8). 1-31.
- Inghelandt, D. V., Reif, J. C., Dhillon, B. S., Flament, P., and 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. , and 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.
- Jannoo , N., Grivet , L., Chantret , N., Garsmeur , O., Glaszmann , J. C., Arruda , P., and 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.
- Khanbo, S., 2020, Association Mapping of Sugar Related Traits in Sugarcane (*Saccharum* spp.). Doctoral Dissertation, Thammasat University, Pathum Thani.
- Kim, Y. , and Nielsen, R. , 2004, Linkage disequilibrium as a signature of selective sweeps, *Genetics*. 167(3). 1513-1524.
- Lu, Y., Shah, T., Hao, Z., Taba, S., Zhang, S., Gao, S., Liu, J., Cao, M., Wang, J., and 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., Jackson, M., Cordeiro, G. M., Amouyal, O., Hermann, S., Aitken, F. S., Elliott, F., Henry, R. J., Casu, R. E., and 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., Kernytzky, A., Garimella, K., Altshuler, D., Gabriel, S., and Daly, M., 2010, The Genome Analysis Toolkit: a Map Reduce framework for analyzing next-generation DNA sequencing data, *Genome research*. 20(9). 1297-1303.
- Nimmakayala, P., Levi, A., Abburi, L., Abburi, V. L., Tomason, Y. R., Saminathan, T., Vajja, V. G. , Malkaram, S. , Reddy, R. , and 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.
- Piriyaopongsa, J., Kaewprommal, P., Vaiwsri, S., Anuntakarun, S., Wirojsirasak, W., Punpee, P., Klomsa-ard, P., Shaw, P. J., Pootakham, W., and Yoocha, T., 2018, Uncovering full-length transcript isoforms of sugarcane cultivar Khon Kaen 3 using single-molecule long-read sequencing, PeerJ. 6(5818). 1-23.
- Qian, L., Qian, W., and 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., and 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., Pauquet, J., Butterfield, M., D'Hont, A., and Glaszmann, J. C., 2008, Analysis of genome-wide linkage disequilibrium in the highly polyploid sugarcane, Theoretical and Applied Genetics. 116(5). 701-714.
- Rozen, S., and Skaletsky, H., 2000, Primer3 on the WWW for general users and for biologist programmers pp. 365-386. In Bioinformatics methods and protocols
- Ryu, M. Y., Cho, S. K., Hong, Y., Kim, J., Kim, J. H., Kim, G. M., Chen, Y.-J., Knoch, E., Møller, B. L., and 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.
- Sela, H., Loutre, C., Keller, B., Schulman, A., Nevo, E., Korol, A., and 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., and 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.
- Song, J., Yang, X., Resende Jr, M. F., Neves, L. G., Todd, J., Zhang, J., Comstock, J. C., and 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.
- Wang, B., Tseng, E., Regulski, M., Clark, T. A., Hon, T., Jiao, Y., Lu, Z., Olson, A., Stein, J. C., and Ware, D., 2016, Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing, Nature communications. 7(1). 1-13.
- Wang, W., Wang, W., Wu, Y., Li, Q., Zhang, G., Shi, R., Yang, J., Wang, Y., and 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., and 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.
- Zhang, W., Wang, L., Wang, Y., Wang, Y., and 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., and 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., and Yu, J., 2008, Status and prospects of association mapping in plants, *The Plant Genome*. 1(1). 5-20.
- Zia, M. A. B., Demirel, U., Nadeem, M. A., and Caliskan, M. E., 2020, Genome-wide association study identifies various loci underlying agronomic and morphological traits in diversified potato panel, *Physiology and Molecular Biology of Plants*. 26. 1003-1020.