



รายงานวิจัยฉบับสมบูรณ์

โครงการ การวิเคราะห์หน้าที่ของโปรตีนแฟมิลี DUF642
บนผนังเซลล์ต่อการเจริญและพัฒนาของพืช

โดย รศ. ดร. ศุภชัย วุฒิพงศ์ชัยกิจ

พฤษภาคม พ.ศ. 2563

รายงานวิจัยฉบับสมบูรณ์

โครงการ การวิเคราะห์หน้าที่ของโปรตีนแฟมิลี DUF642
บนผนังเซลล์ต่อการเจริญและพัฒนาของพืช

รศ. ดร. ศุภชัย วุฒิพงศ์ชัยกิจ
มหาวิทยาลัยเกษตรศาสตร์

สนับสนุนโดยสำนักงานกองทุนสนับสนุนการวิจัย
และ มหาวิทยาลัยเกษตรศาสตร์

(ความเห็นในรายงานนี้เป็นของผู้วิจัย สกว. และมหาวิทยาลัยเกษตรศาสตร์ไม่
จำเป็นต้องเห็นด้วยเสมอไป)

กิตติกรรมประกาศ

ขอขอบคุณสำนักงานกองทุนสนับสนุนการวิจัย (สกว.) และ มหาวิทยาลัยเกษตรศาสตร์ ที่ให้การสนับสนุนทุน
ในการดำเนินงานโครงการวิจัยนี้

ขอขอบคุณ Prof. Simon McQueen-Mason, University of York แห่ง สหราชอาณาจักร
และ

Prof. George Bassel, University of Warwick แห่ง สหราชอาณาจักร
ที่สนับสนุนโครงการวิจัยและให้โอกาสในการดำเนินการวิจัย ณ ห้องแล็บของท่านทั้งสอง

ขอขอบคุณ ดร. อนงค์ภัทร สุทรางกุล และ ผศ. ดร. ภััสสร วรรณพินิจ ภาควิชาพันธุศาสตร์ คณะวิทยาศาสตร์
มหาวิทยาลัยเกษตรศาสตร์ ที่ช่วยสนับสนุนงานด้านเทคนิคและการดำเนินงานในห้องแล็บ

ขอขอบคุณนิสิตและลูกจ้างนักวิจัยที่ช่วยดำเนินงานในโครงการวิจัยนี้

Project Code: RSA6080031

Project Title: การวิเคราะห์หน้าที่ของโปรตีนแฟมิลี DUF642 บนผนังเซลล์ต่อการเจริญและพัฒนาของพืช

Investigator: นายศุภชัย วุฒิพงศ์ชัยกิจ มหาวิทยาลัยเกษตรศาสตร์

E-mail Address: supachai.v@ku.th

Project Period: พฤษภาคม 2560 – พฤษภาคม 2563

บทคัดย่อ

ผนังเซลล์มีหน้าที่สำคัญต่อการมีชีวิตของเซลล์พืช โครงสร้างประกอบด้วยโพลีแซคคาไรด์หลากชนิด และมีโปรตีนและเอนไซม์ทำหน้าที่ควบคุมการทำงาน โปรตีนแฟมิลี Domain of Unknown Function 642 (DUF642) เป็นกลุ่มโปรตีนที่ไม่ทราบหน้าที่ และไม่สามารถคาดเดาหน้าที่จากลำดับกรดอะมิโนได้ แต่พบมากในผนังเซลล์ของเนื้อเยื่อพืชที่กำลังมีการเจริญเติบโต วัตถุประสงค์ของงานวิจัยนี้ คือ การศึกษาหน้าที่ของโปรตีนแฟมิลี DUF642 ในอะราบิดอปซิสและการทำงานในผนังเซลล์พืช โปรตีนแฟมิลี DUF642 ถอดรหัสจากยีน 10 ตำแหน่ง ในอะราบิดอปซิส จากการตรวจสอบการแสดงออกของยีนทั้ง 10 ด้วย Promoter-GUS fusion แสดงให้เห็นว่ายีนทั้ง 10 มีการแสดงออกในต้นอะราบิดอปซิส และน่าจะมีการทำงานทดแทนกันแบบ functional redundancy จากการตรวจสอบการจับระหว่างโปรตีน DUF642 กับโพลีแซคคาไรด์ชนิดต่าง ๆ ที่เป็นองค์ประกอบของผนังเซลล์พืชทั้งหมด 16 ชนิด โดยใช้วิธี ELISA พบว่าโปรตีน DUF642 สามารถจับกับโพลีแซคคาไรด์ทั้งกลุ่มเซลลูโลส เฮมิเซลลูโลส และเพกติน โดยโปรตีนแต่ละตัวมีความจำเพาะที่แตกต่างกัน เมื่อตรวจสอบผลการสูญเสียหน้าที่ของยีนด้วย T-DNA insertion หลายตำแหน่งร่วมกันถึง 7 ตำแหน่ง โดยตรวจสอบจากสายพันธุ์กลายทั้งหมด 70 สายพันธุ์ พบว่าสายพันธุ์กลายมีการเจริญเติบโตลดลงเมื่อทดสอบกับ $AlCl_3$ ที่ส่งผลยับยั้งการทำงานของเพกติน แต่ไม่มีการเปลี่ยนแปลงที่ชัดเจนในด้านการเจริญเติบโตหรือสภาวะทดสอบความเครียดต่าง ๆ ทั้งนี้สายพันธุ์กลายที่ทดสอบยังมียีนอีกอย่างน้อย 3 ตำแหน่งที่ยังทำงานทดแทน เมื่อตรวจสอบการเข้าจับกับโปรตีนอื่น ๆ ในผนังเซลล์ด้วยเทคนิค co-immunoprecipitation และ yeast two hybrids พบว่าโปรตีน DUF642 เข้าจับได้อย่างจำเพาะกับโปรตีน Expansins ที่ควบคุมการขยายขนาดของผนังเซลล์ จากการทดลองสามารถบ่งชี้ได้ว่าโปรตีน DUF642 น่าจะทำหน้าที่เป็นตัวกลางหรือเกี่ยวข้องในการทำงานระหว่างโปรตีน Expansins และโพลีแซคคาไรด์ในผนังเซลล์ ข้อมูลวิจัยที่ได้จากโครงการนี้จะช่วยเพิ่มความเข้าใจของกระบวนการและการทำงานของผนังเซลล์พืชที่เกี่ยวข้องกับการเจริญของเซลล์และต้นพืช

คำสำคัญ: อะราบิดอปซิส, ผนังเซลล์พืช, โดเมนไม่ทราบหน้าที่, การเจริญของพืช, หน้าที่ของยีน

Abstract

Plant cell walls play important roles in the life of plant cells. Cell walls consist of various types of polysaccharides and proteins, which regulates the dynamic of the walls. Domain of unknown function family 642 (DUF642) is an unknown function protein found abundantly in the wall of active growing cells. The objective of this work is to characterize the function of DUF642 family and its role in the plant cell walls in Arabidopsis. DUF642 proteins are encoded from a family of 10 genes in Arabidopsis, and expression analysis of these genes by Promoter-GUS fusion showed different pattern of expression and suggested a highly functional redundancy among the 10 family members. Analysis of protein-polysaccharide binding by an ELISA based approach showed that DUF642s can bind to cellulose, hemicelluloses and pectins, in which each of the members has different binding specificity. Analysis of T-DNA insertion lines for simultaneously knockout up to 7 genes from in total 70 lines showed that the insertions only responded to AlCl_3 treatments, as a pectin inhibitor, with reduced root growth, but did not show changes in general seedling growth or treatments with stresses. It is important to note that there are at least 3 genes remained functioning in the mutant lines as to sustain the septuple knockouts. Protein interaction studies by co-immunoprecipitation and yeast two hybrids showed that DUF642 interact specifically with Expansins, which is known to control cell wall extension. The results indicate that DUF642 is likely to function as a mediator or being involve between Expansins and polysaccharides in the cell wall. Data obtained from this work will expand the understanding of plant cell walls, particularly the process that related to the cell and plant growth.

Keywords: Arabidopsis; plant cell wall; domain of unknown function; plant growth; gene function

บทนำ

ผนังเซลล์เป็นองค์ประกอบที่สำคัญและทำหน้าที่หลายอย่างในเซลล์พืช ในช่วงเวลาหลายปีที่ผ่านมา งานวิจัยของผนังเซลล์พืชได้รับความสนใจจากนักวิจัยหลากหลายสาขาเนื่องจากความสำคัญในด้านการผลิตชีวมวลและการผลิตพลังงานชีวภาพ โดยความรู้ที่ได้ทำมาช่วยในการประเมินศักยภาพและการวางแผนการใช้ชีวมวลที่ผลิตจากพืชชนิดต่าง ๆ ได้อย่างมีประสิทธิภาพมากขึ้น ผนังเซลล์ประกอบด้วยองค์ประกอบหลัก คือ โพลีแซคคาไรด์ และองค์ประกอบย่อย คือ โปรตีนชนิดต่าง ๆ ถึงแม้ว่าโครงสร้างของผนังเซลล์จะแข็งแรงและคงที่ แต่ในระดับเซลล์นั้นผนังเซลล์มีความเปลี่ยนแปลงอยู่ตลอดชีวิตของเซลล์ ทั้งการรักษาภาวะของเซลล์ การยืดขยาย ตลอดจนการเปลี่ยนแปลงรูปร่างเพื่อไปทำหน้าที่ต่าง ๆ โดยที่การเปลี่ยนแปลงเหล่านี้เกิดขึ้นโดยการควบคุมของเอนไซม์หลาย ๆ ชนิดที่อยู่ในผนังเซลล์นั่นเอง จากการศึกษาทางจีโนมิก โปรตีโอมิก และการแสดงออกของยีน ในพืชหลายชนิดได้แสดงให้เห็นว่า มากกว่า 15% ของโปรตีนที่มีการแสดงออกของในเซลล์พืชนั้นเป็นโปรตีนทำงานที่เกี่ยวข้องกับผนังเซลล์ ปัจจุบันโปรตีนโครงสร้างและเอนไซม์หลายชนิดที่ทำงานอยู่ในผนังเซลล์นั้นได้มีการศึกษาหน้าที่และการทำงานไปเพียงบางส่วนเท่านั้น ยังมีโปรตีนในผนังเซลล์อีกจำนวนมากที่ยังไม่ทราบถึงการทำงานและความสำคัญ ทั้งนี้ความเข้าใจเกี่ยวกับการทำงานของโปรตีนในผนังเซลล์ที่เกี่ยวข้องกับการเจริญและพัฒนาของพืชนั้นยังไม่มีเป็นที่เข้าใจมากนัก ความรู้และความเข้าใจที่มากขึ้นจะสามารถนำไปใช้ประโยชน์ของพืชทั้งด้านชีวมวลและการเกษตร

โปรตีนในผนังเซลล์นั้นจัดกลุ่มออกเป็น แฟมิลีตามลำดับกรดอะมิโน และโดเมนของโปรตีน อย่างไรก็ตามการจัดจำแนกนี้ไม่ได้บ่งบอกถึงหน้าที่การทำงานของโปรตีนนั้น ๆ ทั้งนี้มีโปรตีนจำนวนมากถูกจัดอยู่ในกลุ่ม known-unknown proteins ซึ่งหมายความว่าโปรตีนนั้น ๆ มีโครงสร้างเป็นที่รู้จักแต่ยังไม่ทราบการทำงานที่ เช่น เอนไซม์กลุ่ม kinase ซึ่งทราบว่าทำหน้าที่เติมหมู่ฟอสเฟต แต่ไม่รู้ว่ามีผลอย่างไรต่อพืช เป็นต้น และอีกกลุ่ม domains of unknown function (DUFs) ซึ่งมีโครงสร้างไม่เหมือนเอนไซม์ใด ๆ ฐานข้อมูลและไม่สามารถบ่งชี้การทำงานได้ โดยในฐานข้อมูล Pfam database มี DUF ที่จัดจำแนกแล้วจำนวนมากกว่า 3,000 แฟมิลี แฟมิลี DUF642 เป็นเป้าหมายของการศึกษาในข้อเสนอโครงการวิจัยชิ้นนี้เนื่องจากการศึกษาในเบื้องต้นก่อนหน้านี้ โดยการศึกษาเอนไซม์ expansin ซึ่งมีทำหน้าที่ส่งเสริมการยืดขยายของผนังเซลล์เพื่อรองรับการขยายตัวของเซลล์พืช โดเมนเอนไซม์นี้มีโปรตีนแฟมิลีขนาดใหญ่พบในพืชทุกชนิด และมีความสำคัญต่อการเจริญและพัฒนาของพืช จากการศึกษาเพื่อแยกโปรตีน expansin ออกจากผนังเซลล์พบว่ามีแถบโปรตีนที่มักตรวจพบด้วย และเมื่อนำแถบโปรตีนนี้ไปหาลำดับกรดอะมิโนพบว่ามีลำดับเหมือนกับโปรตีนแฟมิลี DUF642 ข้อมูลนี้แสดงให้เห็นว่าโปรตีนแฟมิลี DUF642 นี้่าจะมีการทำงานในผนังเซลล์ที่เกี่ยวข้องกับ expansin และการเจริญของพืช จากข้อมูลของลำดับเบสและลำดับกรดอะมิโนโปรตีนแฟมิลี DUF642 ในอะราบิโดพซิสนั้นมีจำนวนทั้งหมด 10 ยีน ในขณะนี้สมาชิกของแฟมิลีนี้บางตัวได้ถูกทดสอบแล้วในระดับโปรตีนถึงการจับกับผนังเซลล์พืชและการศึกษา gene knockout ในอะราบิโดพซิสที่เกี่ยวกับการเจริญเติบโตของราก อย่างไรก็ตามข้อมูลที่ได้นั้นยังอยู่ในระดับเบื้องต้น และหน้าที่และกลไกการทำงานของโปรตีนแฟมิลีนี้ก็ยังไม่มีที่ทราบ

ดังนั้นข้อเสนอวิจัยนี้จะเป็นการศึกษาหน้าที่และการทำงานของโปรตีนแฟมิลี DUF642 จำนวน 10 ยีนในอะราบิโดพซิส โดยการศึกษาจะประกอบไปด้วยการสร้างรีคอมบิแนนท์โปรตีนเพื่อการทดสอบการจับอย่างจำเพาะต่อโพลีแซคคาไรด์ของผนังเซลล์ชนิดต่าง ๆ และศึกษาการสร้างสายพันธุ์อะราบิโดพซิสที่ยับยั้งการทำงานของยีนแฟมิลีนี้โดยใช้ insertion line เพื่อใช้ศึกษาผลของการขาดหายไปของโปรตีนเหล่านี้ โปรตีน DUF642 ที่ได้จะนำมาใช้ทำ immuno-precipitation เพื่อหาโปรตีนชนิดอื่น ๆ ที่อาจจะทำงานร่วมกับโปรตีน DUF642 ในผนังเซลล์ ข้อมูลที่ได้จะรวบรวมเพื่อบ่งชี้หน้าที่และการทำงานของโปรตีนแฟมิลี DUF642 ในผนังเซลล์ของอะราบิโดพซิส

วัตถุประสงค์ของโครงการวิจัย

วัตถุประสงค์หลักของโครงการวิจัยนี้ คือ การศึกษาเพื่อเข้าใจหน้าที่และการทำงานของโปรตีนแฟมิลี DUF642 ที่ทำงานอยู่ในผนังเซลล์พืช โดยใช้อะราบิโดพซิสเป็นพืชต้นแบบในการศึกษา มีจุดประสงค์ย่อยดังต่อไปนี้

1. เพื่อตรวจสอบคุณสมบัติของโปรตีน DUF642 ในการเข้าจับโพลีแซคคาไรด์ของผนังเซลล์พืช
2. เพื่อตรวจหาหน้าที่การทำงานของโปรตีน DUF642 ในพืช โดยการศึกษาต้นอะราบิโดพซิสพันธุ์กลายหลายตำแหน่ง
3. เพื่อตรวจสอบคุณสมบัติการกระตุ้นการเจริญของพืชโดยการถ่ายยีน *DUF642* แบบชั่วคราว และการเติมโปรตีนโดยตรง
4. การตรวจสอบคุณสมบัติของโปรตีน DUF642 ในการเข้าจับและการทำงานร่วมกับโปรตีนอื่น ๆ ในผนังเซลล์พืช

การทบทวนวรรณกรรม

ผนังเซลล์พืช (Plant cell wall) และการทำหน้าที่เกี่ยวข้องกับการเจริญและพัฒนาของพืช

เซลล์พืชส่วนใหญ่จะต้องผ่านกระบวนการขยายขนาดของเซลล์ ซึ่งเป็นกระบวนการที่ควบคุมและบ่งชี้รูปร่าง การพัฒนา และการทำหน้าที่เซลล์นั้น ๆ โดยเซลล์พืชนั้นสามารถขยายได้ในอัตรา 10-20% ต่อชั่วโมง ในขณะที่กำลังมีการเจริญเติบโต และในขณะเดียวกันผนังเซลล์พืชก็ต้องยืดขยายเพื่อรองรับการขยายขนาดของเซลล์ โดยต้องคงสภาวะที่สมดุลระหว่างโครงสร้างที่แข็งแกร่งกับการยืดขยายตัว ทั้งนี้จากการศึกษาการทำงานของผนังเซลล์พบว่าโครงสร้างของผนังเซลล์นี้เองที่เป็นตัวควบคุมและกำหนดการขยายตัวของเซลล์ และ ควบคุมการเจริญและพัฒนาของพืช (Cosgrove, 1999) ผนังเซลล์เป็นตัวกลางระหว่างเซลล์พืช มีความสำคัญต่อการเชื่อมต่อโครงสร้างและการส่งสัญญาณต่าง ๆ ระหว่างเซลล์ (Wolf et al., 2012) ทั้งนี้โพลีเมอร์ที่เป็นองค์ประกอบของผนังเซลล์จะทำหน้าที่เป็นโมเลกุลตัวส่งสัญญาณ การส่งสัญญาณระหว่างเซลล์นี้มีความสำคัญต่อชีวิตของพืช เช่น การ differentiation เป็นเนื้อเยื่อต่าง ๆ การตายของเซลล์ และการตอบสนองต่อ pathogens และ symbionts (Baker et al., 1997).

ถึงแม้ว่าผนังเซลล์เป็นโครงสร้างที่ให้ความแข็งแรงและบ่งชี้รูปร่างของเซลล์พืช แต่ในตัวโครงสร้างเองนั้นสามารถบิด เบน โอนตาม หรือยืดขยายเพื่อรองรับการแบ่งเซลล์ได้ โดยการขยายของเซลล์พืชนั้นเกิดจากแรงดัน turgor ภายในของเซลล์และการตอบสนองของผนังเซลล์ร่วมกัน และส่วนผนังเซลล์เองเป็นตัวควบคุมอัตราและทิศทางการขยายของเซลล์ ทั้งนี้กระบวนการต่าง ๆ ในการเปลี่ยนแปลงของผนังเซลล์นั้นมีการควบคุมผ่านเอนไซม์หลายชนิด ตัวอย่างที่เป็นไคนามิคของผนังเซลล์ที่สามารถเห็นได้อย่างชัดเจนคือ การยืดขยายของไฮโปคอติลอย่างรวดเร็วของต้นอ่อนในสภาวะมืดซึ่งเมื่อมีแสงกระบวนการยืดขยายอย่างรวดเร็วนี้จะหยุดลงภายใน 1 นาที (Spalding and Cosgrove, 1992) กระบวนการหยุดนี้เกิดจากการแข็งขึ้นอย่างรวดเร็วของผนังเซลล์โดยการทำงานของเอนไซม์กลุ่ม cell wall loosening/rigidifying enzymes ซึ่งถูกกระตุ้นได้เมื่อมีแสง เอนไซม์ expansin และ xyloglucan endo-transglycosylase/hydrolase (XTHs) เป็นเอนไซม์สองตัวหลักที่ควบคุมผนังเซลล์ (Van Sandt et al., 2007)

ตั้งแต่ปี ค.ศ. 1665 เมื่อโรเบิร์ต ฮุก ส่องได้กล้องจุลทรรศน์แล้วพบว่าเนื้อเยื่อพืชประกอบด้วยเซลล์เป็นล้าน ๆ เซลล์ นักวิทยาศาสตร์ส่วนใหญ่หันจุดสนใจไปยังองค์ประกอบภายในของเซลล์มากกว่าผนังเซลล์ซึ่งเป็นสิ่งมองเห็นได้ในขณะนั้น จนถึงปัจจุบันถึงแม้ว่าความรู้ต่าง ๆ เกี่ยวกับผนังเซลล์พืชได้มีการค้นพบแล้วในระดับหนึ่ง แต่ความเข้าใจก็ยังไม่เพียงพอ โดยนิตยสาร Science ฉบับครบรอบ 125 ปี ได้รวมเรื่องของผนังเซลล์พืชเป็นหนึ่งใน 125 หัวข้อทางวิทยาศาสตร์ที่ยังต้องการคำตอบหรือความเข้าใจที่มากขึ้น หนึ่งในเหตุผลและ

ความสำคัญของผนังเซลล์คือการเป็นแหล่งตั้งต้นของพลังงานทางเลือกของโลก เนื่องจากผนังเซลล์เป็นส่วนย่อยที่สุดของชีวมวลซึ่งสามารถนำมาใช้เป็นพลังงานได้หลายรูปแบบ ตั้งแต่เผาจนถึงการเปลี่ยนเป็นพลังงานรูปแบบของเหลวหรือเอทานอล ความเข้าใจเกี่ยวกับผนังเซลล์ที่มากขึ้นจะทำให้สามารถผลิตชีวมวลที่มีคุณสมบัติที่ดี เหมาะสมในการผลิตพลังงาน และนำไปใช้ได้อย่างมีประสิทธิภาพสูงสุด และปลอดภัยกับสิ่งแวดล้อม (Gomez et al., 2008) ดังนั้นเพื่อความเข้าใจที่มากขึ้นจึงต้องมีการศึกษาองค์ประกอบของผนังเซลล์พืชและหน้าที่การทำงานระหว่างองค์ประกอบต่าง ๆ ที่ทำให้ผนังเซลล์ทำหน้าที่ได้

องค์ประกอบของผนังเซลล์ (Plant cell wall compositions)

ผนังเซลล์พืชเป็นโครงสร้างเชิงซ้อนประกอบด้วยโมเลกุลโพลีแซคคาไรด์หลายชนิดที่มาเชื่อมต่อกันเป็นโครงสร้างที่ทำหน้าที่ได้ โดยมีเซลลูโลสไมโครไฟบริลที่ประกอบด้วยสาย beta-glucan เล็ก ๆ จำนวนมากเป็นโครงสร้างหลัก โครงสร้างหลักนี้จะถูกจับด้วยโพลีแซคคาไรด์ชนิดอื่นในกลุ่มเฮมิเซลลูโลสและเชื่อมโยงเซลลูโลสไมโครไฟบริลเข้าด้วยกัน สร้างเป็นโครงสร้างที่ต่อเนื่องที่ทำหน้าที่รองรับแรงกระแทกต่าง ๆ ที่เกิดขึ้นทั้งจากภายในเซลล์และจากภายนอกเพื่อปกป้องเซลล์พืช (Delmer, 1999) โครงสร้างเซลลูโลสและเฮมิเซลลูโลสนี้จะถูกล้อมหรือหุ้มด้วยเมทริกซ์ของโพลีแซคคาไรด์ชนิดเพกตินอีกชั้นหนึ่ง เพื่อสร้างโครงสร้างผนังเซลล์ได้มีรูปแบบ นอกจากนี้ภายในโครงสร้างเชิงซ้อนก็จะมีกลุ่มของโปรตีน และสารประกอบอื่น ๆ แทรกอยู่เพื่อทำหน้าที่ (Carpita and Gibeaut, 1993)

เส้นใยเซลลูโลส (Cellulose microfibrils): เซลลูโลสเป็นโพลีแซคคาไรด์ที่พบมากที่สุดในโลก เป็นองค์ประกอบหลักของผนังเซลล์ โดยมีสัดส่วนประมาณ 40% อยู่ในรูปแบบ โครงสร้างพื้นฐานประกอบด้วย linear homopolymer of 1,4-linked beta-D-glucan โดยแต่ละสายจับกันด้วยพันธะไฮโดรเจนตามความยาวของโมเลกุลเกิดเป็นโครงสร้าง semi-crystalline microfibrils (McCann et al., 1990) ในผนังเซลล์เซลลูโลสสร้างโครงสร้างที่แข็งแรงผ่านการจับหรือเชื่อมต่อกับ cross-linking glycans หรือ hemicelluloses

เฮมิเซลลูโลส (Hemicelluloses): เป็นกลุ่มของโพลีแซคคาไรด์ที่แยกได้จากผนังเซลล์ด้วยคุณสมบัติการละลายในเบสสูง เฮมิเซลลูโลสชนิดหลักมี 2 ชนิด คือ xyloglucans (XyGs) และ glucuronoarabinosylans (GAXs) นอกจากนี้ ยังมีเป็นส่วนน้อย คือ MLGs ซึ่งพบเฉพาะในพืชกลุ่มหญ้าและชนิด glucomannans, galactoglucomannans, and galactomannans ที่พบในผนังเซลล์ชั้นปฐมภูมิ

XyGs นั้นพบในผนังเซลล์ของพืชทุกชนิด (Popper and Fry, 2005) XyGs เป็นองค์ประกอบหลักของผนังเซลล์ปฐมภูมิโดยมีสัดส่วนเท่า ๆ กันกับเซลลูโลส โดยเคลือบหรือล้อมรอบเซลลูโลสไมโครไฟบริลและเป็นตัวเชื่อมโยงระหว่างสายเซลลูโลสเข้าด้วยกัน XyG เป็นโพลีแซคคาไรด์ในผนังเซลล์ที่มีการศึกษามากที่สุด เนื่องจากเป็นองค์ประกอบที่น่าจะมีส่วนสำคัญในการควบคุมการทำงานของผนังเซลล์ในช่วงการยืดขยายของเซลล์ และมีเอนไซม์หลายชนิดที่ทำงานโดยตรงกับ XyG โดยการเปลี่ยนแปลงแก้ไขโมเลกุล เช่น การตัดต่อและการแยกพันธะระหว่างโมเลกุล (Cosgrove, 2005) เช่น expansin ซึ่งเป็นเอนไซม์ที่ควบคุมการยืดขยายของผนังเซลล์โดยการทำลายพันธะไฮโดรเจนระหว่าง XyG กับเซลลูโลส (McQueen-Mason et al., 1992) และ xyloglucan transglycosylase/hydrolases (XTHs) ที่ตัดต่อโมเลกุล XyGs ในผนังเซลล์ให้ยาวขึ้น

เพกติน (Pectins): เป็นโพลีแซคคาไรด์ที่มีคุณสมบัติเป็นกรดอ่อน (acidic polysaccharides) ซึ่งมีองค์ประกอบหลักคือ D-galacturonic acid (GalA) (Carpita and McCann, 2000) โดยทั่วไปเพกตินมีคุณสมบัติละลายละลายน้ำได้ และสามารถสกัดออกจากผนังเซลล์ได้ง่ายโดยใช้ความร้อนหรือ calcium chelators โครงสร้างหลักของเพกติน แบ่งออกเป็น 3 ชนิด ตามองค์ประกอบน้ำตาล คือ homogalacturonan (HGA) rhamnogalacturonan I (RG-I) และ rhamnogalacturonan II (RG-II) ซึ่งจะพอร์มเครือข่ายด้วยพันธะโควา

เลนซ์และไอออนิกด้วย cation ion หลายชนิด กระบวนการเชื่อมต่อเครือข่ายเพกตินนั้นได้มีการพิสูจน์ว่ามีความสำคัญต่อการสร้างผนังเซลล์ การเกิดช่องของเซลล์ และความหนาของผนังเซลล์พืช ในผนังเซลล์พืชที่กำลังจะมีการเจริญพบว่า 75% ของ HGA มีการเติมหมู่เมทิลบนโมเลกุลเพื่อป้องกันการเกิดประจุ และไม่ให้เกิดการฟอร์มเครือข่ายเพกติน เพื่อให้เพกตินนั้นยังอยู่ในสภาวะที่ยังเปลี่ยนแปลงรูปร่างได้ แต่เมื่อผนังเซลล์มีการขยายเรียบร้อยแล้วหมู่เมทิลนั้นจะถูกกำจัดออกด้วยเอนไซม์ methylsterases และเครือข่ายเพกตินก็จะถูกสร้างขึ้นด้วย ion bridge สร้างเป็นเจลเพกติน (Carpita and Gibeaut, 1993) เพกตินมีหน้าที่ในผนังเซลล์หลายอย่าง เช่น wall porosity, cell adhesion ที่มีดีเซลลามลลา, ควบคุมสภาวะ pH และ ionic buffers, และเป็นโมเลกุลตัวส่งสัญญาณในขณะเจริญเติบโต หรือการบุกรุกจากโรคต่าง ๆ (Willats et al., 1999; Ridley et al., 2001) องค์ประกอบอื่น ๆ ของผนังเซลล์ ประกอบด้วย โปรตีน สารประกอบฟีนอลิก และลิกนิน

โปรตีนในผนังเซลล์และหน้าที่การทำงาน (Cell wall proteins and their function)

โปรตีนโครงสร้าง (Structural proteins): โปรตีนโครงสร้างเป็นองค์ประกอบส่วนย่อย แต่มีความสำคัญต่อการทำงานของผนังเซลล์ แบ่งออกเป็น 5 กลุ่ม ตามโครงสร้างขององค์ประกอบ โดย 3 กลุ่มแรกแบ่งด้วยองค์ประกอบของกรดอะมิโน คือ hydroxyproline-rich glycoproteins (HRGPs) หรือ เรียกอีกชื่อหนึ่งว่า “extensins” (Kieliszewski and Lamport, 1994; Xu et al., 2008), proline-rich proteins (PRPs; Kieliszewski and Lamport, 1994) และ glycine-rich proteins (GRPs) ส่วนกลุ่มที่ 4 คือ arabinogalactan proteins (AGPs; Kieliszewski and Lamport, 1994; Showalter et al., 2010) โดยโครงสร้างโมเลกุลประกอบด้วย glycosylated side chain มากกว่า 95% และกลุ่มที่ 5 คือ Ictin proteins (Showalter, 1993) โปรตีน 3 กลุ่มแรกมีการทำงานคล้าย ๆ กันคือ การจับกันเองสร้างเป็นเครือข่าย isodityrosine bridges (Epstein and Lamport, 1984) และเข้าจับกับเพกตินประกอบกันเป็น polysaccharide matrix (Keegstra et al., 1973) นอกจากนี้ยังทำหน้าที่เป็นตัวตรวจจับหรือตรวจสอบสภาพแวดล้อม ส่งสัญญาณระหว่างเซลล์ และกลไกการต่อต้านต่อโรคต่าง ๆ (Fincher et al., 1983; Showalter, 1993)

เอนไซม์: ทำหน้าที่สำคัญในการควบคุมการเจริญ การส่งสัญญาณ การต่อต้านการบุกรุก และกิจกรรมต่าง ๆ ของผนังเซลล์ เอนไซม์กลุ่มจะทำงานโดยตรงกับโพลีแซคคาไรด์ในผนังเซลล์ ด้วยการทำปฏิกิริยาต่าง ๆ คือ การย่อย (polysaccharide hydrolases) การตัดหมู่เมทิล, (pectin methyl esterase) การตัดและต่อสายโพลีแซคคาไรด์ (xyloglucan transglycosylase/hydrolases :XTHs) และ ทำลายพันธะไฮโดรเจนระหว่างสายโพลีแซคคาไรด์ (expansins) นอกจากนี้ยังมีเอนไซม์อื่น ๆ ที่ทำงานกับโมเลกุลอื่น ๆ เช่น peroxidases, invertases, phosphatases และ dehydrogenases (Fry, 1995).

ตามการศึกษาในปัจจุบันมีเอนไซม์ 3 กลุ่มที่ทำงานกับโมเลกุล XyGs และมีหน้าที่ควบคุมการยืดขยายของผนังเซลล์ในขณะที่มีการขยายขนาดเซลล์ กลุ่มแรกคือ Endoglucanases ทำหน้าที่ย่อย glucan chains และ XyGs เพื่อลดความแข็งแรงของโครงสร้างเซลล์ลูโลส-เฮมิเซลล์ลูโลส ทำให้ผนังเซลล์ยืดขยายได้ และคล้าย ๆ กันคือ XTHs ซึ่งทำหน้าที่ตัดและต่อสาย XyG เพื่อยอมให้ผนังเซลล์อ่อนลง และต่อสาย XTH ให้ยาวขึ้นเพื่อสามารถเชื่อมโครงสร้างผนังเซลล์ที่ขยายตัวออกแล้วได้ (Fry et al., 1992) และ Expansins เป็นตัวเร่งปฏิกิริยาการยืดขยายผนังเซลล์ในสภาวะกรด “acid induced extension” โดยทำปฏิกิริยาล้ำกับเอนไซม์สองชนิดที่กล่าวไปแล้ว แต่ในปฏิกิริยานั้นไม่มีการตัดพันธะของน้ำตาลในโพลีแซคคาไรด์ เอนไซม์นี้จะทำลายพันธะไฮโดรเจนระหว่างสายโพลีแซคคาไรด์ของ XyG กับเซลล์ลูโลสไมโครไฟบริล ทำให้โครงสร้างของผนังเซลล์ในลดความแข็งแรงลงและสามารถยืดขยายจากแรงดันภายในได้ (McQueen-Mason et al., 1992)

โปรตีนในผนังเซลล์ที่ยังไม่ทราบหน้าที่ (Cell wall-related proteins of unknown function)

นอกเหนือจากโปรตีนที่ได้กล่าวไปแล้วข้างต้น ในผนังเซลล์ยังมีโปรตีนอื่น ๆ อีกเป็นจำนวนมากที่ยังไม่ทราบหน้าที่และความสำคัญ โดยในอะราบิโดพซิสได้มีการประมาณว่า 10-15% ของยีนที่เป็นรหัสสำหรับโปรตีนนั้นมีความเกี่ยวข้องกับผนังเซลล์ ตั้งแต่การสังเคราะห์จนถึงการทำหน้าที่ในผนังเซลล์ (McCann and Carpita, 2008) จากการวิเคราะห์จีโนมของพืชและการแสดงออกของยีนด้วยเทคนิคทาง Bioinformatic ได้มีการบ่งชี้ขึ้นที่ยังไม่ทราบหน้าที่ แต่มีความเป็นไปได้สูงที่จะทำงานเกี่ยวกับผนังเซลล์ และได้มีการแบ่งกลุ่มโปรตีนเหล่านี้เป็น cell wall-related proteins of unknown function (PUFs) (Mewalal et al., 2014) เมื่อมีการศึกษาและจัดกลุ่มเพิ่มเติมพบว่า สามารถแบ่งโปรตีนที่ยังไม่ทราบหน้าที่เหล่านี้ออกเป็น 2 กลุ่มใหญ่ ๆ คือ 1) proteins of known function and 2) PUFs โดยกลุ่มแรกเป็นโปรตีนที่รู้หน้าที่จากการทำนายด้วยข้อมูลลำดับอะมิโนและโครงสร้างที่เหมือนหรือคล้ายกับโปรตีนที่มีการศึกษามาก่อนหน้านี้แต่ยังไม่มีการทดลองยืนยันว่ามีหน้าที่ใดในผนังเซลล์พืช กลุ่มนี้อาจจะเรียกได้อีกอย่างว่า “cell wall-known unknown proteins” ส่วนกลุ่มที่ 2 (PUF group) นั้นเป็นโปรตีนที่มีลำดับเบสไม่เหมือนกันโปรตีนที่เคยมีการศึกษา ไม่สามารถเปรียบเทียบหน้าที่การทำงานกับโปรตีนอื่น ๆ ได้ ซึ่งกลุ่มนี้อาจแบ่งออกได้อีกเป็น hidden Markov model (HMM)-recognised domain of unknown function (DUF) ซึ่งมีการจัดกลุ่มโปรตีนที่มีความคล้ายคลึงกันเข้าด้วยกัน โดยจัดอยู่ใน Pfam databases และอีกกลุ่มที่ไม่มีสามารถหาโดเมนที่จดจำได้และตั้งชื่อเป็น proteins of obscure feature (POFs)

โปรตีนกลุ่ม cell wall-known unknown เป็นกลุ่มโปรตีนที่ได้จากการทำนายทางทฤษฎีว่ามีหน้าที่ทางชีวเคมีหรือทางเซลล์ชีววิทยา ยกตัวอย่างเช่น โปรตีนบางตัวได้ถูกจัดกลุ่มอยู่ใน kinase superfamilies, methyl esterase, methyltransferase หรือ GTP binding protein ทั้งนี้โปรตีนเหล่านี้อาจทำหน้าที่ตามการจัดกลุ่มดังกล่าว แต่การทำงานหรือหน้าที่ที่เกี่ยวข้องกับผนังเซลล์นั้นไม่สามารถทำนายได้ อย่างไรก็ตามข้อมูลการทำนายเบื้องต้นนี้ก็สามารถใช้เป็นแนวทางในการออกแบบการทดลองเพื่อพิสูจน์หน้าที่ของโปรตีนที่ง่ายขึ้น (Galperin and Koonin, 2010) ตัวอย่างของกรณีนี้คือ การศึกษา eukaryotic translation elongation factor (EF-1B β 1) ซึ่งทราบมาก่อนแล้วว่าเป็นส่วนหนึ่งของ translation complex ในการสร้างโปรตีน แต่เมื่อมีการศึกษาอะราบิโดพซิสพันธุ์กลายที่ทำลายหน้าที่ของยีนนี้พบผลที่เกี่ยวข้องกับผนังเซลล์โดยมีสัดส่วนของเซลล์ลูลอสและลิกลินที่ลดลง (Hossain et al., 2012) ดังนั้นยังมีโปรตีนอีกจำนวนมากที่มีการทำนายหน้าที่โดยใช้ความใกล้เคียงทางลำดับกรดอะมิโน แต่ยังไม่ทราบหน้าที่การทำงานในระดับเซลล์ ซึ่งต้องมีการศึกษาในรายละเอียดต่อไป

Domains of unknown function (DUFs) เป็นกลุ่มของโปรตีนที่มีการจัดแบ่งโดย Pfam database โดยที่ยังไม่สามารถหาความคล้ายคลึงกับโปรตีนอื่น ๆ ที่มีการศึกษามากแล้ว และไม่สามารถทำนายการทำงานได้ ปัจจุบันกลุ่ม DUFs ได้มีการจัดแบ่งออกเป็น มากกว่า 2,000 แฟมิลี และมีการค้นพบมากขึ้นเรื่อย ๆ ตามการศึกษาทางจีโนมและโปรตีโอมิกส์ ถึงแม้ว่าการจัดกลุ่มนี้จะไม่ได้ช่วยบ่งชี้หน้าที่ของโปรตีน แต่ก็จะมีโดยประโยชน์โดยเมื่อมีการศึกษาสมาชิกตัวใดตัวหนึ่งความรู้ความเข้าใจที่ได้ก็จะสามารถนำมาใช้ช่วยในการวิเคราะห์ศึกษาสมาชิกตัวอื่น ๆ ได้ สะดวกและรวดเร็วยิ่งขึ้นได้ (Galperin and Koonin, 2010; Punta et al., 2011) ทั้งนี้ DUF266 เป็นแฟมิลีแรกที่มีรายงานการทำหน้าที่ที่เกี่ยวข้องกับผนังเซลล์พืช โดยค้นพบจากการศึกษาข้าวพันธุ์กลายที่ต้านทานต่อโรค rice brittle และจากกาค้นหาดำแหน่งยีนก็ตรวจพบยีน *BC10* ที่เป็นสมาชิกของ DUF266 โดยยีนนี้มีความสำคัญเกี่ยวข้องกับการควบคุมระดับของเซลล์ลูลอสกับ arabinogalactan proteins ซึ่งคาดว่ามีส่วนต่อความแข็งแรงของผนังเซลล์ และต้นข้าว ส่งผลให้ทำให้ข้าวโตช้าลง และต้านทานโรคมากขึ้น (Zhou et al., 2009) แฟมิลี DUF อื่น ๆ ที่มีการศึกษาแล้วพบว่าเกี่ยวข้องกับผนังเซลล์ ได้แก่ DUF246 ซึ่งพบว่าเป็นเอนไซม์ glycosyltransferase ที่สังเคราะห์โพลีแซคคาไรด์ที่เป็นองค์ประกอบของผนังเซลล์ (Hansen et al., 2012) DUF231 เป็นเอนไซม์ esterase ที่ทำหน้าที่ตัดหมู่อะซิติกบนโมเลกุลโพลีแซคคาไรด์ในผนังเซลล์พืช (Bischoff et al., 2010) และ DUF579 ซึ่งเป็นเอนไซม์ methyltransferase ทำหน้าที่เติมหมู่เมทิล

เข้าสู่โพลีแซคคาไรด์ (Lee et al., 2012, Urbanowicz et al., 2012) และล่าสุด DUF642 ซึ่งมีการศึกษาเพียงบางส่วนและยังไม่สามารถสรุปถึงหน้าที่ของโปรตีนได้ DUF642 เป็นแฟมิลีที่จะศึกษาในครั้งนี้และจะกล่าวถึงในส่วนถัดไป ในขณะที่แฟมิลี DUF อื่น ๆ ที่อาจมีความเกี่ยวข้องกับผนังเซลล์กำลังมีการศึกษาอยู่ อย่างไรก็ตามเนื่องจากกลุ่มโปรตีน DUF นั้นไม่มีลำดับกรดอะมิโนที่จดจำได้ หรือมีโมดูลโปรตีนที่ไม่เหมือนโปรตีนอื่นที่มีการศึกษามาก่อน ทำให้การศึกษาหน้าที่ของโปรตีนแฟมิลี DUF เป็นไปค่อนข้างช้าและลำบาก

DUF642 family

โปรตีนแฟมิลี DUF642 ในอะราบิดอปซิสประกอบด้วยสมาชิกทั้งหมด 10 ตัว ซึ่งถอดรหัสจากยีนทั้งหมด 10 ยีน จากรายงานตีพิมพ์โปรตีนแฟมิลีนี้ได้ถูกจัดจำแนกจากการศึกษาทางโปรตีโอมิกส์ของผนังเซลล์ของเนื้อเยื่อพืชที่กำลังมีการเจริญเติบโต (Bayer et al., 2006; Irshad et al., 2008) โดยมีสมาชิก 2 ยีน (At1g80240 และ At5g25460) ที่มีการแสดงออกที่ตอบสนองกับการเติมสารเคมี L-galactono-1,4-lactone ในอะราบิดอปซิสพันธุ์กลายที่ไม่ตอบสนองต่อการดัดแปลง (Gao et al., 2011) และการศึกษาพบว่าการทำลายหน้าที่ของยีน At5g25460 ส่งผลให้พืชมีต้นขนาดเล็กลง และรากสั้น อย่างไรก็ตามการทำลายยีน At1g80240 นั้นยังไม่พบว่ามีผลต่อพืช (Gao et al., 2012) นอกจากนี้สมาชิกอีก 2 ยีน คือ At5g11420 และ At4g32460 ได้มีการศึกษาการทำงานร่วมกับโปรตีนอื่น ๆ พบว่ามีการจับกับ pectic methylesterase protein จึงมีสมมุติฐานว่าโปรตีน DUF642 นี้ น่าจะทำงานเกี่ยวข้องกับเพกติน (Zuniga-Sanchez and Gamboa de Buen, 2012) การศึกษาริคอมบิแนนท์โปรตีนของยีน At3g08030 พบว่ามีคุณสมบัติจับกับ cellulose, cellobiose และ hemicellulose ซึ่งแสดงให้เห็นว่าโปรตีนแฟมิลีนี้อาจทำงานกับองค์ประกอบอื่น ๆ ด้วย นอกจากเพกติน (Vazquez-Lobo et al., 2012) ส่วนยีนสมาชิกอื่น ๆ นั้นนั้นไม่มีการศึกษา ถึงแม้ว่าตอนนี้จะมีข้อมูลของยีน DUF642 ว่ามีการทำหน้าที่ในผนังเซลล์ แต่ข้อมูลนั้นยังไม่เพียงพอที่จะบ่งชี้ถึงหน้าที่ทางชีววิทยาของโปรตีนแฟมิลีนี้ในระดับโมเลกุล

วิธีการทดลอง

การสืบหาข้อมูลเบื้องต้นของโปรตีนแฟมิลี DUF642

เป็นการสืบค้นข้อมูลเกี่ยวกับโครงสร้างของยีนและลำดับเบสที่เป็นรหัสของโปรตีนแฟมิลี DUF642 จำนวน 10 ยีน ในอะราบิดอปซิส โดยใช้ฐานข้อมูล Phytozome database (<http://phytozome.jgi.doe.gov>) ส่วนข้อมูลการแสดงออกของยีนทั้ง ระดับการแสดงออกและตำแหน่งของเนื้อเยื่อที่เกิดการแสดงออกจะสืบค้นจากฐานข้อมูล microarray และฐานข้อมูลวิเคราะห์การแสดงออกของยีนในอะราบิดอปซิส

การตรวจหาตำแหน่งการทำงานของโปรตีนแฟมิลี DUF642 ด้วยวิธี Promoter-GUS fusion

โคลนส่วนโปรโมเตอร์ของยีน DUF642 ทั้ง 10 ยีน ในช่วงขนาด 1500-2000 คู่เบส จากนั้นนำไปเชื่อมต่อเข้าเวกเตอร์ pCXGUS-P (Chen et al., 2009) ที่ตำแหน่งด้านหน้ายีน GUS 9 รวบรวมความถูกต้องโดยการวิเคราะห์ลำดับเบส จากนั้นนำไปถ่ายโอนเข้าสู่อะราบิดอปซิส สายพันธุ์ col-0 ผ่านวิธีอะโกรแบคทีเรีย โดยการทำให้ floral dipping คัดเลือกต้นที่ได้รับยีนโดยการเพาะเลี้ยงบนอาหาร 1/2MS เติมน้ำ 1% และสารปฏิชีวนะคัดเลือก hygromycin B จากนั้นนำต้นที่คัดเลือกมาปลูกและเก็บเนื้อเยื่อที่ระยะต่างมาย้อมสีด้วยสารตั้งต้น X-glucA ล้างเนื้อเยื่อที่ไม่ติดสีด้วยเอทานอล

การผลิตและตรวจสอบการทำงานของโปรตีนรีคอมบิแนนท์ DUF642 โดยการจับกับโพลีแซคคาไรด์ของผนังเซลล์

เป็นการสร้างเวกเตอร์สำหรับการผลิตรีคอมบิแนนท์โปรตีน DUF642 ทั้ง 10 ตัว ด้วยระบบการแสดงออกของโปรตีนด้วยเซลล์ *E. coli* โดยใช้ pGEX expression system ซึ่งระบบนี้จะใช้โปรตีน GST สำหรับการทำให้รีคอมบิแนนท์โปรตีนให้บริสุทธิ์ จากนั้นนำโปรตีนที่ได้ไปตรวจสอบการจับอย่างจำเพาะกับโพลีแซคคาไรด์ชนิดต่าง ๆ ประมาณ 20 ชนิด ที่เป็นองค์ประกอบของผนังเซลล์พืช ในกลุ่ม เซลลูโลส เฮมิเซลลูโลส และ เพกติน

นอกจากนี้รีคอมบิแนนท์โปรตีนที่ได้จะนำไปทดสอบ competitive digestion assays กับโพลีแซคคาไรด์ที่จับได้ เพื่อดูว่าโปรตีนที่เติมลงไปจะลดความสามารถในการย่อยของ hydrolysis enzyme ได้หรือไม่

การศึกษาผลของ gene disruption ในยีนแฟมิลี DUF642 โดยใช้ T-DNA insertion lines

สายพันธุ์อะราบิดอปซิสชนิด insertion line ที่มีชิ้น T-DNA แทรกอยู่ในตำแหน่งต่าง ๆ บนยีน DUF642 ทั้ง 10 ยีน จะส่งมาจาก Arabidopsis collection centre และคัดเลือกสายพันธุ์ที่เป็นโฮโมไซกัสเพื่อใช้ศึกษาต่อไป ส่วนยีนใดที่ไม่มีสายพันธุ์ที่มีชิ้น T-DNA ก็จะสามารถสร้างสายพันธุ์กลายเฉพาะยีนด้วยวิธีแก้ไขจีโนมเนื่องจาก DUF642 เป็นยีนแฟมิลีที่มีสมาชิกอยู่ถึง 10 ยีน ทำให้มีความเป็นไปได้สูงว่าจะมี functional redundancy ในกลุ่มยีนนี้ ซึ่งส่งผลให้เมื่อยับยั้งการทำงานของยีนใดยีนหนึ่งก็อาจไม่เกิดผลใดต่อต้นพืช เนื่องจากยังมียีนสมาชิกอื่นทำหน้าที่แทนได้ ดังนั้นในการศึกษาครั้งนี้จึงต้องสร้างสายพันธุ์ที่มีการเสียการทำงานของยีน DUF642 พร้อม ๆ กันหลายยีน (multiple mutation) โดยการผสมข้ามสายพันธุ์กลายที่ได้เข้าด้วยกันและคัดเลือกต้นที่เป็นโฮโมไซกัสของการกลายหลาย ๆ ตำแหน่ง

เมื่อได้สายพันธุ์กลายหลายตำแหน่งแล้ว สายพันธุ์ที่ได้จะนำไปวิเคราะห์เพื่อตรวจหาความเปลี่ยนแปลงต่าง ๆ ดังนี้ 1) การเจริญเติบโต ในระดับต้นอ่อน ราก ไฮโปคอติล ก้านช่อดอก และขนาดของต้น 2) ตรวจสอบรูปร่างของเซลล์จากผลของการเปลี่ยนแปลงในผนังเซลล์ 3) ตรวจสอบองค์ประกอบของโพลีแซคคาไรด์ในผนังเซลล์ที่ได้เพื่อตรวจดูความเปลี่ยนแปลงขององค์ประกอบและโครงสร้าง

การตรวจสอบการย่อยสลายของผนังเซลล์ด้วยเอนไซม์

อบแห้งตัวอย่างเนื้อเยื่อจากดอกที่อุณหภูมิ 60 องศาเซลเซียสจนแห้ง นำไปบดเป็นผงด้วย bead beater ตวงผงตัวอย่าง 20 กรัม ใส่หลอดทดลอง เติมน้ำ 0.3 M NaOH สำหรับขั้นตอน pre-treatment เป็นเวลา 30 นาที ล้างสาร pre-treatment ออก แล้วเติม sodium acetate buffer pH 4.5 ที่ผสมชุดเอนไซม์สำหรับการย่อยชีวมวล ปั่นที่อุณหภูมิ 50 องศาเซลเซียส เป็นเวลา 8 ชั่วโมง จากนั้นนำสารละลายที่ได้มาตรวจสอบด้วยสาร MBTH วิเคราะห์ปริมาณโดยใช้กราฟมาตรฐานของน้ำตาลกลูโคส ทดลองโดยใช้ตัวอย่าง 3 ซ้ำทางชีวภาพ และแต่ละตัวอย่างวิเคราะห์อย่างน้อย 2 ซ้ำ คำนวณการปลดปล่อยน้ำตาลเป็นหน่วย นาโนโมลน้ำตาลต่อมิลลิกรัมตัวอย่าง ต่อชั่วโมง

การวิเคราะห์องค์ประกอบน้ำตาลโมเลกุลเดี่ยวโดยการย่อยผนังเซลล์

สกัดผนังเซลล์โดยการบดในไนโตรเจนเหลว แล้วล้างด้วยเมทานอล เอทานอล และอะซิโตน ตามลำดับ จากนั้นอบแห้งที่อุณหภูมิ 60 องศาเซลเซียส นำไปย่อยด้วยกรด trifluoroacetic acid ที่อุณหภูมิ 100 องศาเซลเซียสเป็นเวลา 4 ชั่วโมง จากนั้นกำจัดกรดด้วยการ freeze-dry จากนั้นละลายด้วยน้ำวิเคราะห์ด้วย Dionex HPLC CarboPac column วิเคราะห์ปริมาณด้วยกราฟมาตรฐานจากน้ำตาลโมเลกุลเดี่ยว 9 ชนิด

การตรวจสอบการเจริญเติบโตร่วมกับสภาวะเครียดและสารยับยั้งการทำงานของผนังเซลล์

นำเมล็ดอะราบิโดพซิสที่ฆ่าเชื้อแล้วมาเพาะลงบนเพลทอาหาร 1/2MS ที่เติมวัน 1% น้ำตาลซูโรส 1% และเติมสารสำหรับทดสอบ ดังนี้ 100 mM NaCl, 100 mM Sorbitol, $AlCl_3$, Isoxaben, และ Oryzalin จากนั้นเพาะเลี้ยงเป็นเวลา 7-10 วัน ตรวจวัดขนาดของราก

การศึกษาผลของโปรตีนแฟมิลี DUF642 ต่อการเจริญและพัฒนาของต้นอ่อนอะราบิโดพซิส

ยีนสมาชิกของแฟมิลี DUF642 ทั้ง 10 ยีน จะนำมาแสดงออกแบบชั่วคราวในต้นอ่อนของอะราบิโดพซิส โดยต้นที่ได้รับการถ่ายยีนจะนำมาตรวจสอบผลของโปรตีนต่อการเจริญและพัฒนาของต้นอ่อน นอกจากนี้รีคอมบิแนนท์โปรตีนทั้ง 10 ตัวที่ผลิตได้จะนำมาทดสอบโดยการแช่ต้นอ่อนผสมกับอะโกรแบคทีเรียที่มีเวกเตอร์สำหรับการถ่ายโอนยีน โดยใช้วิธี Agrobast (Wu et al., 2014) แล้วตรวจสอบการยืดขนาดของรากต้นอ่อน

การตรวจหาโปรตีนอื่น ๆ ที่ทำหน้าที่ร่วมกับโปรตีนแฟมิลี DUF642 ด้วยวิธี immunoprecipitation

เป็นการตรวจหาโปรตีนที่น่าจะทำงานร่วมกันกับโปรตีน DUF642 โดยใช้ต้นอ่อนอะราบิโดพซิสที่มีการถ่ายโอนเวกเตอร์ให้มีการแสดงออกแบบชั่วคราวแล้วสกัดโปรตีน คัดแยกโปรตีน DUF642 ด้วยวิธี immunoprecipitation ด้วยแอนติบอดีจำเพาะกับ epitope tag แล้วตรวจสอบจาก SDS-PAGE เพื่อหาโปรตีนที่ติดตามออกมาจากการคัดแยกด้วย แล้วบ่งชี้โปรตีนที่ได้ด้วยการหาลำดับกรดอะมิโน

Yeast two hybrids

สร้างเวกเตอร์สำหรับการแสดงออกของยีนในยีสต์ โดยที่มีส่วนประกอบของรหัสผลิต activation domain และ binding domain เวกเตอร์ที่ใช้คือ pGAD และ pGBK จากนั้นนำไปถ่ายโอนเข้าสู่ยีสต์ 2 สายพันธุ์ คัดเลือกด้วยอาหารที่ขาดอาหารสำหรับการคัดเลือก จากนั้นนำยีสต์ 2 สายพันธุ์ที่ต้องการทดสอบการทำงานร่วมกันระหว่างโปรตีน มาผสมกันโดยบ่มที่อุณหภูมิ 28 องศาเซลเซียสเป็นเวลา 1 คืน จากนั้นนำไปเพาะเลี้ยงในอาหารคัดเลือก ยีสต์ที่เจริญในอาหารที่ขาดสารสำหรับการคัดเลือกร่วมกันแสดงว่าโปรตีนที่ทดสอบสามารถเข้าจับกันได้

ผลการทดลองและวิจารณ์

1. ข้อมูลเบื้องต้นของยีน *DUF642* family จากอะราบิโดพซิส

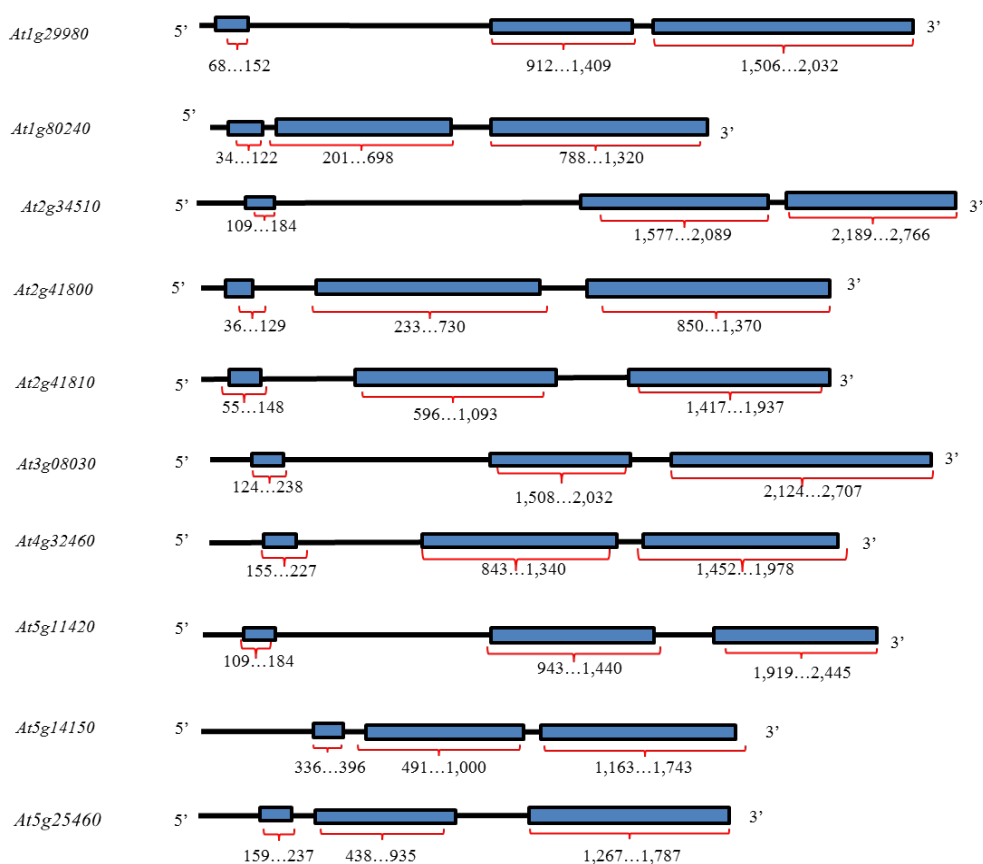
DUF642 family ของอะราบิโดพซิส ประกอบด้วยสมาชิก 10 ยีน แต่ละยีนมีโครงสร้างยีนที่คล้ายกัน ประกอบด้วยเอกซอน 3 ส่วน และ อินทรอน 2 ส่วน ตามภาพที่ 1 รวมมีรหัสสำหรับสังเคราะห์โปรตีน (coding sequence) เฉลี่ยยาวประมาณ 1,400 คู่เบส แปลรหัสเป็นโปรตีนประมาณ 350-400 กรดอะมิโน และมีน้ำหนักโมเลกุลประมาณ 40 กิโลดัลตัน (ตารางที่ 1) เอกซอนที่ 1 เมื่อแปลรหัสเป็นโปรตีนจะได้เป็นเปปไทด์ส่งสัญญาณ (signal peptide) ขนาดประมาณ 20-30 กรดอะมิโน ซึ่งทำหน้าที่เป็นสัญญาณขนส่งโปรตีนเป้าหมายออกสู่ภายนอกเซลล์ (apoplasmic space)

จากการรวบรวมข้อมูลการแสดงออก *DUF642* family ทั้ง 10 ยีน โดยใช้ฐานข้อมูลเว็บไซต์ Arabidopsis eFP Browser แสดงข้อมูลระดับการแสดงออกของยีน *DUF642* ทั้ง 10 ยีน ในอวัยวะต่าง ๆ พบว่า 5 สมาชิก มีการแสดงออกทุกเนื้อเยื่อและระดับการแสดงออกที่สูง และ อีก 5 สมาชิกที่เหลือ มีการแสดงออกเฉพาะบางเนื้อเยื่อและมีระดับการแสดงออกที่ต่ำ โดยรวมพบว่าการแสดงออกของ *DUF642* family ในทุก ๆ เนื้อเยื่อและตลอดเวลาในต้นอะราบิโดพซิส ตามตารางที่ 2

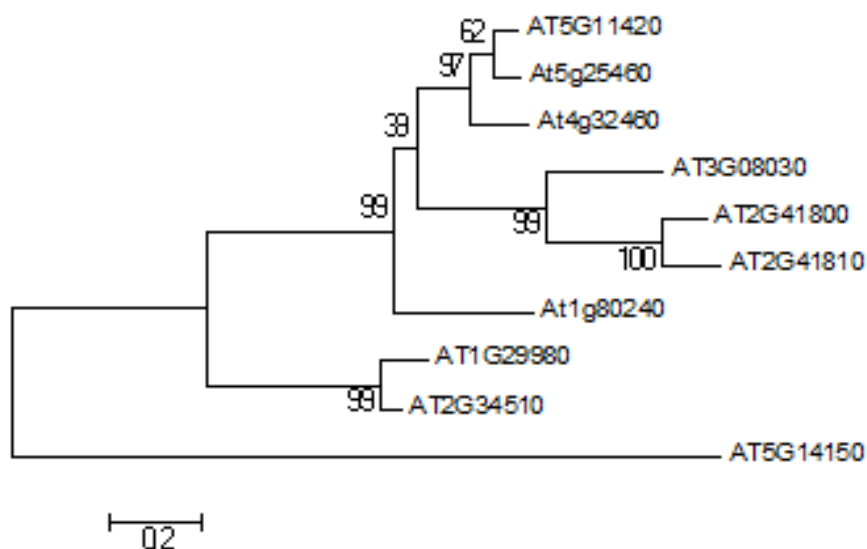
จากการนำลำดับกรดอะมิโนของทั้ง 10 โปรตีน มาเปรียบเทียบ แล้วสร้างต้นไม้วิวัฒนาการด้วยโปรแกรม Mega6 พบว่า ลำดับอะมิโนของทั้ง 10 โปรตีนมีความเหมือนกันสูง ซึ่งค่า distant ไกลกันซึ่งสามารถแบ่งได้ 5 กลุ่ม ได้แก่ กลุ่มที่ 1 คือโปรตีน *DUF642-8*, *DUF642-10*, *DUF642-7* กลุ่มที่ 2 คือโปรตีน *DUF642-6*, *DUF642-4*, *DUF642-5* กลุ่มที่ 3 คือโปรตีน *DUF642-2* กลุ่มที่ 4 คือโปรตีน *DUF642-1* และ *DUF642-3* และกลุ่มที่ 5 คือ โปรตีน *DUF642-9* ตามภาพที่ 2

ตารางที่ 1 ข้อมูล *DUF642* family จำนวน 10 ยีน

ชื่อยีน	ชื่อยีน	ขนาดนิวคลีโอไทด์ (คู่เบส)	ขนาดโพลีเปปไทด์ (กรดอะมิโน)	น้ำหนักโมเลกุล (กิโลดัลตัน)
<i>At1g29980</i>	<i>DUF642-1</i>	1,473	407	44.29
<i>At1g80240</i>	<i>DUF642-2</i>	1,316	370	40.22
<i>At2g34510</i>	<i>DUF642-3</i>	1,470	401	43.94
<i>At2g41800</i>	<i>DUF642-4</i>	1,290	370	40.37
<i>At2g41810</i>	<i>DUF642-5</i>	1,207	370	40.26
<i>At3g08030</i>	<i>DUF642-6</i>	1,519	365	39.06
<i>At4g32460</i>	<i>DUF642-7</i>	1,489	365	39.81
<i>At5g11420</i>	<i>DUF642-8</i>	1,513	366	39.64
<i>At5g14150</i>	<i>DUF642-9</i>	1,416	383	40.90
<i>At5g25460</i>	<i>DUF642-10</i>	1,460	369	39.97



ภาพที่ 1 โครงสร้างยีนของ DUF642 family จำนวน 10 ยีน แต่ละยีนประกอบด้วยเอกซอน 3 ส่วน และ อินทรอน 2 ส่วน



ภาพที่ 2 ต้นไม้วิวัฒนาการของครอบครัวโปรตีน DUF642 ทั้ง 10 โปรตีน วิเคราะห์จากโปรแกรม Mega6 สามารถแบ่งออกเป็น 5 กลุ่ม กลุ่มที่ 1 คือโปรตีน DUF642-8, DUF642-10, DUF642-7 กลุ่มที่ 2 คือโปรตีน DUF642-6, DUF642-4, DUF642-5 กลุ่มที่ 3 คือโปรตีน DUF642-2 กลุ่มที่ 4 คือโปรตีน DUF642-1 และ DUF642-5 และกลุ่มที่ 5 คือ โปรตีน DUF642-9

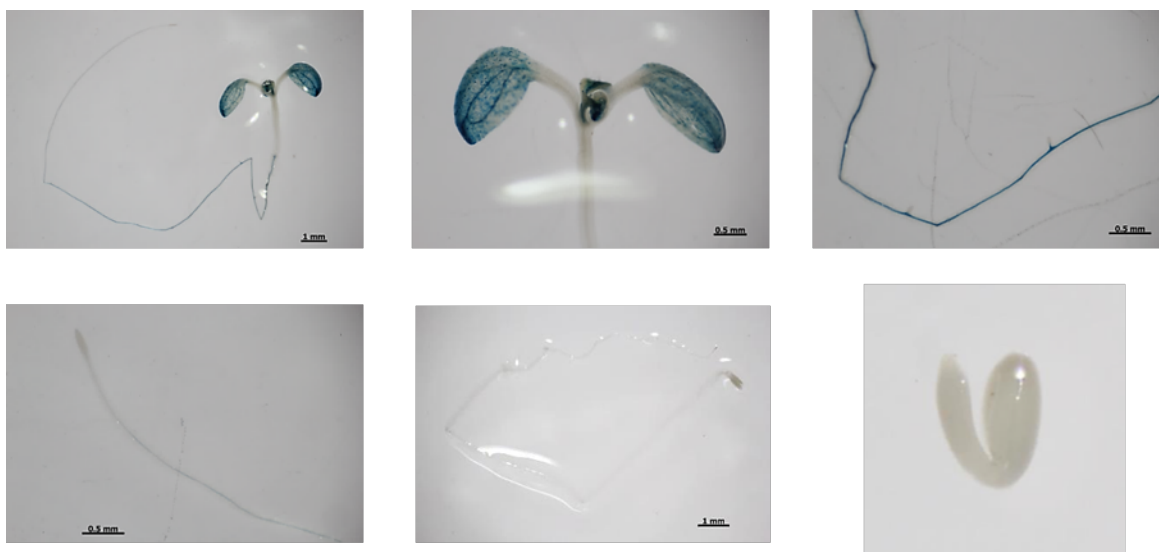
2. การแสดงออกของยีนแฟมิลี *DUF642* ด้วยวิธี *Promoter-GUS fusion*

ข้อมูลการแสดงออกของยีนมีความสำคัญต่อการแปลผลการทำงานของยีน ดังนั้นจึงต้องตรวจสอบการแสดงออกของยีนแฟมิลี *DUF642* ทั้ง 10 ว่ามีรูปแบบการแสดงออกอย่างไร จากการตรวจสอบข้อมูลการแสดงออกของยีน *DUF642-1* - *DUF642-10* ในฐานข้อมูลการแสดงออกของยีนในอะราบิโดพซิสพบว่า ทั้ง 10 ยีนมีรูปแบบการแสดงออกที่แตกต่างกัน (ตารางที่ 2) บางยีนมีการแสดงออกสูง บางยีนมีการแสดงออกต่ำ หรือมีรูปแบบการแสดงออกเฉพาะเนื้อเยื่อ อย่างไรก็ตาม ข้อมูลดังกล่าวได้มาจากการศึกษาขนาดใหญ่ อาจมีข้อมูลผิดพลาดได้ ดังนั้นจึงต้องมีการตรวจสอบแบบจำเพาะเพื่อบ่งชี้รูปแบบการแสดงออกของยีนอย่างละเอียด

การตรวจสอบการแสดงออกแบบจำเพาะนี้ใช้วิธี *Promoter::GUS fusion* โดยการนำโปรโมเตอร์ของยีน *DUF642* ทั้ง 10 ยีน ขนาด 1.5-2.0 กิโลเบส มาเชื่อมต่อบนเวกเตอร์ที่บริเวณด้านหน้ายีน *GUS* (*beta-glucuronidase*) แล้วถ่ายโอนชุดการแสดงออกของยีนเข้าสู่ต้นอะราบิโดพซิส คัดเลือกต้นที่ได้รับยีน แล้วจึงตรวจสอบการแสดงออกด้วยปฏิกิริยาของเอนไซม์ต่อสารตั้งต้น X-gluc บริเวณเนื้อเยื่อพืชที่มีสีฟ้าแสดงว่ามียีนแสดงออกที่ตำแหน่งนั้น ๆ ตามภาพที่ 3- แสดงผลการแสดงออกของยีน *DUF642* ทั้ง 10 ยีน

ตารางที่ 2 ระดับการแสดงออกของ *DUF642* family ในต้นอะราบิโดพซิสจากฐานข้อมูล eFP-browser

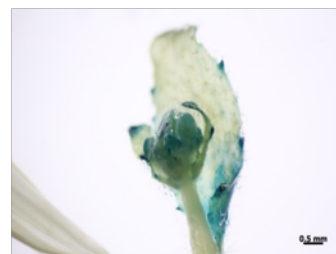
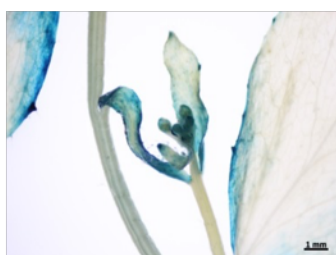
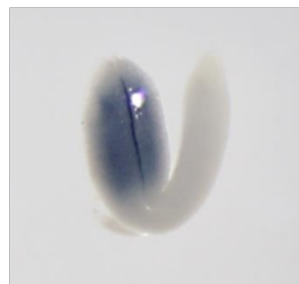
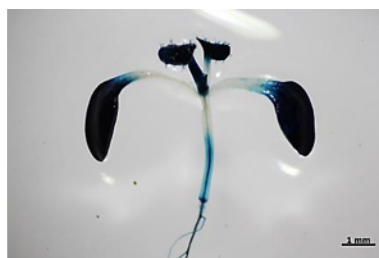
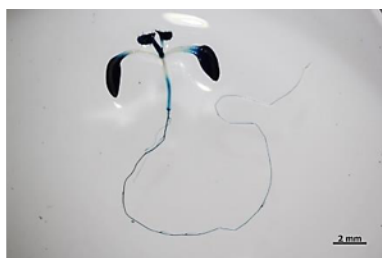
แสดงระดับการแสดงออกของ DUF gene family ทั้ง 10 ยีน ในระยะการเจริญของเมล็ด								แสดงระดับการแสดงออกของ DUF gene family ทั้ง 10 ยีน ในระยะต้น เจริญ		แสดงระดับการแสดงออก ของ DUF gene family ทั้ง 10 ยีน ในระยะเจริญของ ดอก				
ลำดับ ที่	ชื่อ	Dry seed	24h Imbibed seed	root	hypocotyl	Cotyledon	Rosette	st 1 Node	nd 2 Node	leaf	Sepal	Petal	stamen	carpel
1	<i>DUF642-1</i>	0	1027.63	411.0	411.05	0	205.52	411.0	0	0	0	205.52	0	411.05
2	<i>DUF642-2</i>	0	0	161.7	0	0	0	0	0	0	0	0	0	0
3	<i>DUF642-3</i>	0	0	0	0	527.48	316.49	0	0	316.4	0	0	0	210.99
4	<i>DUF642-4</i>	0	84.14	420.7	0	0	0	0	0	0	0	0	0	294.49
5	<i>DUF642-5</i>	0	19.09	95.45	0	0	0	0	0	0	0	0	0	0
6	<i>DUF642-6</i>	0	1460.99	730.4	1460.99	730.49	730.49	730.4	438.2	146.0	438.29	438.29	146.09	876.59
7	<i>DUF642-7</i>	0	1115.83	0	278.95	0	1115.8	278.9	278.9	0	1115.8	1952.71	1115.8	1115.8
8	<i>DUF642-8</i>	0	1176.12	0	1008.1	672.07	1008.1	168.0	168.0	168.0	0	1680.18	336.03	0
9	<i>DUF642-9</i>	0	0	80.03	35.21	8.8	0	8.8	8.8	8.8	0	44.01	61.62	61.62
10	<i>DUF642-10</i>	0	1673.75	0	836.87	557.91	1115.8	557.9	557.9	0	1394.7	1952.71	836.87	1394.7



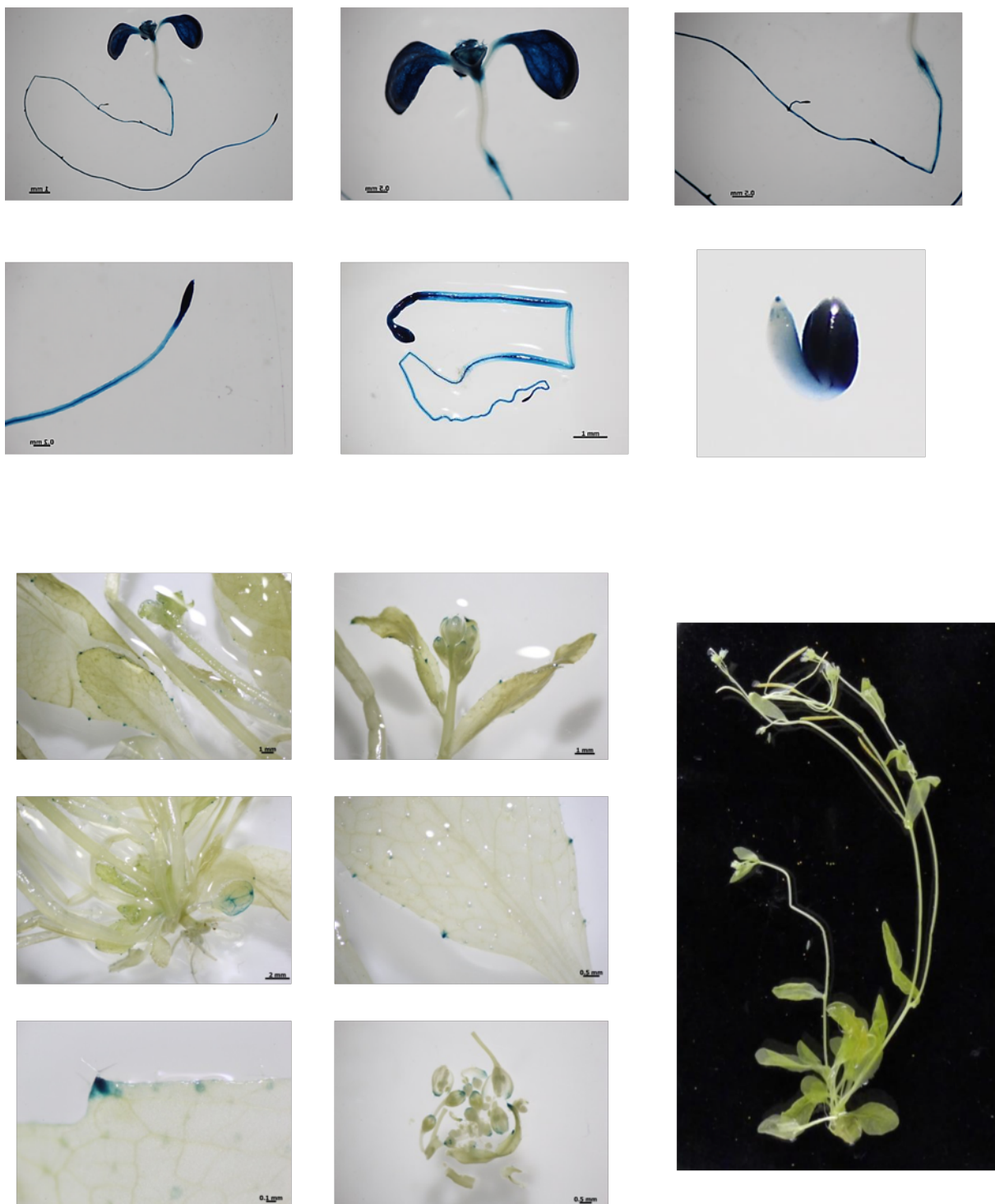
ภาพที่ 3 การแสดงออกของยีน *DUF642-1* ในต้นอะราบิโดพซิส



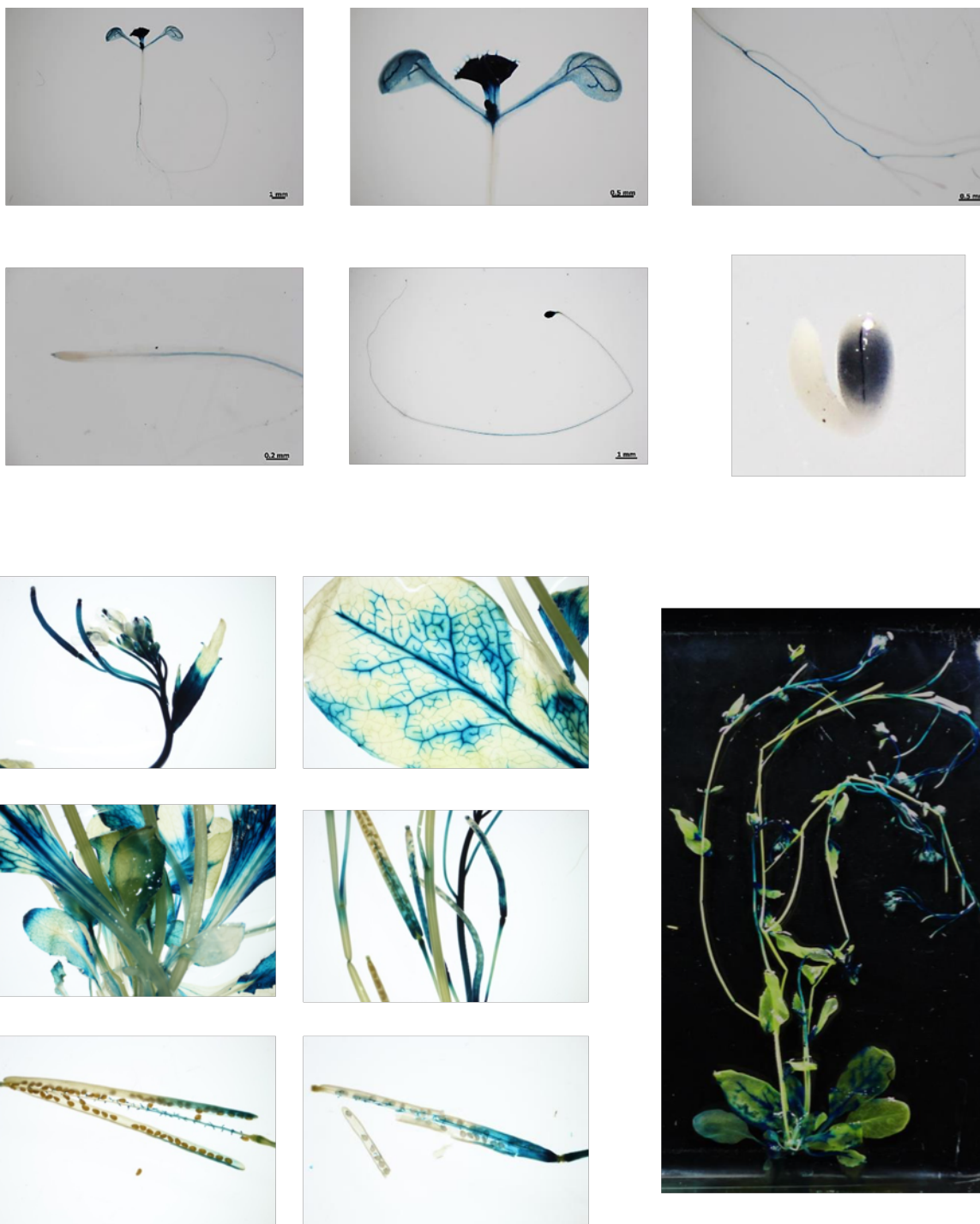
ภาพที่ 4 การแสดงออกของยีน *DUF642-2* ในต้นอะราบิโดพซิส



ภาพที่ 5 การแสดงออกของยีน *DUF642-3* ในต้นอะราบิโดพซิส



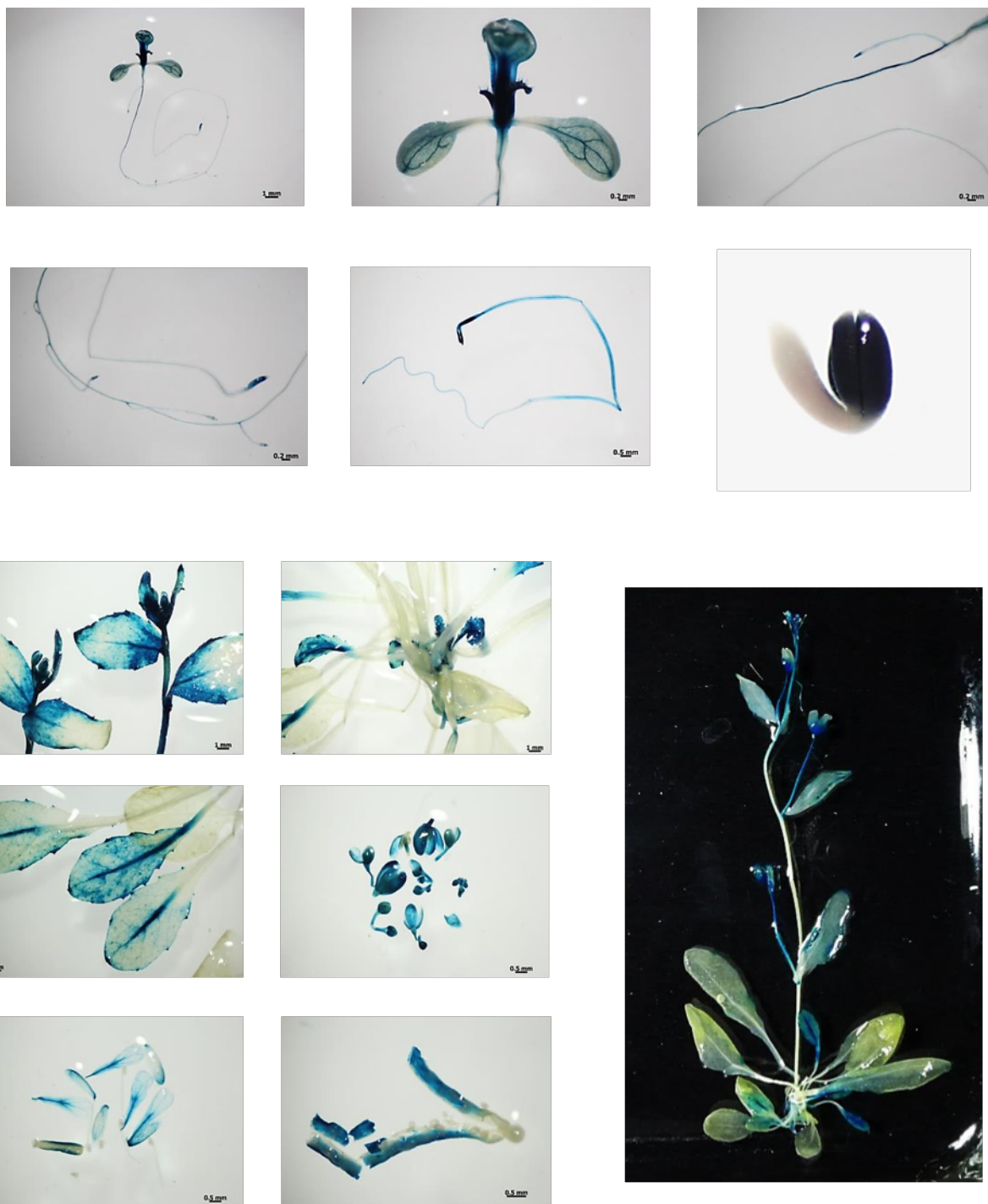
ภาพที่ 5 การแสดงออกของยีน *DUF642-4* ในต้นอะราบิโดพซิส



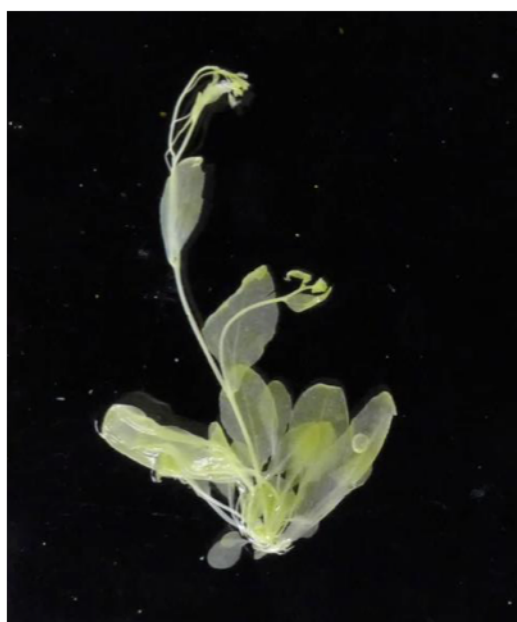
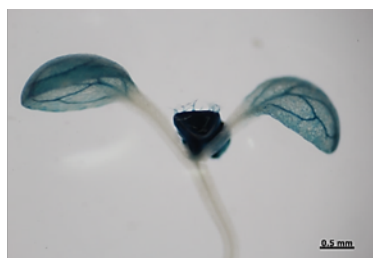
ภาพที่ 6 การแสดงออกของยีน *DUF642-5* ในต้นอะราบิโดพซิส



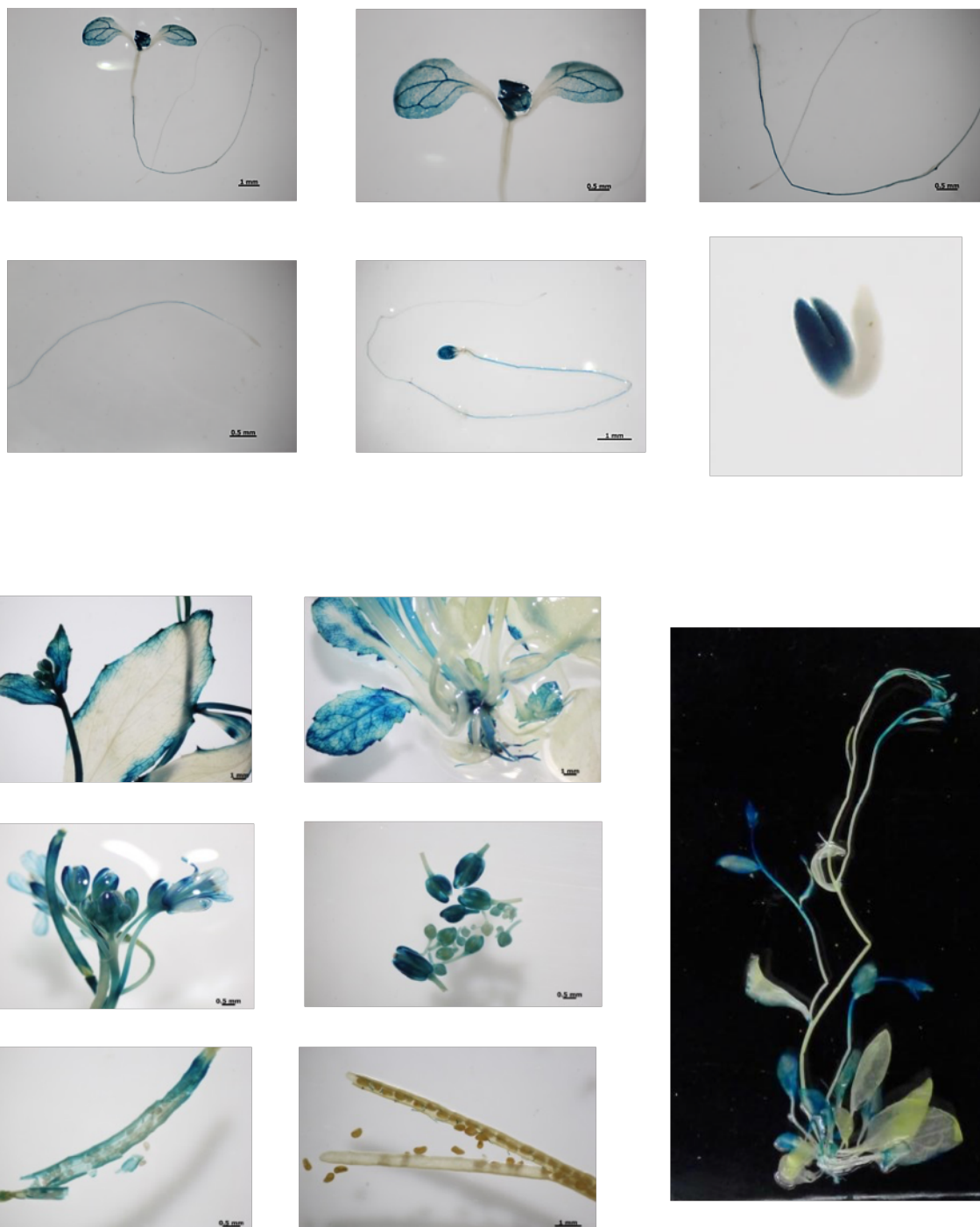
ภาพที่ 7 การแสดงออกของยีน *DUF642-6* ในต้นอะราบิโดพซิส



ภาพที่ 8 การแสดงออกของยีน *DUF642-7* ในต้นอะราบิโดพซิส



ภาพที่ 9 การแสดงออกของยีน *DUF642-8* ในต้นอะราบิโดพซิส



ภาพที่ 10 การแสดงออกของยีน *DUF642-9* ในต้นอะราบิโดพซิส



ภาพที่ 11 การแสดงออกของยีน *DUF642-10* ในต้นอะราบิโดพซิส

3. การตรวจสอบการทำงานของโปรตีนรีคอมบิแนนท์ DUF642 โดยการจับกับโพลีแซคคาไรด์ของผนังเซลล์

3.1 การโคลนยีน *DUF642* ทั้ง 10 ยีน จาก cDNA ของอะราบิโดพซิส

จากการรวบรวมลำดับเบสของยีน *DUF642-1* - *DUF642-10* จากฐานข้อมูล NCBI ได้ออกแบบไพรเมอร์สำหรับการโคลนยีนด้วยวิธีเพิ่มปริมาณดีเอ็นเอหรือพีซีอาร์ โดยใช้ cDNA ที่สังเคราะห์ขึ้นจากอาร์เอ็นเอที่สกัดจากต้นอ่อนอะราบิโดพซิส จากการทำพีซีอาร์พบว่าสามารถเพิ่มปริมาณยีนได้ทั้ง 10 ยีน และเมื่อนำไปเชื่อมต่อกับเวกเตอร์ pGem-Teasy ด้วยวิธี TA cloning นำเข้าสู่เซลล์ *E. coli* แล้วคัดเลือกโคลนด้วยวิธีโคลนนิ่งพีซีอาร์ สกัดพลาสมิด และวิเคราะห์ลำดับเบสพบว่ามีความถูกต้อง 100% ตามฐานข้อมูล

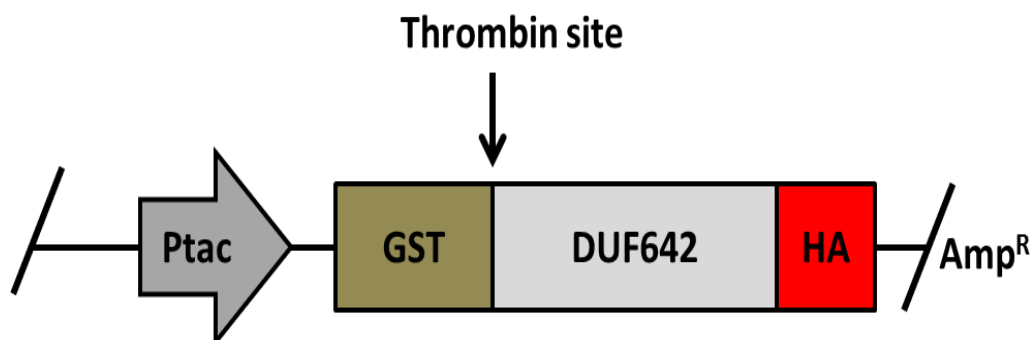
จากโครงสร้างของโปรตีน *DUF642* ที่มีลำดับเปปไทด์ส่งสัญญาณที่บริเวณ N-terminus ทำให้มีความเป็นไปได้ที่ลำดับเปปไทด์นี้จะรบกวนการฟิสิกส์คอมบิแนนท์โปรตีนใน *E. coli* ดังนั้นการสร้างพลาสมิดสำหรับการผลิตโปรตีนในครั้งนี้ จะไม่รวมลำดับเบสประมาณ 50 เบสในช่วงต้นของรหัสสำหรับการสังเคราะห์โปรตีน การเลือกโคลนเฉพาะส่วนของยีนใช้วิธีออกแบบไพรเมอร์สำหรับการทำพีซีอาร์

เมื่อออกแบบไพรเมอร์ที่ปลาย 5' ที่ตำแหน่งด้านหลังเปปไทด์ส่งสัญญาณ แล้วทำพีซีอาร์จากพลาสมิด pGEM-T-DUF642 ที่สร้างขึ้นข้างต้น ได้ชิ้นส่วนยีน *DUF642* เพื่อใช้เชื่อมต่อเข้ากับเวกเตอร์ pGEX ด้วยวิธี ligation independent cloning (LIC) ซึ่งหลักการ LIC คือ การอาศัยลำดับเบสที่เข้าคู่จำเพาะระหว่างชิ้นส่วนดีเอ็นเอกับเวกเตอร์ ถึง 17 ลำดับเบสในการเชื่อมต่อกันแบบจำเพาะ โดยวิธีขั้นแรกออกแบบไพรเมอร์ 5' และ 3' ให้มีส่วนปลายนอกเหนือบริเวณยีนที่ปลาย 5' และ 3' โดยลำดับเบสที่เพิ่มที่ปลาย 5' และ 3' คือลำดับเบสของเวกเตอร์บริเวณที่เชื่อมต่อกับชิ้นส่วนดีเอ็นเอ เอนไซม์ที่ใช้คือ T4 DNA polymerase มีสมบัติย่อยดีเอ็นเอ 3' – 5' exo-nuclease โดยเอนไซม์จะทำหน้าที่ย่อยปลาย 3' ทั้ง 2 สาย ของเวกเตอร์และของชิ้นส่วนดีเอ็นเอ และเมื่อหยุดปฏิกิริยาแล้วถ่ายพลาสมิดและชิ้นส่วนดีเอ็นเอ เข้าสู่เซลล์ *E. coli* บริเวณที่ถูกย่อยทั้งคู่จะสามารถจับกันได้อย่างจำเพาะเจาะจงและซ่อมแซมภายในเซลล์แบคทีเรีย แผนภูมิโครงสร้างของเวกเตอร์ที่สร้างขึ้นประกอบด้วย glutathione-s-transferase (GST) เชื่อมกับลำดับเบส *DUF642* ตามภาพที่ 12 และมีส่วนโปรโมเตอร์และระบบควบคุมการแสดงออกของยีนจาก Lac operon ซึ่งจะถูกเหนี่ยวนำด้วยการเติมสาร IPTG

3.1 การสร้างพลาสมิด การผลิตโปรตีน GST-DUF642-HA และการตัดโปรตีนด้วย thrombin

เนื่องจากทำงานของโปรตีนรีคอมบิแนนท์ *DUF642* ที่อาจถูกรบกวนจากโปรตีน GST ดังนั้นจึงได้ออกแบบการทดลองใหม่เพื่อแก้ไขปัญหา โดยการสร้างรีคอมบิแนนท์พลาสมิดใหม่ที่มีการเพิ่มโปรตีนติดตามขนาดเล็กชนิด HA tag เข้าที่ด้านปลาย C-terminal ของโปรตีน *DUF642* เพื่อให้สามารถกำจัดโปรตีน GST ด้วยเอนไซม์ thrombin protease และใช้ HA tag ในการตรวจสอบ (ภาพที่ 12)

การสร้างพลาสมิดที่ยีน *DUF642* เชื่อมต่อกับ HA tag ทำโดยออกแบบไพรเมอร์ด้านปลาย 3' ของยีน *DUF642* ทั้ง 10 ยีน ให้มีลำดับเบสสำหรับเป็นรหัสของเปปไทด์ HA จากนั้นนำไปเพิ่มปริมาณยีนด้วยวิธีพีซีอาร์โดยใช้เวกเตอร์ pGEX-DUF642 ทั้ง 10 ตัวเป็นต้นแบบ จากนั้นเชื่อมต่อกับชิ้นส่วนรีคอมบิแนนท์ดีเอ็นเอเข้ากับเวกเตอร์ pGEX ภาพที่ 8 แสดงแผนภูมิโครงสร้างของเวกเตอร์ ประกอบด้วย glutathione-s-transferase (GST) เชื่อมกับลำดับเบส *DUF642* ซึ่งต่อดำเนินลำดับเบสสำหรับรหัสเปปไทด์ HA tag



ภาพที่ 12 ภาพรีคอมบิแนนต์พลาสมิดในระบบ pGEX ประกอบด้วย ยีน glutathione-s-transferase (GST) เชื่อมต่อกับ ชิ้นส่วนยีน DUF642 เชื่อมต่อกับยีน HA และเมื่อแปลรหัสเป็นโปรตีนจะมีจุด Thrombin protease สำหรับตัดแยกระหว่างโปรตีน GST-tag กับ รีคอมบิแนนต์โปรตีน DUF642-HA

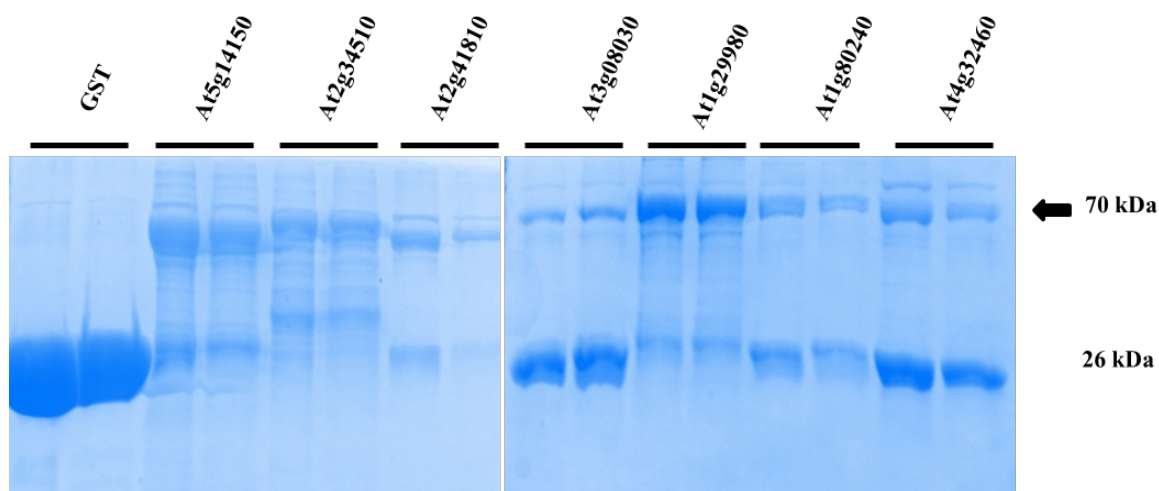
3.2 การเหนี่ยวนำผลิตโปรตีนและการทำรีคอมบิแนนต์โปรตีนให้บริสุทธิ์

จากการที่มีพลาสมิด pGEX-DUF642-HA ทั้ง 10 ยีน ถ่ายเข้าสู่เซลล์ *E. coli* เหนี่ยวนำผลิตโปรตีน ทำโปรตีนให้บริสุทธิ์ และแยกขนาดโปรตีนด้วยเจลโพลีอะครีลาไมด์ที่มีสาร SDS ผลการวิเคราะห์จากเจลโพลีอะครีลาไมด์ พบแถบด้านบน คือ โปรตีน DUF642 ด้าน N-terminus เชื่อมต่อกับโปรตีน GST tag และด้าน C-terminus เชื่อมต่อกับ โปรตีน HA มีขนาดรวมประมาณ 71 กิโลดัลตัน และแถบล่างคือโปรตีน GST tag อีกระยะมีขนาดประมาณ 26 กิโลดัลตัน ตามภาพที่ 13 ดังนั้นผลการแยกโปรตีนบนเจลนี้จึงเป็นการยืนยัน การผลิตรีคอมบิแนนต์โปรตีนที่ถูกต้อง โดยสามารถสกัดโปรตีนได้ความเข้มข้นประมาณ 200 ไมโครกรัม ต่อปริมาตรอาหารเลี้ยงเชื้อ 200 มิลลิลิตร

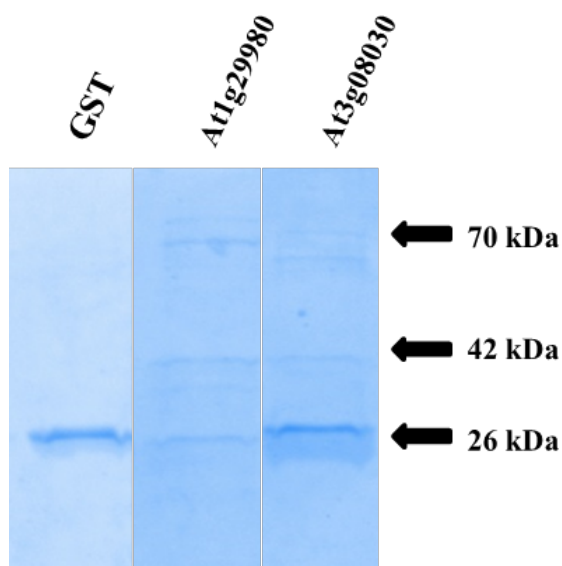
3.3 การตัด GST tag ด้วย thrombin

เมื่อนำโปรตีนที่ทำให้บริสุทธิ์มาทดสอบการตัดด้วยเอนไซม์ thrombin protease โดยการบ่มที่อุณหภูมิ 20 เซลเซียส เป็นเวลา 20 ชั่วโมง พบว่าสามารถตัด GST ออกจากรีคอมบิแนนต์โปรตีนได้ ภาพที่ 14 แสดงผลการตัดรีคอมบิแนนต์โปรตีน GST- DUF642-1-HA และ GST- DUF642-6-HA และตรวจสอบการแยกขนาดโปรตีนด้วย SDS-PAGE โดยแถบบนบนเป็นโปรตีน DUF642-1-HA และ DUF642-6-HA ซึ่งมีขนาดประมาณ 40-42 กิโลดัลตัน และแถบล่างเป็นส่วนของโปรตีน GST tag ซึ่งถูกตัดออกมา มีขนาดประมาณ 26 กิโลดัลตัน

หลังจากนั้น ตรวจสอบการแยกของขนาดโปรตีนด้วย SDS-PAGE และ ย้อมสีด้วย Coomassie blue พบว่า แถบบนบนของโปรตีน DUF642 มีขนาดประมาณ 39-44 กิโลดัลตัน และแถบล่างของโปรตีน GST tag จะมีขนาดประมาณ 26 กิโลดัลตัน และอาจมีแถบบนสุดบาง ๆ ของรีคอมบิแนนต์โปรตีนเชื่อมกับโปรตีน GST tag ขนาดประมาณ 70 กิโลดัลตัน เนื่องจากเกิดการตัดด้วยเอนไซม์ thrombin proteases ไม่หมด



ภาพที่ 13 การเหนี่ยวนำผลิตโปรตีนและการทำรีคอมบิแนนต์โปรตีนให้บริสุทธิ์ หลังจากนั้นตรวจสอบการแยกของขนาดโปรตีนด้วย SDS-PAGE และ ย้อมสีด้วย Coomassie blue ผลคือ แถบด้านบนของโปรตีน DUF642 เชื่อมติดกับโปรตีน GST tag และเชื่อมติดกับโปรตีน HA tag มีขนาดประมาณ 71 กิโลดัลตัน และแถบล่างของโปรตีน GST tag จะมีขนาดประมาณ 26 กิโลดัลตัน

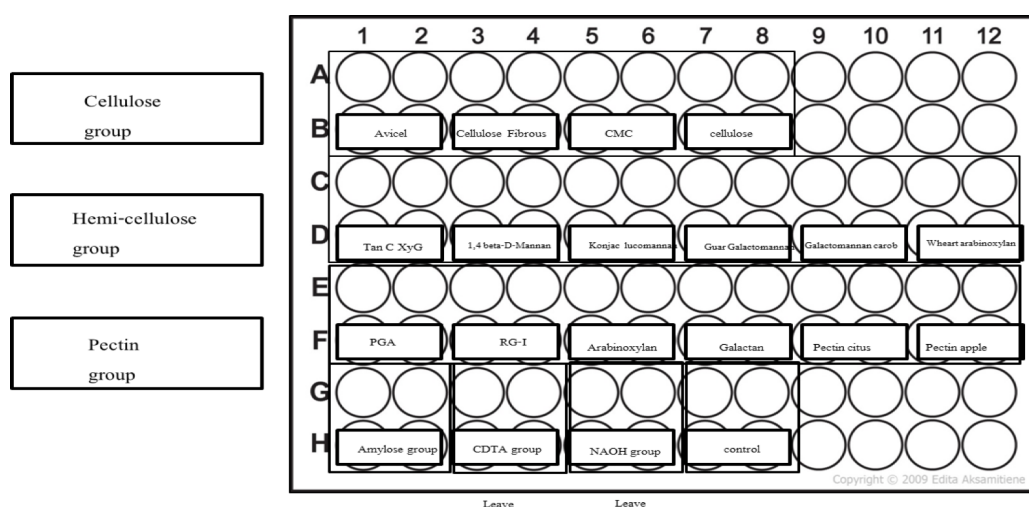


ภาพที่ 14 การตัดรีคอมบิแนนต์โปรตีน ด้วยเอนไซม์ Thrombin protease

4. การทดสอบการจับระหว่างโปรตีน DUF642-HA และโพลีแซคคาไรด์ด้วยวิธี ELISA

ตรวจสอบการเข้าจับระหว่างรีคอมบิแนนต์โปรตีน DUF642-HA กับโพลีแซคคาไรด์ ด้วยวิธี The enzyme-linked immunosorbent assay (ELISA) โดยแสดงผลเป็นค่าการจับกันระหว่างรีคอมบิแนนต์โปรตีน DUF642-HA กับโพลีแซคคาไรด์ และสามารถนำข้อมูลมาวิเคราะห์เป็นกราฟแสดงถึงระดับการจับระหว่างกันได้

โพลีแซคคาไรด์ที่นำมาทดสอบแบ่งออกเป็น 3 กลุ่ม คือ เซลลูโลส, เฮมิเซลลูโลส และ เพคติน รวมทั้งหมด 16 ชนิด ทั้งหมดนำมาละลายในน้ำกลั่นและหยดลงบนเพลท 96 ช่อง ชนิดละ 2 ปริมาณ คือ 0.5 ไมโครกรัม และ 0.05 ไมโครกรัม ตามภาพที่ 15



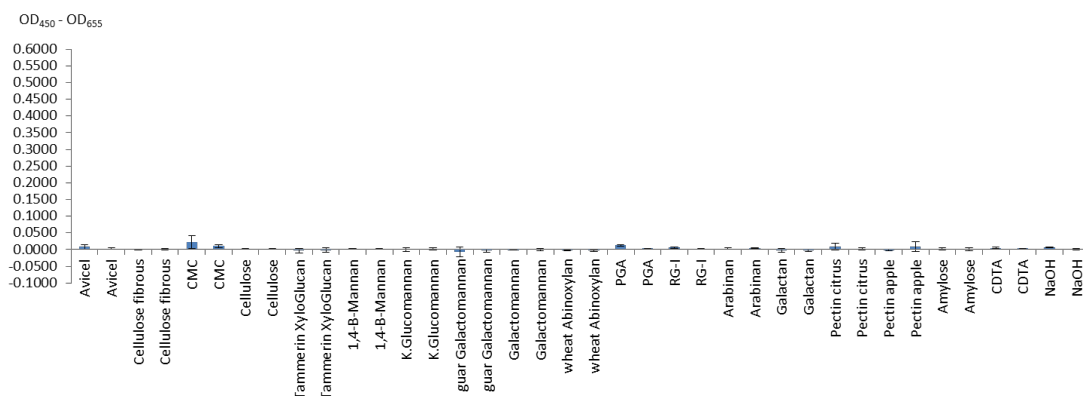
ภาพที่ 15 การเตรียมโพลีแซคคาไรด์ ลงบนเพลท 96 ช่อง ซึ่งโพลีแซคคาไรด์แบ่งออกเป็นกลุ่มเซลลูโลส, เฮมิเซลลูโลส และ เพคติน รวมทั้งหมด 16 ชนิด ที่ความเข้มข้น 0.5 ไมโครกรัม กับ 0.05 ไมโครกรัมโดยทำ 2 ซ้ำ

จากการทดลองตรวจพบสัญญาณการจับกันระหว่าง รีคอมบิแนนต์โปรตีน DUF642-HA และโพลีแซคคาไรด์ (ภาพที่ 16) โดยเปรียบเทียบกับผลการทดลองควบคุมที่ใช้โปรตีน GST ซึ่งไม่พบปฏิกิริยาการจับหรือมีการจับน้อยมากกับโพลีแซคคาไรด์ทั้งหมดที่นำมาทดสอบ สัญญาณที่ได้จากการทดลองนำมาสรุปเป็นตารางที่ 3

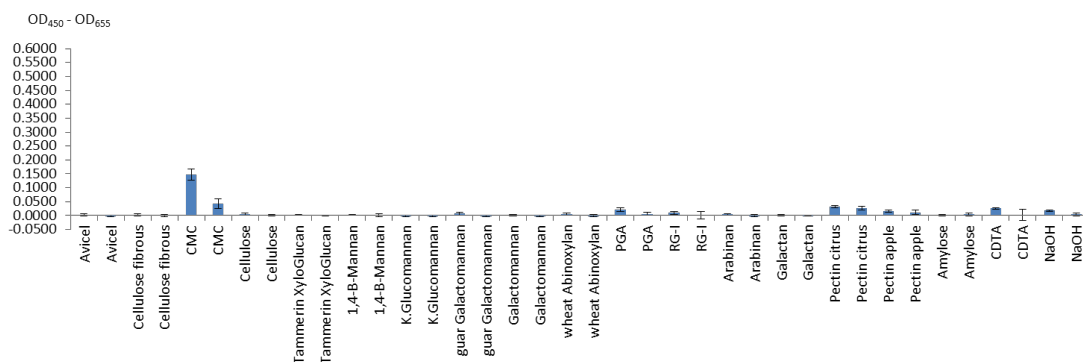
รีคอมบิแนนต์โปรตีน DUF642-1-HA, DUF642-2-HA, DUF642-3-HA, DUF642-6-HA, DUF642-7-HA ตรวจพบสัญญาณการจับกับโพลีแซคคาไรด์ กลุ่มเซลลูโลส ได้แก่ carboxy methyl cellulose และกลุ่มเพคติน ได้แก่ PGA, RG-I from soybean, pectin citrus, pectin apple

รีคอมบิแนนต์โปรตีน DUF642-4-HA, DUF642-5-HA จับกับโพลีแซคคาไรด์กลุ่มเซลลูโลส ได้แก่ carboxy methyl cellulose และกลุ่มเฮมิเซลลูโลส ได้แก่ wheat arabinoxylan และกลุ่มเพคติน ได้แก่ PGA, RG-I from soybean, pectin citrus, pectin apple รีคอมบิแนนต์โปรตีน DUF642-8-HA, DUF642-10-HA จับกับโพลีแซคคาไรด์กลุ่มเพคติน ได้แก่ PGA และสุดท้าย รีคอมบิแนนต์โปรตีน DUF642-9-HA จับกับโพลีแซคคาไรด์ทุกชนิดในกลุ่มโพลีแซคคาไรด์เซลลูโลส, กลุ่มโพลีแซคคาไรด์เฮมิเซลลูโลส, กลุ่มโพลีแซคคาไรด์เพคติน

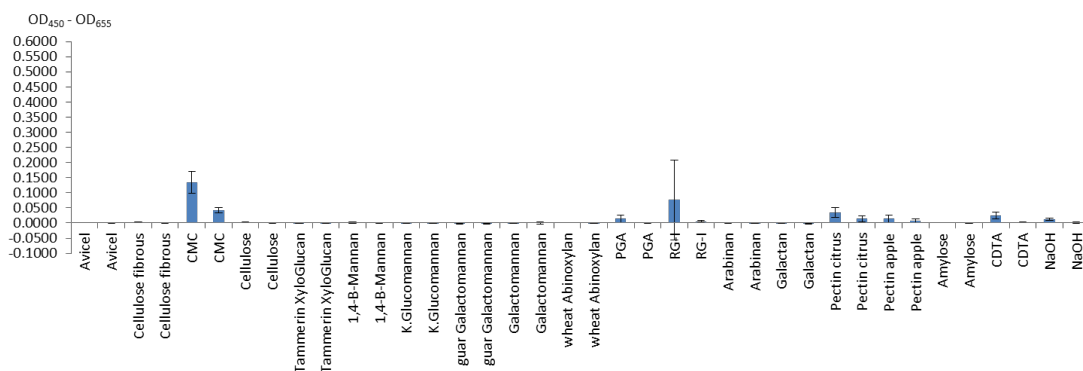
Control-GST



At1g29980 DUF642-1

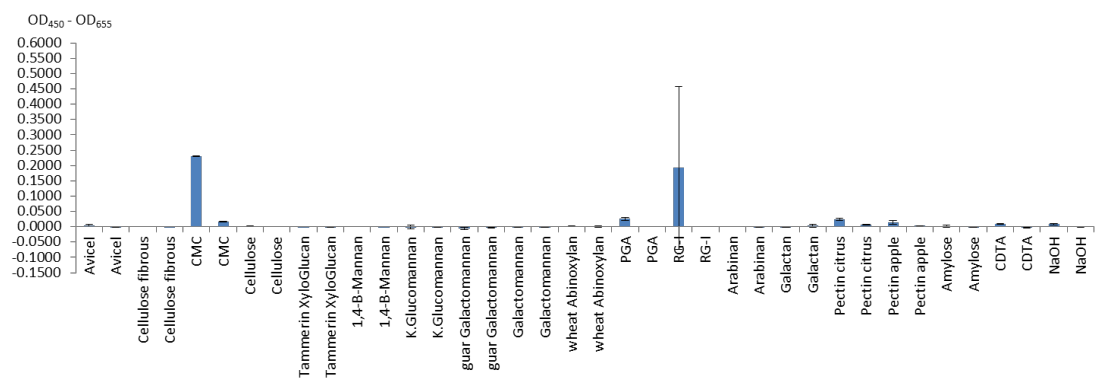


At1g80240 DUF642-2

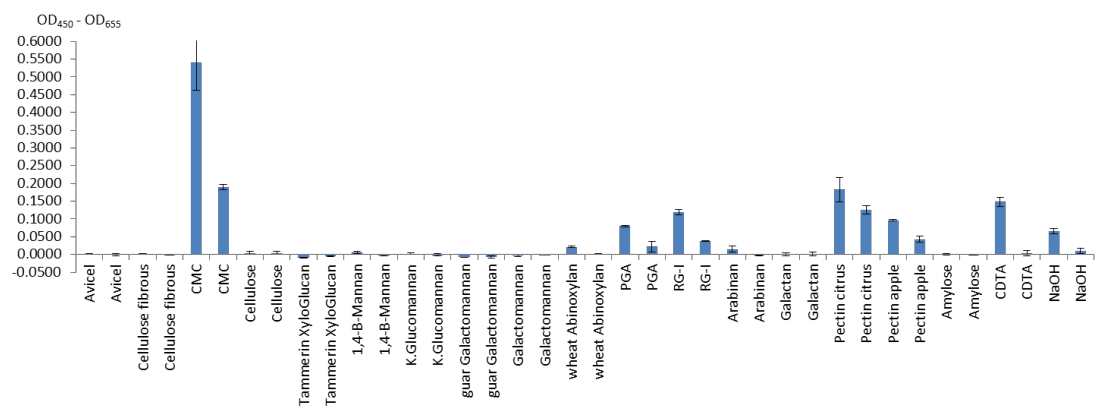


ภาพที่ 16 กราฟจากการทดลอง ELISA ซึ่งแสดงการเข้าจับกันระหว่าง รีคอมบิแนนต์โปรตีน GST-DUF642 กับโพลีแซคคาไรด์ ที่ความเข้มข้น 0.5 นาโนกรัม และ 0.05 นาโนกรัม

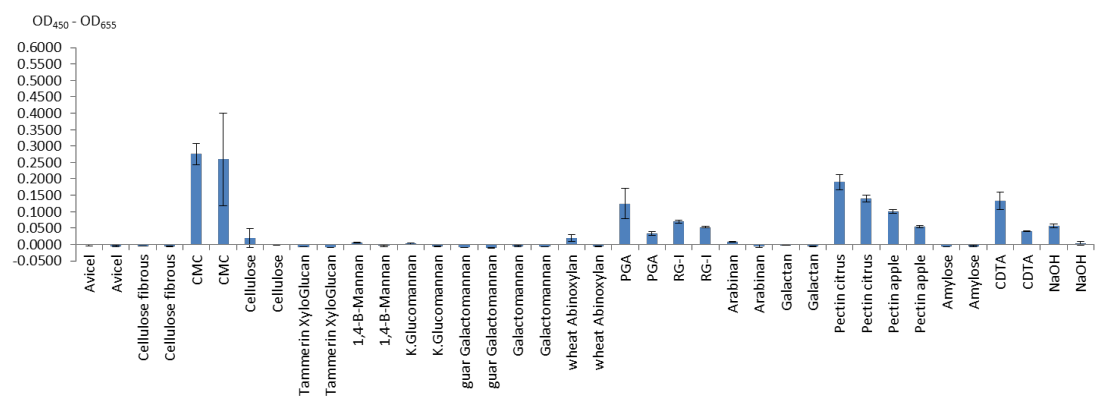
At2g34510 DUF642-3



At2g41800 DUF642-4

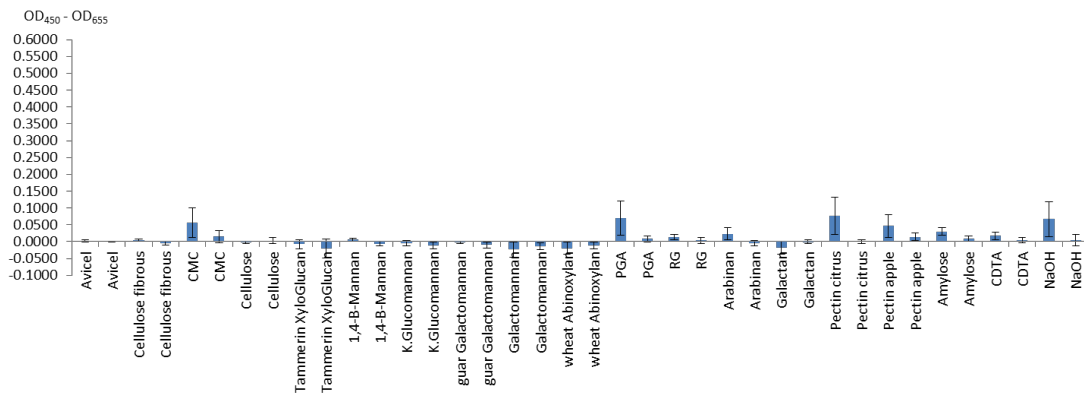


At2g41810 DUF642-5

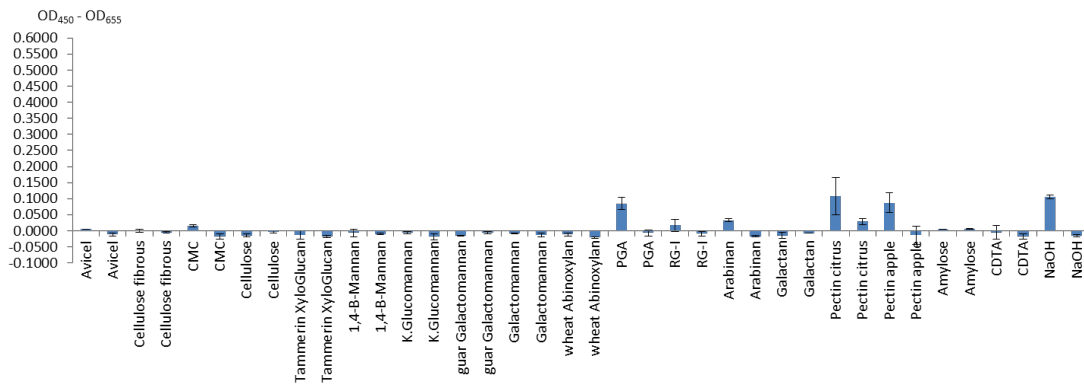


ภาพที่ 16 (ต่อ)

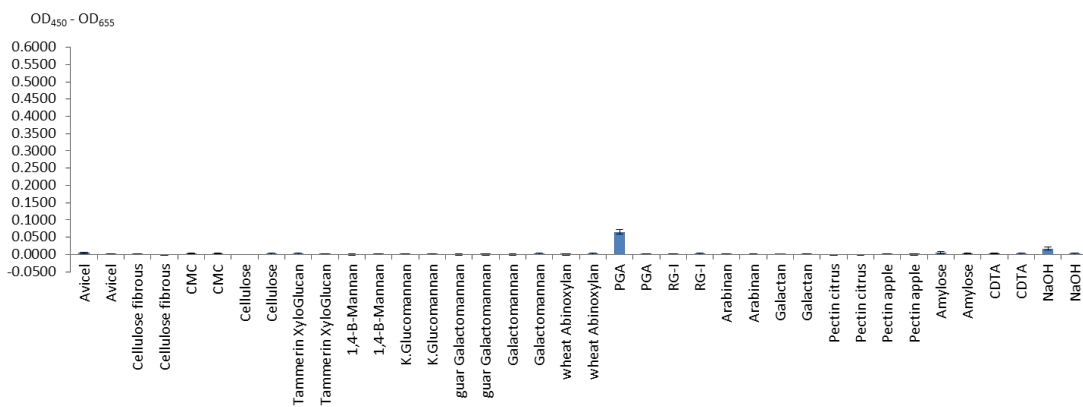
At3g08030 DUF642-6



At4g32460 DUF642-7

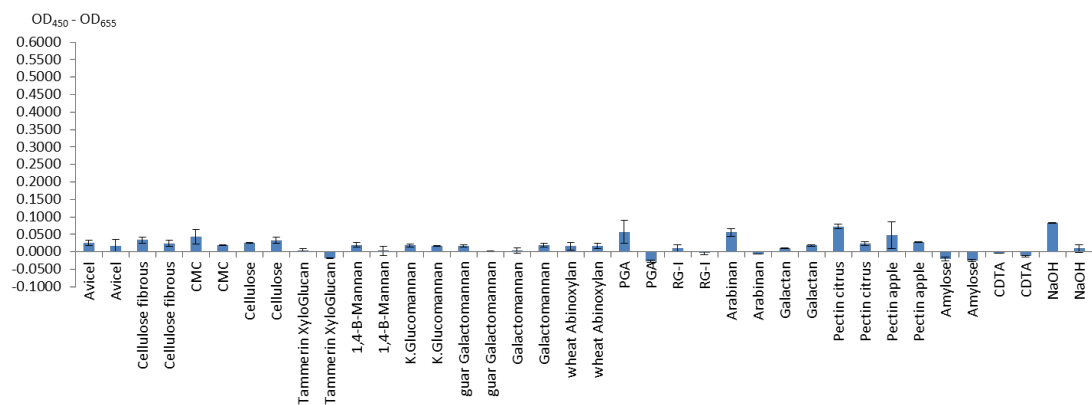


At5g11420 DUF642-8

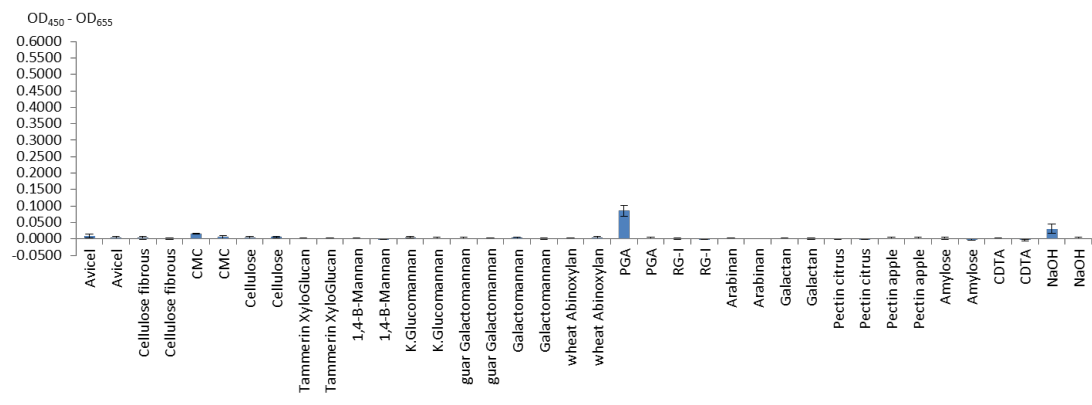


ภาพที่ 16 (ต่อ)

At5g14150 DUF642-9



At5g25460 DUF642-10



ภาพที่ 16 (ต่อ)

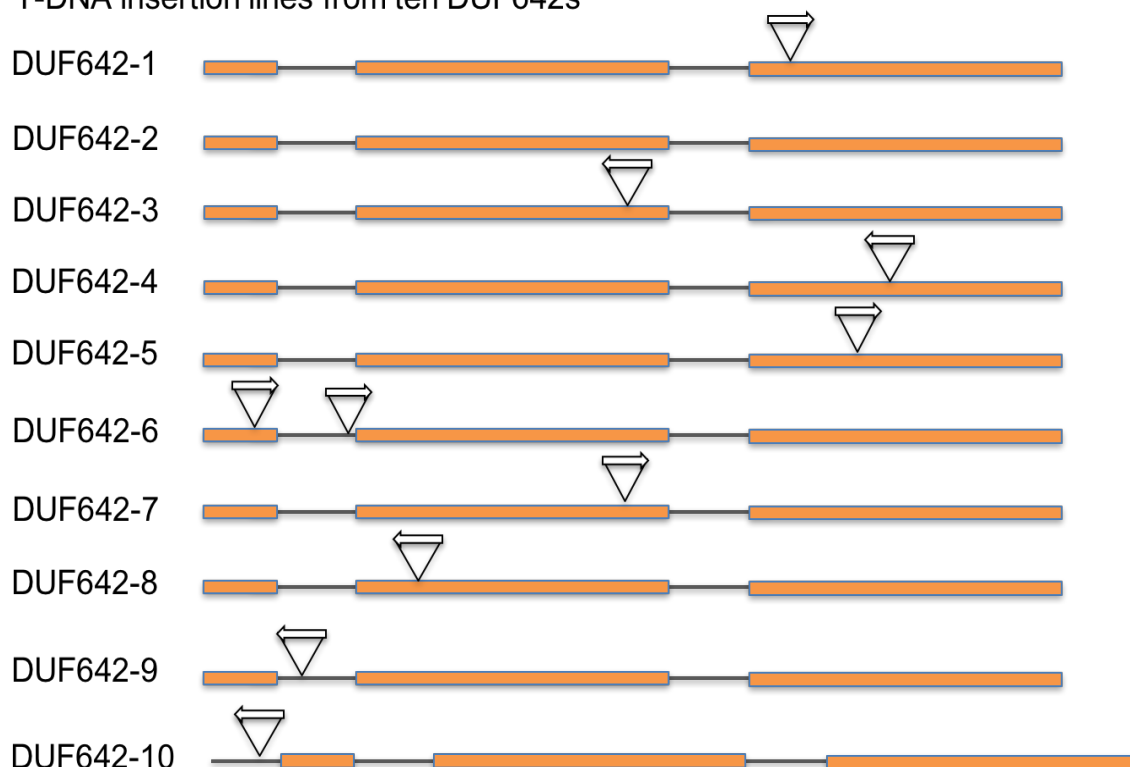
ตารางที่ 3 ข้อมูลสรุปปฏิกิริยาการจับระหว่างรีคอมบิแนนต์โปรตีน DUF642-HA ทั้ง 10 ตัว กับโพลีแซคคาไรด์ 17 ชนิด ทดสอบด้วย ELISA assays

	Cellulose groups			Hemicellulose group				Pectin group									
	Avicel	Cellulose Fibrous	Carboxy methyl cellulose	Cellulose	tamarind XyG	1,4- β -D- mannan	Konjac glucoman nan	Guar Galactom annan	Galactom annan	Wheat arabinoxyl an	PGA	RG-I from Soybean	Arabinoxyl an	Galactan	Pectin citrus	Pectin apple	Amylose
DUF642-1			+								+	+			+	+	
DUF642-2			+								+	+			+	+	
DUF642-3			+								+	+			+	+	
DUF642-4			+							+	+	+			+	+	
DUF642-5			+							+	+	+			+	+	
DUF642-6			+								+	+			+	+	
DUF642-7			+								+	+			+	+	
DUF642-8											+						
DUF642-9	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
DUF642-10											+						

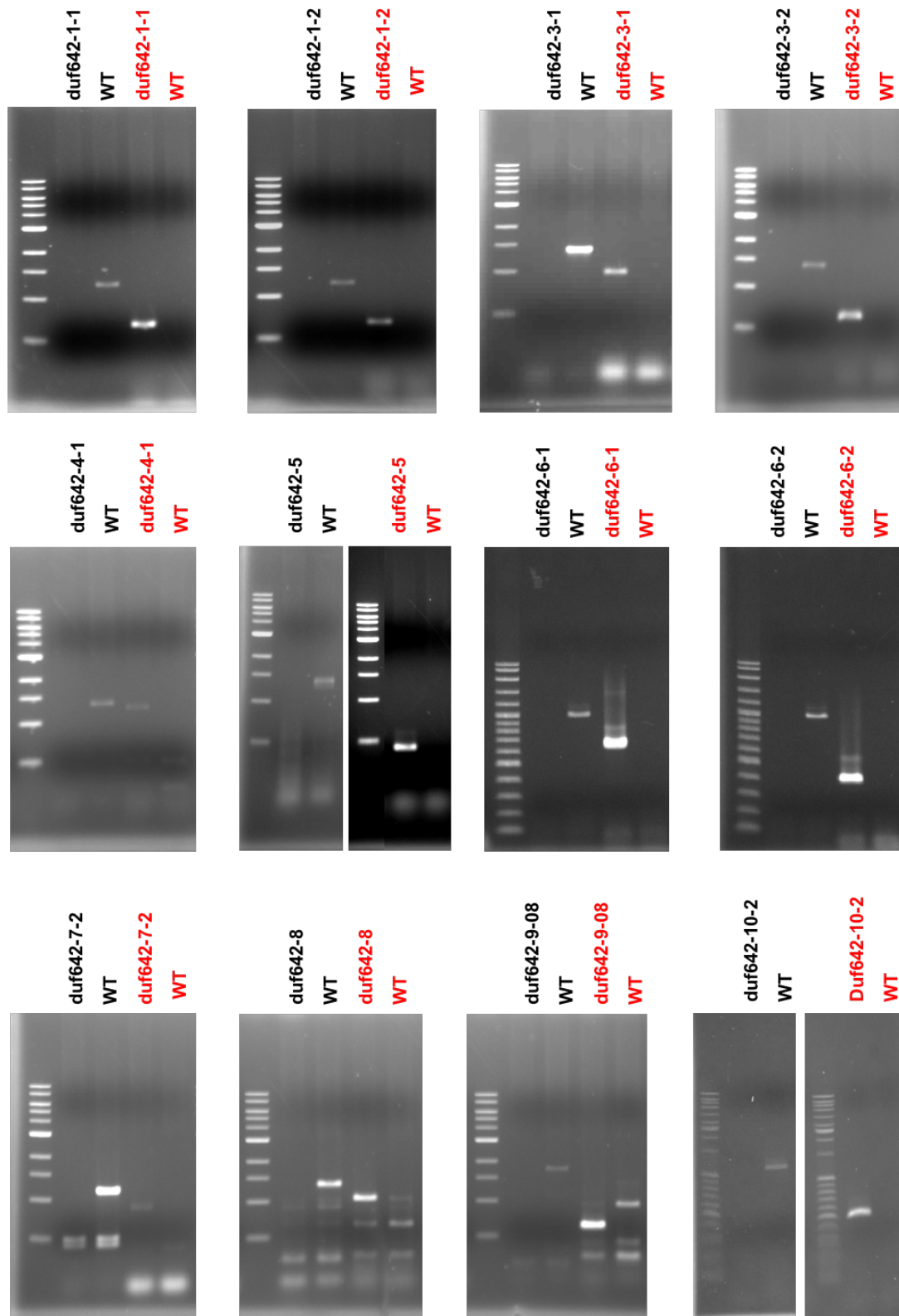
5. การตรวจสอบผลของการสูญเสียหน้าที่ในยีน *DUF642* ในต้นอะราบิโดพซิส

เพื่อตรวจสอบหน้าที่ของโปรตีน DUF642 ในพืชจึงได้สร้างอะราบิโดพซิสสายพันธุ์กลายชนิด T-DNA insertion line ของทั้ง 10 ยีนจาก www.arabidopsis.org จากการตรวจสอบฐานข้อมูลพบว่าจาก 10 ยีน มี 1 ยีน คือ *DUF642-2* (At1g80240) ที่ไม่มีสายพันธุ์กลาย ส่วน 9 ยีนที่เหลือมีพันธุ์กลายอย่างน้อย 1 สายพันธุ์ เมื่อได้เมล็ดพันธุ์กลายมาแล้วนำมาหาสายพันธุ์ที่เป็นโฮโมไซกัสและยืนยันตำแหน่งการกลายของยีนด้วยวิธีพีซีอาร์ ภาพที่ 17 แสดงตำแหน่งของ T-DNA ของแต่ละสายพันธุ์ ภาพที่ 18 แสดงผลยืนยันการได้สายพันธุ์ที่เป็นโฮโมไซกัสและมีตำแหน่ง T-DNA ที่ถูกต้อง และภาพที่ 19 แสดงผลยืนยันการสูญเสียหน้าที่ของยีนในระดับทรานสคริปชันด้วยวิธี RT-PCR อย่างไรก็ตามสายพันธุ์กลายของยีน *DUF642-10* ที่มีตำแหน่ง T-DNA insertion ที่ตำแหน่ง 5' UTR มีรูปแบบการทำงานของยีนแบบลดการแสดงออก หรือ knockdown

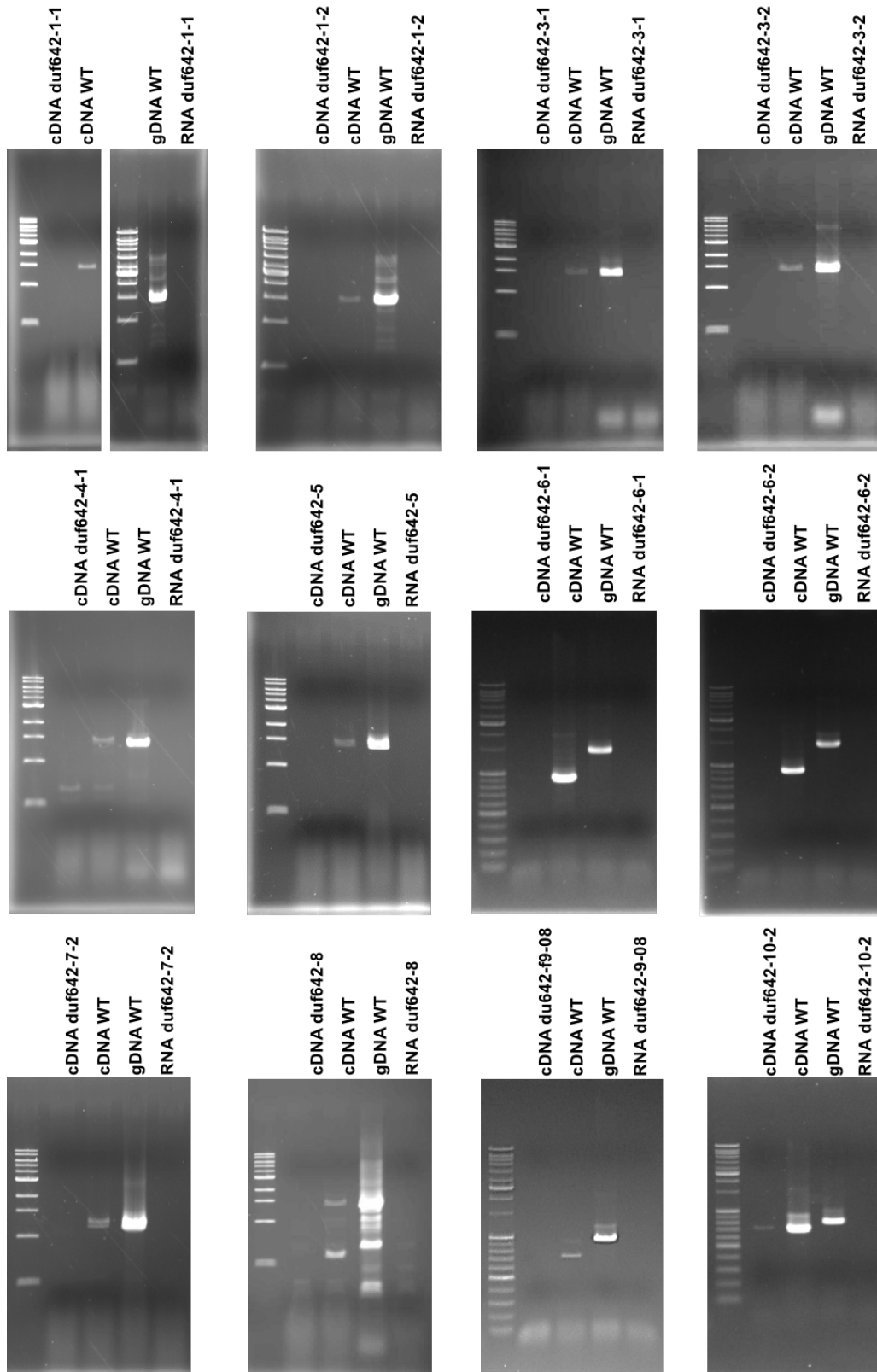
T-DNA insertion lines from ten DUF642s



ภาพที่ 17 ตำแหน่งของ T-DNA ของยีนแฟมิลี *DUF642* ทั้ง 10 ยีน



ภาพที่ 18 การยืนยันการเป็นโฮโมไซกัสของสายพันธุ์ T-DNA ของยีน *DUF642-1* – *DUF642-10* (ยกเว้น *DUF642-9*)

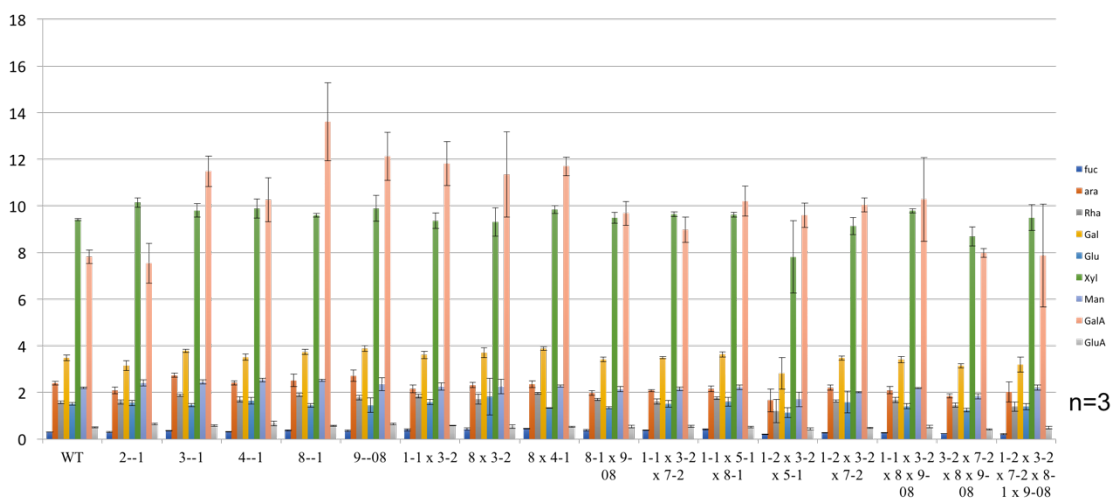


ภาพที่ 19 การตรวจสอบการแสดงออกของยีนในสายพันธุ์กลาย T-DNA insertion ด้วยวิธี RT-PCR

5.1 การตรวจสอบผลการสูญเสียหน้าที่ของยีน DUF642 เบื้องต้น

จากการตรวจสอบการเปลี่ยนแปลงทางฟีโนไทป์ขั้นต้นโดยการปลูกและตรวจสอบขนาดต้น รูปร่างใบ ความยาวลำต้น และการติดฝักและเมล็ด พบว่าสายพันธุ์กลายทั้งหมดไม่แตกต่างจากต้น wild type จากการวิเคราะห์ปริมาณน้ำตาลรวมของผนังเซลล์ที่ได้จากลำต้นของพันธุ์กลายด้วยวิธี HEAPC โดยใช้เครื่อง ion-exchange chromatography (ภาพที่ 20) และการทดสอบความสามารถในการย่อยสลายของชีวมวลด้วยเอนไซม์ (ภาพที่ 21) ก็ไม่พบความแตกต่างจากต้น wildtype อย่างมีนัยสำคัญ ทำให้คาดว่าหน้าที่ที่ไม่สามารถตรวจพบลักษณะทางฟีโนไทป์ที่เปลี่ยนแปลงไปน่าจะเกิดจากผลของ functional redundancy ซึ่งเกิดจากยีนแฟมิลี DUF642 มีจำนวนถึง 10 ยีน เมื่อทำลายหน้าที่ของยีนหนึ่ง ๆ ยีนที่เหลือสามารถทำหน้าที่แทนกันได้ และไม่สามารถตรวจวิเคราะห์หน้าที่ของยีนได้

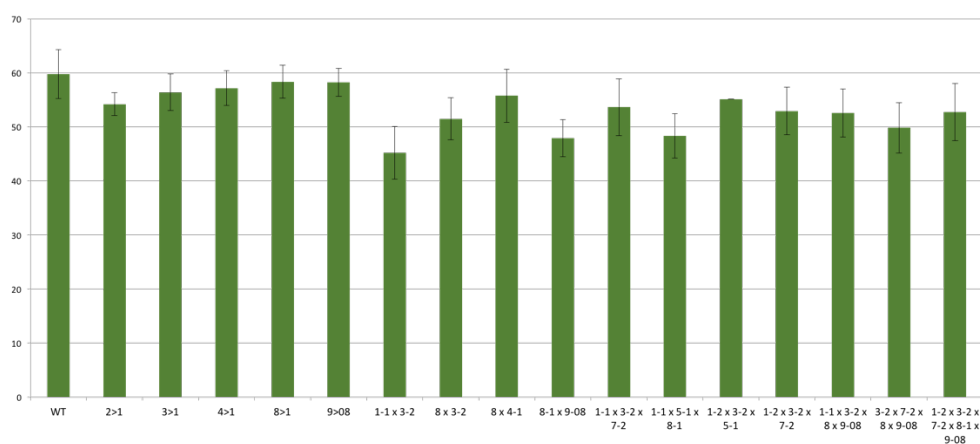
Cell wall monosaccharide analysis



TFA hydrolysis and analyzed by DIONEX HPAEC.

ภาพที่ 20 การวิเคราะห์องค์ประกอบน้ำตาลโมเลกุลเดี่ยวในผนังเซลล์ของสายพันธุ์กลาย

Cell wall saccharification- reducing sugars (nmoles/hour/mg DW)



Treatment: 0.5 M NaOH 90 C 20 min and enzymatic digestion at 50 C 8 hours using Celluclast and Novozyme 188

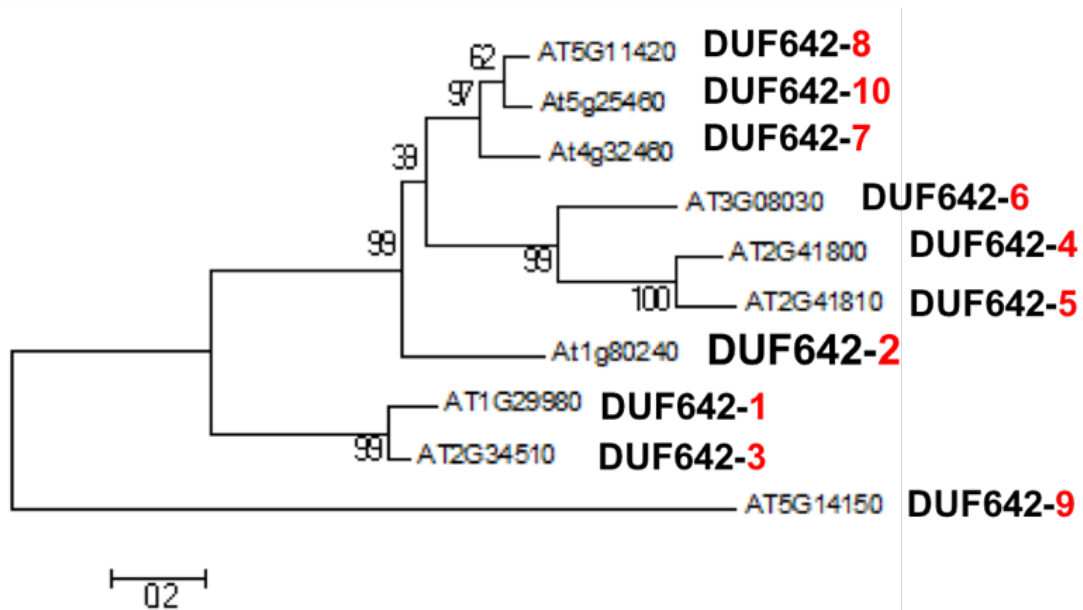
n=6

ภาพที่ 21 การวิเคราะห์ประสิทธิภาพการย่อยสลายผนังเซลล์ด้วยเอนไซม์

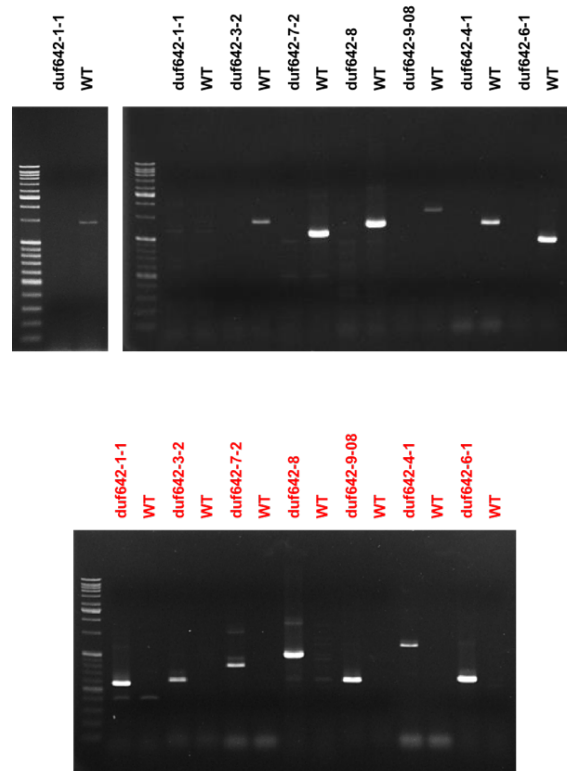
5.2 การตรวจสอบผลของการสูญเสียหน้าที่ในยีน *DUF642* พร้อมกันหลายตำแหน่ง

เนื่องจากผลของ functional redundancy จึงได้ทดลองผสมข้ามอะราบิโดพซิสสายพันธุ์กลายเข้าด้วยกัน เพื่อสร้างสายพันธุ์ที่มีการสูญเสียหน้าที่ของยีน *DUF642* พร้อมกันหลาย ๆ ยีน โดยเริ่มจับคู่ผสมระหว่างยีนที่มีความใกล้เคียงกัน ตามแผนภาพ phylogenetic tree (ภาพที่ 22) จากการผสมข้ามและคัดเลือกในรุ่น F2 หรือ F3 ด้วยวิธีพีซีอาร์ได้สายพันธุ์สูญเสียหน้าที่ของยีนหลากหลายตำแหน่งตามตารางที่ 4 โดยได้ตำแหน่งการกลายมากที่สุด 7 ตำแหน่งพร้อมกัน (สายพันธุ์หมายเลข 58 และ 65) โดยแสดงการยืนยันการกลาย 7 ตำแหน่ง จริงโดยการพีซีอาร์และ RT-PCR ตามภาพที่ 23 และ 24

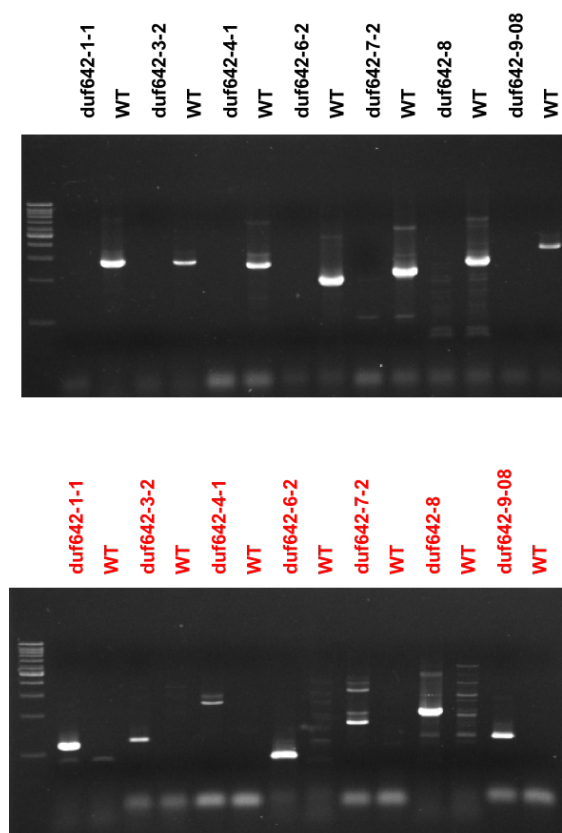
เมื่อทดสอบปลูกเปรียบเทียบระหว่างสายพันธุ์เปรียบเทียบกับ wild type พบทั้งหมดมีลักษณะไม่แตกต่างจาก wildtype ยกเว้น 1 สายพันธุ์ คือ *At1g22980* X *At2g41800* หรือต่อไปจะเรียกว่า *duf642-1* X *duf642-4* ที่มีลักษณะต้นเล็กกว่าปกติ จากการทดลองผสมพันธุ์กลายของทั้งหมด 9 ยีน แล้วมีเพียงยีน *DUF642-1* และ *DUF642-4* ที่มีการแสดงลักษณะที่เปลี่ยนแปลงไปในทางลบ คือ มีผลต่อการเจริญเติบโต แสดงให้เห็นว่ายีนแฟมิลี *DUF642* มีความเป็น functional redundancy สูง และยีน *DUF642-1* และ *DUF642-4* น่าจะเป็นยีนหลักของโปรตีนแฟมิลีนี้ที่ทำหน้าที่ในพืช



ภาพที่ 22 คู่ผสมของยีนสายพันธุ์กลายหลายตำแหน่ง



ภาพที่ 23 การยืนยันการเป็นโฮโมไซกัสของสายพันธุ์กลาย 7 ตำแหน่ง พันธุ์กลายหมายเลข 58



ภาพที่ 24 การยืนยันการเป็นโฮโมไซกัสของสายพันธุ์กลาย 7 ตำแหน่ง พันธุ์กลายหมายเลข 65

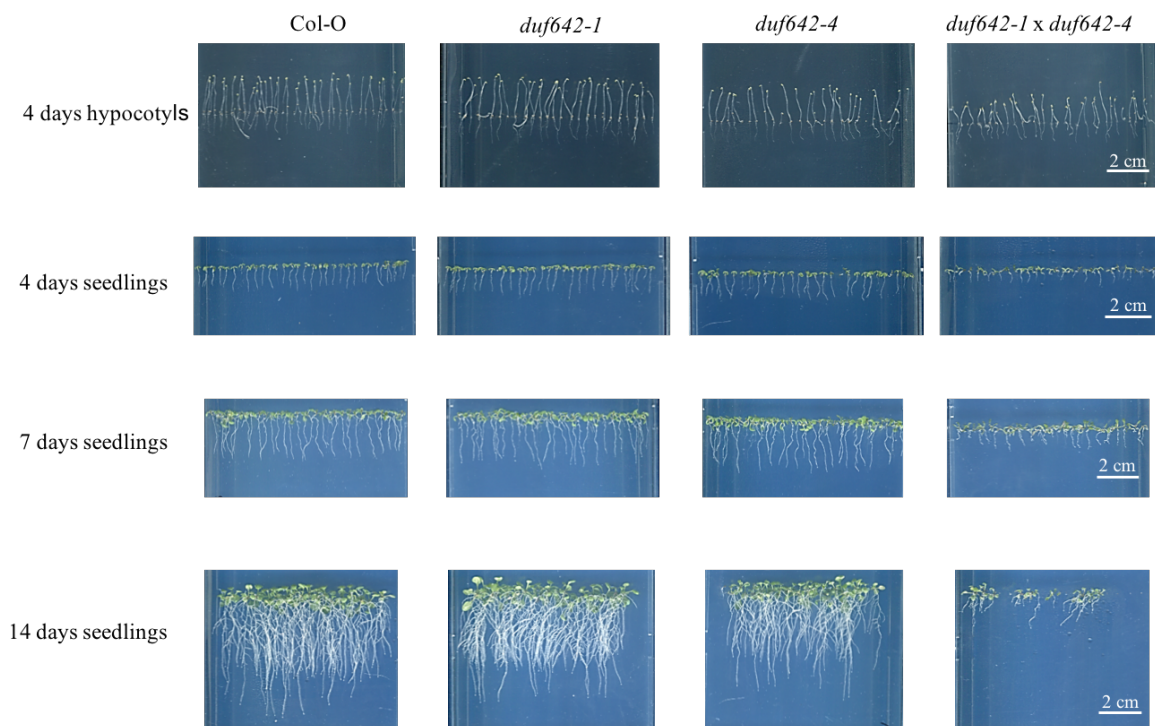
33				X					X	X		
34				X						X	X	
35						X				X	X	
36	X			X						X	X	
37	X								X	X	X	
38				X					X	X	X	
39	X			X					X	X	X	
40		X		X					X	X	X	
45												X
46					X		X					
48					X			X				
55	X			X					X	X	X	X
57	X			X		X			X	X	X	
58	X			X	X		X		X	X	X	
60	X				X							
61	X			X			X		X	X	X	
62	X				X		X		X	X	X	
63	X				X			X	X	X	X	
64	X				X				X	X	X	
65	X			X	X			X	X	X	X	
67					X					X	X	
68	X				X						X	
69	X				X			X		X	X	
70	X				X			X			X	

5.3 การวิเคราะห์ผลของการสูญเสียหน้าที่ของยีน *DUF642-1 X DUF642-4*

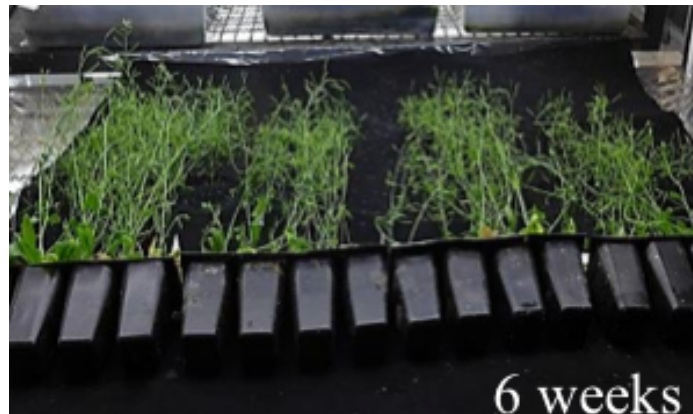
จากการทดสอบการปลูกเปรียบเทียบสายพันธุ์ *duf642-1 X duf642-4* ที่ระยะต้นอ่อน ระยะไฮโปคิไทล จนถึงระยะเจริญเต็มที่ พบว่ามีการเจริญเติบโตช้าและมีขนาดเล็กกว่า *wildtype* และต้นพันธุ์กลาย 1 ตำแหน่ง *duf642-1* และ *duf642-4* ภาพที่ 25 26 และ 27 โดยต้นกลายพันธุ์ 2 ตำแหน่งมีการเจริญที่ผิดปกติและโตช้ามากในระยะ 10-14 วันแรก ซึ่งเห็นอย่างชัดเจนเมื่อเปรียบเทียบกับสายพันธุ์อื่น ๆ และมีแนวโน้มที่จะไม่สามารถเจริญเติบโตต่อไปได้ อย่างไรก็ตามเมื่อย้ายไปปลูกบนดินกลับมีการพัฒนาการดีขึ้น และสามารถเจริญและพัฒนาเป็นต้นสมบูรณ์ได้ ได้ การเจริญหลังย้ายปลูกในดินอาจเกิดขึ้นเนื่องจากการแสดงออกของยีน *DUF642* อื่น ๆ และช่วยให้พันธุ์กลายนี้สามารถพัฒนาได้

จากข้อมูลทางฟีโนไทป์ที่ได้แสดงให้เห็นว่าโปรตีน *DUF642-1* และ *DUF642-4* มีความสำคัญต่อการเจริญเติบโต และการขยายขนาดของเซลล์ที่ปกติของอะราบิโดพซิส แต่เมื่อทดลองตรวจสอบยืนยันหน้าที่ของยีนด้วยวิธี complementation โดยการถ่ายโอนยีน *DUF642-1* และ *DUF642-4* เข้าสู่ต้นสายพันธุ์ *duf642-1 X duf642-4* กลับพบว่าไม่สามารถได้ลักษณะดั้งเดิมกลับคืนมา แสดงว่าการทำลายหน้าที่ของยีนทั้ง 2 ตำแหน่งนี้ ไม่ได้ให้ลักษณะทางฟีโนไทป์ที่ตรงพบในขั้นต้น ดังนั้น จึงไม่ได้ศึกษาสายพันธุ์กลาย 2 ตำแหน่งนี้ ต่อไป ลักษณะที่พบน่าจะเกิดจากกลายในยีนตำแหน่งอื่นที่ซ่อนอยู่ในต้นสายพันธุ์กลายนี้

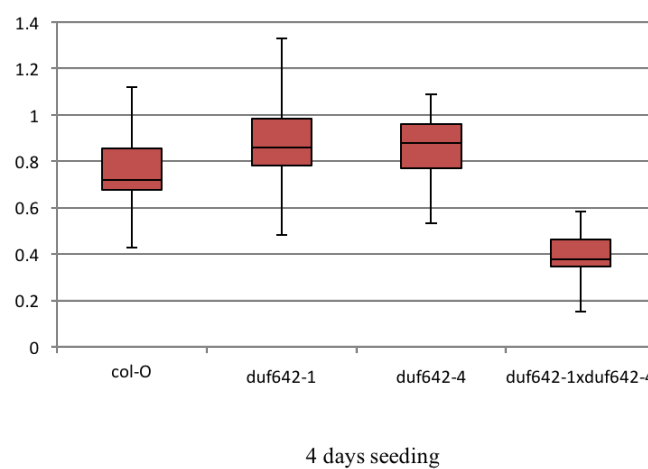
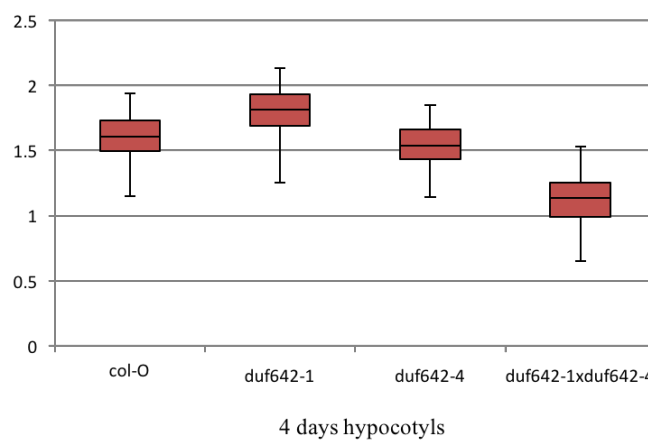
ทั้งนี้ ในโครงการได้ทดสอบลักษณะอื่น ๆ ของสายพันธุ์นี้ในรายละเอียดแล้ว เช่น ปฏิกริยาของเอนไซม์ในผนังเซลล์ องค์ประกอบของผนังเซลล์ และศึกษาทรานสคริปโทมของพันธุ์กลายนี้ แต่ว่าข้อมูลสายพันธุ์กลาย 2 ตำแหน่งที่ศึกษานี้ นั้นจะไม่ได้จัดอยู่ในการศึกษาหน้าที่ของโปรตีนแฟมิลี *DUF642* สายพันธุ์กลายนี้ยังมีคุณค่าสำหรับการศึกษายีนกลายที่ซ่อนอยู่เพื่อการบ่งชี้หน้าที่เกี่ยวกับการเจริญเติบโตของพืชต่อไปได้



ภาพที่ 25 การเจริญเติบโตของสายพันธุ์กลาย *duf642-1 X duf642-4* ระยะ hypocotyl และระยะต้นอ่อนเปรียบเทียบกับสายพันธุ์ *duf642-1*, *duf642-4* และ wildtype



ภาพที่ 26 การเจริญเติบโตระยะ 6 อาทิตย์ เปรียบเทียบ wildtype, *duf642-1*, *duf642-4* และ *duf642-1 X duf642-4* ตามลำดับ



ภาพที่ 27 ขนาดเปรียบเทียบระหว่างสายพันธุ์แสดงด้วย Box plot (n=20)

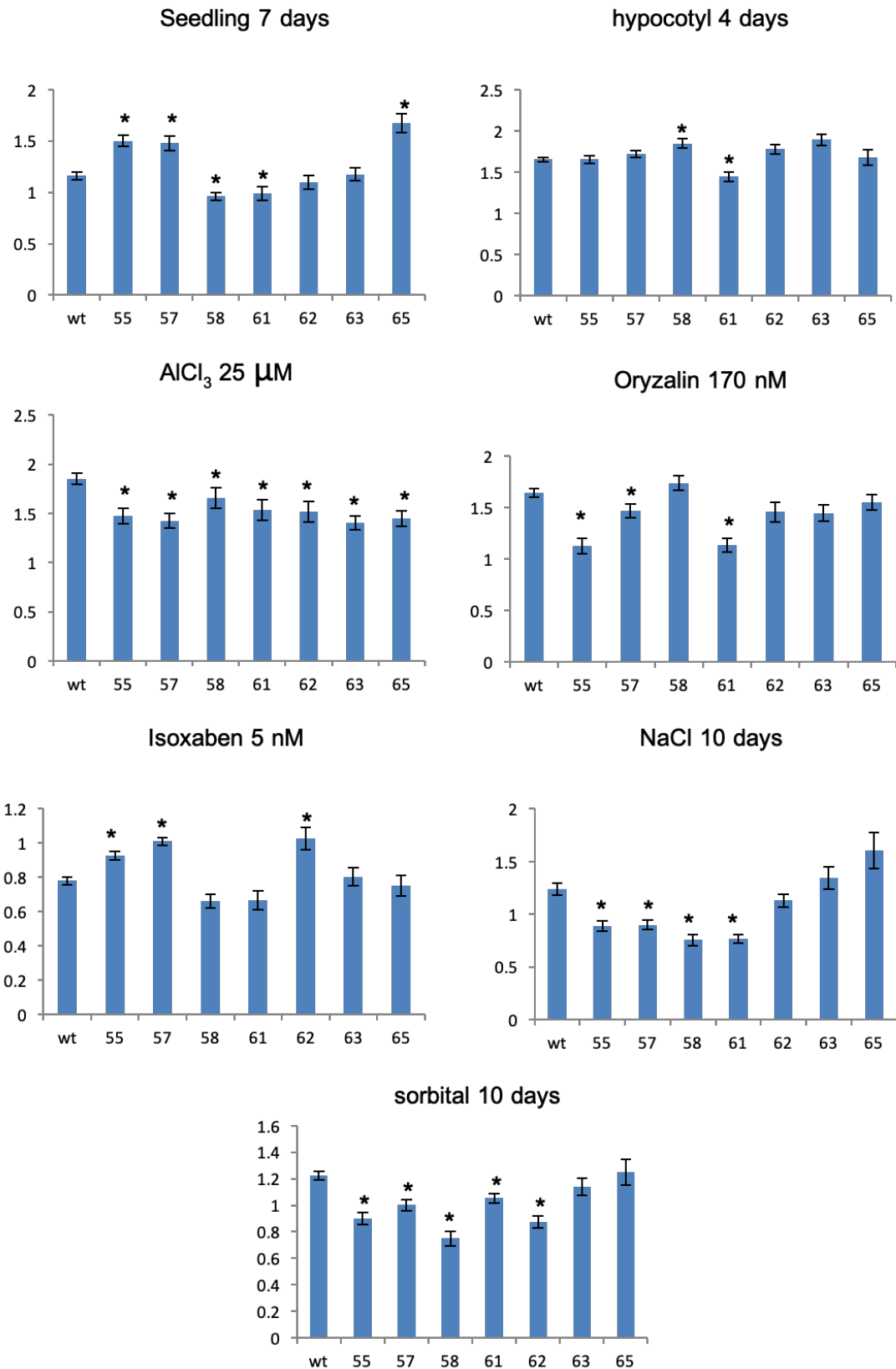
5.3 การวิเคราะห์ผลของการสูญเสียหน้าที่ของยีน *DUF642* จำนวน 6 และ 7 ตำแหน่ง พร้อมกัน

เนื่องจากการสูญเสียหน้าที่ของยีน *DUF642* จำนวน 1 หรือ 2 ตำแหน่งไม่อาจส่งผลให้เห็นความเปลี่ยนแปลง ทั้งนี้เนื่องมาจาก functional redundancy ดังนั้นจึงได้ตรวจสอบความเปลี่ยนแปลงทางฟีโนไทป์ของสายพันธุ์กลายที่มีการสูญเสียหน้าที่ของยีน *DUF642* พร้อมกันหลากหลายยีนโดยทดสอบการเจริญเติบโตระดับต้นอ่อน และการทดสอบกับสภาวะเครียดและสารยับยั้งการทำงานของผนังเซลล์พืช อย่างไรก็ตามจากการทดสอบสายพันธุ์ที่สร้างขึ้นจำนวนมากก็ยังไม่สามารถตรวจพบความเปลี่ยนแปลงทางฟีโนไทป์ได้

ในส่วนนี้จะขอแสดงผลของสายพันธุ์กลายที่มีการสูญเสียหน้าที่ของยีน จำนวน 6 และ 7 ยีน ได้แก่ สายพันธุ์กลายหมายเลข 55, 57, 58, 61, 62, 63 และ 65 โดยการตรวจสอบการเจริญของปลายราก ความยาวของไฮโปคอทิล และการเจริญที่ตอบสนองต่อสภาวะเครียดด้วยเกลือ sorbitol และสารยับยั้งการทำงานของผนังเซลล์พืช ได้แก่ Isoxaben และ oryzalin ที่ยับยั้งการสังเคราะห์เซลลูโลส และ AlCl_3 ที่ยับยั้งการทำงานของเพกติน

ทั้งการพิจารณาผลการทดลองนี้จะมุ่งไปที่สายพันธุ์หมายเลข 58 และ 65 ที่มีการสูญเสียหน้าที่ของยีน 7 ยีน จากการตรวจสอบการเจริญของต้นอ่อนด้วยสารทั้งหมด ยกเว้นการทดสอบกับสาร AlCl_3 พบว่ามีการเปลี่ยนขนาดความยาวรากเพียงเล็กน้อย ไม่มีรูปแบบที่ชัดเจน และผลที่ได้จากสายพันธุ์หมายเลข 58 และ 65 ไม่ได้ไปในทางเดียวกัน แต่การทดสอบด้วยสาร พบว่าทุกสายพันธุ์ที่ทดสอบมีขนาดของรากที่สั้นลงอย่างมีนัยสำคัญเมื่อเปรียบเทียบกับขนาดที่ได้จากทดสอบต้น wild type

จากผลการทดลองนี้แสดงให้เห็น 2 ส่วนคือ 1) โปรตีน *DUF642* น่าจะมีความเกี่ยวข้องกระบวนการขยายขนาดของเซลล์ที่เกี่ยวข้องกับเพกติน และ 2) ยังไม่สามารถสรุปได้ว่าโปรตีน *DUF642* จะเกี่ยวข้องกับการขยายขนาดเซลล์ผ่านองค์ประกอบของผนังเซลล์ด้านอื่น ๆ หรือไม่ เนื่องจากสายพันธุ์กลาย 7 ตำแหน่ง ยังมียีน *DUF642* เหลืออีก 3 ยีนที่ยังทำงานได้

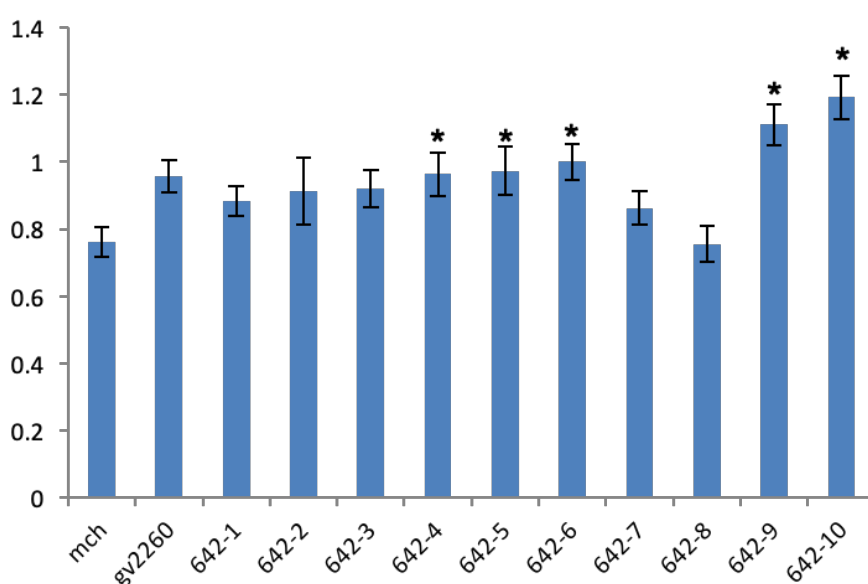


ภาพที่ 28 การตรวจสอบลักษณะทางฟีโนไทป์ของพันธุ์กลายของยีน DUF642 พร้อมกัน 6 และ 7 ตำแหน่ง (อ้างอิงสายพันธุ์ ตามตารางที่ 4) แกน Y ทั้งหมดมีหน่วยเป็นเซนติเมตร เครื่องหมาย * แสดงความแตกต่างอย่างมีนัยสำคัญที่ $P < 0.05$ ($n \geq 20$)

6. การศึกษาผลของโปรตีนแฟมิลี DUF642 ต่อการเจริญและพัฒนาของต้นอ่อนอะราบิโดพซิส

ยีนสมาชิกของแฟมิลี DUF642 ทั้ง 10 ยีน นำมา ถ่ายโอนยีนเพื่อแสดงออกแบบชั่วคราว กับต้นอ่อนของอะราบิโดพซิส แล้วตรวจสอบผลของโปรตีนต่อการเจริญของต้นอ่อนในระยะ 7 วัน มีตัวควบคุมโดยการถ่ายโอนยีน mCherry และการใช้เฉพาะเซลล์อะโกรแบคทีเรียสายพันธุ์ GV2260

หลังจากทดสอบการแสดงออกชั่วคราวจำนวน 10 ยีน ได้ตรวจสอบผลต่อการเปลี่ยนแปลงการเจริญเติบโตของต้นอ่อนพบว่า สามารถเพิ่มความยาวของรากได้อย่างมีนัยสำคัญจากการถ่ายยีน 5 ยีน และมีแนวโน้มที่จะเพิ่มความยาวรากได้อีก 3 ยีน และไม่มีการเปลี่ยนแปลงความยาวราก จำนวน 2 ยีน เมื่อเปรียบเทียบกับ การถ่ายโอนยีน mCherry จากผลดังกล่าวแสดงให้เห็นว่าการแสดงออกที่เพิ่มขึ้นของยีน DUF642 น่าจะส่งผลไปในทางการกระตุ้นความยาวของรากให้เพิ่มขึ้น อย่างไรก็ตามการทำงานของยีน DUF642 แต่ละยีนอาจมีการงานที่แตกต่างกัน จึงทำให้บางยีนไม่ส่งผลเพิ่มความยาวของรากได้



ภาพที่ 29 การตรวจสอบการเจริญโดยการแสดงออกของยีน DUF642-1-10 แบบชั่วคราวด้วยเทคนิค Agrobast เครื่องหมาย * แสดงความแตกต่างอย่างมีนัยสำคัญที่ $P < 0.05$ ($n \geq 20$)

7. การตรวจสอบการเข้าจับของโปรตีน DUF642 กับโปรตีนชนิดอื่น ๆ ในผนังเซลล์

เนื่องจากก่อนหน้านี้มีรายงานว่าโปรตีน At5g11420 และ At4g32460 สามารถเข้าจับกับเอนไซม์ PME (pectic methylesterase protein; Zuniga-Sanchez and Gamboa de Buen, 2012) และจากการทำวิจัยร่วมกับ Professor Simon McQueen-Mason (University of York, UK) ได้ข้อมูลว่าในระหว่างการแยกโปรตีน expansin บริสุทธิ์จากเนื้อเยื่อพืชผักมีแถบโปรตีนชนิดหนึ่งปะปนมาด้วย และเมื่อนำแถบโปรตีนไปวิเคราะห์ลำดับกรดอะมิโนพบว่าโปรตีนกลุ่ม DUF642 จึงมีสมมุติฐานว่าโปรตีน DUF642 นี้จะทำงานเกี่ยวข้องกับการจับกับโปรตีนหรือเอนไซม์ในผนังเซลล์ พร้อมกับเข้าจับกับโพลีแซคคาไรด์ชนิดต่าง ๆ โดยเป็นตัวกลางการนำพาระหว่างเอนไซม์ในผนังเซลล์และโพลีแซคคาไรด์เป้าหมาย นอกจากนี้ยังมีความเป็นไปได้ว่าโปรตีน DUF642 น่าจะสามารถจับกับโปรตีนหรือเอนไซม์ในผนังเซลล์ชนิดอื่น ๆ ได้อีก ดังนั้นการทดลองนี้จึงเป็นการตรวจสอบหาโปรตีนชนิดอื่น ๆ ในผนังเซลล์ที่สามารถเข้าจับกับโปรตีน DUF642

7.1 การตรวจสอบการเข้าจับของโปรตีน DUF642 กับโปรตีนชนิดอื่นด้วยวิธี Proteomic analysis

การทดลองนี้ใช้โปรตีน DUF642 ทั้ง 10 ชนิด ที่ถูกยึดอยู่บนเม็ดเจลสำหรับการทำโปรตีนให้บริสุทธิ์ (immobilized-protein) นำมาผสมและบ่มกับโปรตีนสกัดรวมจากผนังเซลล์อะราบิโดพซิส จากนั้นล้างโปรตีนที่ไม่จับออก และนำโปรตีนที่ได้มาวิเคราะห์โปรตีนด้วย LC/MS/MS ทั้งนี้ใช้ตัวควบคุมเป็นโปรตีน GST สกัดจาก *E. coli*

ผลการทดลองแสดงในตารางที่ 5 พบว่า เมื่อเปรียบเทียบกับโปรตีนที่ตรวจพบในตัวควบคุมที่ใช้ GST เป็นตัวจับ พบว่ามีจำนวนชนิดโปรตีนน้อยกว่าที่ตรวจพบกับโปรตีน DUF642 ทั้ง 10 ชนิด และโปรตีนที่ตรวจพบกับโปรตีน DUF642 มีจำนวนมากกว่า 50 ชนิด โดยส่วนใหญ่พบเป็นโปรตีนที่มีความสำคัญในการควบคุมการทำงานของผนังเซลล์ เช่น expansin, xyloglucan endotransglycosylase และเอนไซม์ polysaccharide hydrolase ซึ่งรวมเรียกว่า cell wall modifying enzymes อย่างไรก็ตามจากการทดลองยังมีโปรตีนชนิดอื่น ๆ จากอะราบิโดพซิสจำนวนมากที่ตรวจสอบพบเนื่องด้วยประสิทธิภาพของเครื่อง mass spectroscopy ทำให้ยังไม่สามารถยืนยันได้แน่ชัดว่าโปรตีนที่ตรวจพบเหล่านี้ทุกชนิดจะเข้าจับกับโปรตีน DUF642 ได้จริง ดังนั้น ต้องมีการทดลองยืนยันผลเพิ่มเติมต่อไป

จากการทดลองนี้สามารถสร้างสมมุติฐานได้ว่าโปรตีน DUF642 น่าจะทำงานในผนังเซลล์โดยการเข้าจับกับเอนไซม์กลุ่ม cell wall modifying enzymes และนำพาไปยังโพลีแซคคาไรด์เป้าหมาย

ตารางที่ 5 การวิเคราะห์โปรตีนที่เข้าจับกับโปรตีน DUF642 ตัวเลขแสดงจำนวนการอ่านพบโปรตีนชนิดต่าง ๆ

	DUF-1	DUF-2	DUF-4	DUF-5	DUF-6	DUF-7	DUF-10
rbcl	31	36	25	31	31	35	34
GSTU19	12	14	17	15	12	25	18
XYL1	18	18	17	16	17	20	18
BGLU44	12	16	10	13	9	26	10
RBCS-1A	11	9	10	7	9	11	10
RBCS-2B	12	11	9	8	8	8	11
PAE7	12	15	13	13	12	4	16
PME18	10	9	10	8	10	20	7
CWINV1	14	15	10	12	10	18	13
MS1	14	19	12	8	9	5	11
HEXO3	12	10	9	9	7	5	7
FBA2	12	13	5	9	10	5	15
GER3	5	6	6	6	6	6	4
LAC7	8	4	8	12	7	4	10
CRA1	4	9	6	5	8	6	6
CRC	7	4	3	10	5	9	5
PER57	4	7	4	8	7	5	8
At2g43610	5	3	2	3	3	20	2
PME17	5	6	6	7	4	7	8
SBT1	7	7	1	3	2	15	6
PER72	3	3	6	2	3	2	8
TCTP1	5	4	4	4	5	5	3
RPL6B	4	4	3	5	3	6	4
PER12	4	6	6	4	5	1	5
GLL23	3	4	6	3	4	8	5
BXL7	2	3	1	3	0	18	2
GAPC1	8	10	1	5	2	0	4
atpA	4	4	5	3	3	2	3
PME3	5	4	3	3	5	1	7
BGLU23	6	4	3	3	2	2	4
GDH2	2	3	3	5	6	5	2
SBT11	4	3	5	2	2	1	4
BCA1	8	7	1	4	5	0	4
NIT1	5	3	2	2	4	4	4

CPN60A1	3	3	2	4	2	9	1
TGG4	2	0	4	0	3	2	6
AGP31	1	4	2	4	1	3	2
JAL34	3	4	2	2	3	4	4
SBT4	2	1	2	0	2	8	1
LOS1	5	3	0	2	4	3	1
EXPA3	1	2	2	1	2	7	2
GAPA1	3	1	0	8	2	0	3
BGAL10	3	2	2	4	2	2	4
GLP4	3	4	2	2	2	0	3
PER27	1	5	2	2	2	0	5
GSTF10	2	2	2	2	0	6	3
XTH4	0	1	0	0	0	10	1
RCA	3	2	1	2	2	0	2
APA1	1	4	1	3	0	0	3
SCPL25	2	1	1	3	1	0	4
LHCB1	1	1	3	1	4	0	1
SCPL34	4	0	3	1	2	4	1
PCKA	1	1	0	1	1	8	1
GLP10	1	3	2	3	1	0	2
GSTF5	1	1	2	2	2	3	3
A1	2	2	3	3	3	0	3
EXL4	0	1	0	1	0	7	1
NDPK3	2	2	2	1	2	4	0
CSP41B	2	3	1	3	3	0	1
FBA3	2	0	1	3	0	0	2
PGIP2	2	2	1	1	3	0	2
SCPL24	2	2	2	3	2	0	2
ACLB-1	2	2	1	1	2	2	1
PME2	1	2	2	1	2	0	3
RPS15AA	1	2	1	3	1	0	2
RPS12A	3	3	0	1	1	0	1
CAT2	0	3	2	2	1	0	2
EXL2	0	0	0	0	0	5	0
NSP1	1	0	1	0	0	4	0
PER69	0	0	0	0	0	8	0
GLP1	1	1	1	1	0	2	2

PRS4	0	1	0	2	2	2	1
RGP1	1	1	3	1	1	0	2
EXPA1	0	1	0	1	1	5	1
PER45	0	0	0	0	0	7	0
BGAL1	0	0	0	0	0	8	0
BGAL4	0	0	0	0	0	8	0
BXL1	0	0	0	0	0	8	0
CRB	0	1	1	3	1	3	0
GSTU20	0	0	0	0	1	5	1
NADP-ME2	0	0	0	0	0	8	0
RD21A	2	4	0	0	0	1	1
PGK1	0	1	1	1	1	0	2
PER30	1	1	1	2	0	2	1
AGT1	1	0	0	1	2	0	2
WHY3	0	1	0	2	0	1	1
EXPA11	0	0	0	0	0	4	1
GSTU16	0	0	0	0	0	7	0
At4g24780	0	1	0	0	0	5	0
MSD2	0	1	0	1	0	5	0
PLAT2	2	2	2	0	0	0	2
NDK1	1	2	1	1	0	0	2
ADPG2	2	0	0	0	0	4	0
AIR3	1	1	0	0	0	4	0
GSTU24	0	1	0	0	0	4	0
MDH1	0	0	0	0	0	6	0
LTP2	2	1	1	0	0	2	1
PGIP1	1	2	2	1	0	0	1
GLO2	1	2	0	1	0	0	1
At5g07830	0	0	0	0	0	4	0
PAP85	0	1	0	0	0	4	0
DEGP2	0	0	0	3	0	2	0
LTP1	0	1	1	1	2	0	1
RD21B	1	1	0	0	0	2	1
HSP90-2	1	2	0	2	0	0	0
ACA1	1	2	1	1	1	0	0
RPS20A	0	3	0	1	1	0	0
BAS1	0	1	0	1	0	3	1

PRXQ	1	0	0	1	0	3	0
TKL-1	3	2	0	0	0	0	0
DIR6	0	0	1	0	1	0	2
GSTU27	0	0	0	0	0	2	0
RPL14A	2	2	0	1	0	0	0
RPL12A	0	3	1	0	0	0	0
NADP-ME4	1	0	0	0	0	2	0
WHY1	0	0	0	0	2	2	0
XTH7	0	0	0	0	1	0	2
PSBO1	1	1	0	0	0	0	2
ENO2	0	0	0	1	0	2	0
GLDP1	3	0	0	0	0	0	0
At1g02335	1	0	1	1	0	0	1
THF1	1	2	0	0	0	0	1
At4g01130	3	0	0	0	0	0	0
AGO4	0	0	0	0	0	3	0
At4g11175	1	0	0	0	2	0	0
RPL11	0	2	0	0	0	0	0
RPL12A	0	2	0	0	0	0	0
FUC95A	0	0	0	0	0	2	0
CTIMC	0	2	0	0	1	0	0
EXPA15	0	0	0	0	0	2	0
PER16	0	0	0	0	0	1	0
At1g32060	0	2	0	1	0	0	0
TSS	0	2	1	0	0	0	0
RRF	0	0	0	0	0	2	0
EXO	0	0	0	0	0	2	0
RPL6	0	2	0	0	0	0	0
SPATA21	0	2	0	0	0	0	0
NDPK2	2	0	0	0	0	0	0
AT2S2	2	0	0	0	0	0	0
GLL22	0	0	0	1	0	1	0
BGLU3	0	0	0	0	1	0	0

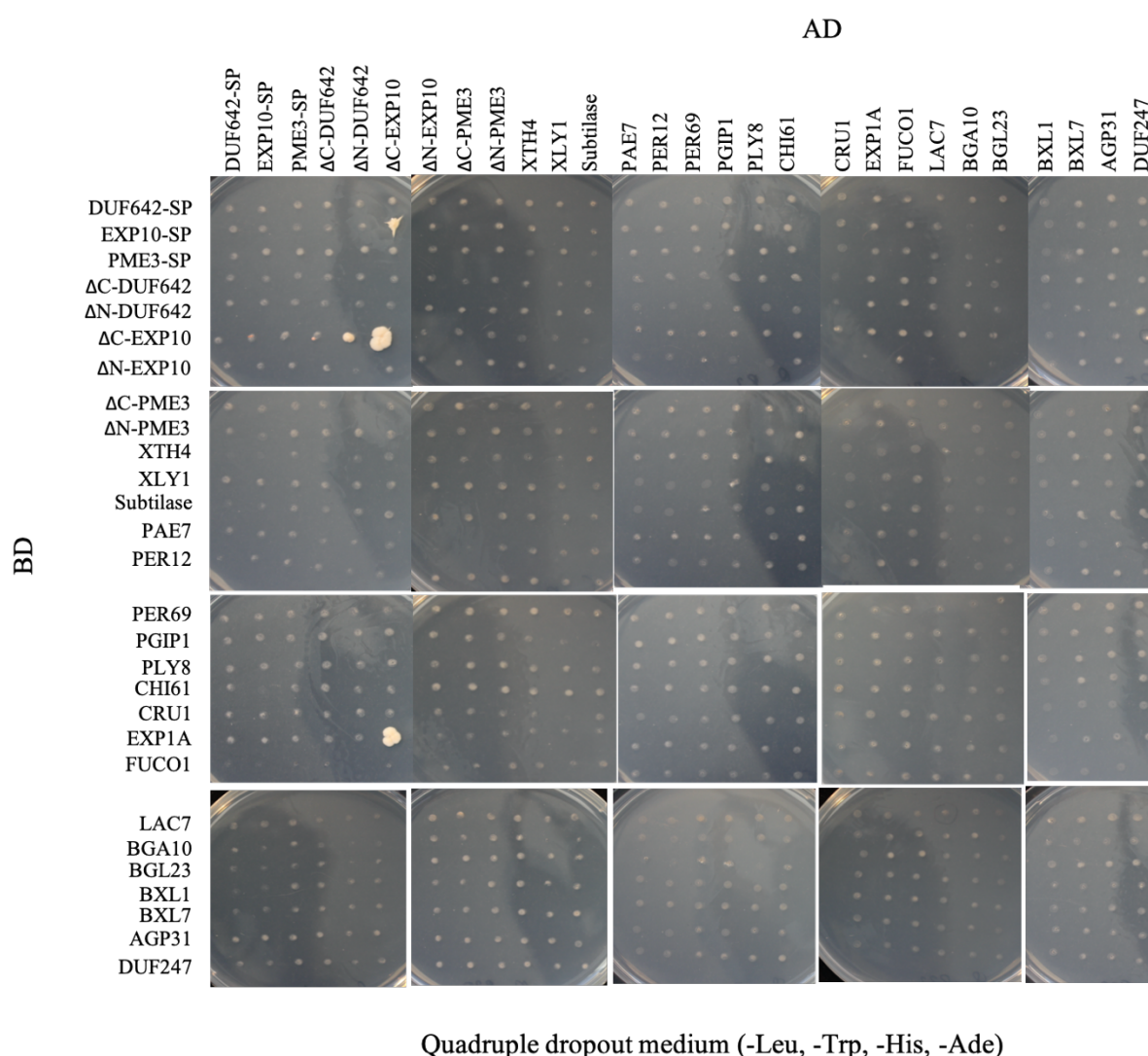
ตารางที่ 5 (ต่อ)

	DUF-3	DUF-8	DUF-9	GST control
GSTU19	17	16	16	13
RBCS-2B	16	12	10	9
XYL1	15	14	15	2
RBCS-1A	15	9	8	10
PAE7	12	9	12	3
PME18	11	7	10	8
FBA2	15	9	10	4
BGLU44	11	10	9	10
MS1	14	12	8	0
CWINV1	11	6	9	2
GER3	5	6	7	4
HEXO3	10	11	6	0
CRC	11	5	7	5
LAC7	8	4	7	0
PER57	7	5	5	0
GAPC1	11	3	2	0
PER72	3	2	5	1
CRA1	5	4	6	0
PME3	6	6	6	0
BCA1	6	4	1	0
PER12	3	3	2	0
GDH2	5	2	3	0
NIT1	4	3	3	0
PER27	5	4	0	0
GLP4	6	2	4	0
TCTP1	4	3	4	0
NADP-ME2	0	0	0	6
PME17	1	5	3	0
RPL6A	5	3	1	0
MDH1	1	0	0	7
At2g43610	4	2	0	3
GLL23	4	0	3	0
NSP1	1	2	0	3
EXL4	0	0	2	4
SBT1	5	1	0	1
LOS1	3	2	2	0

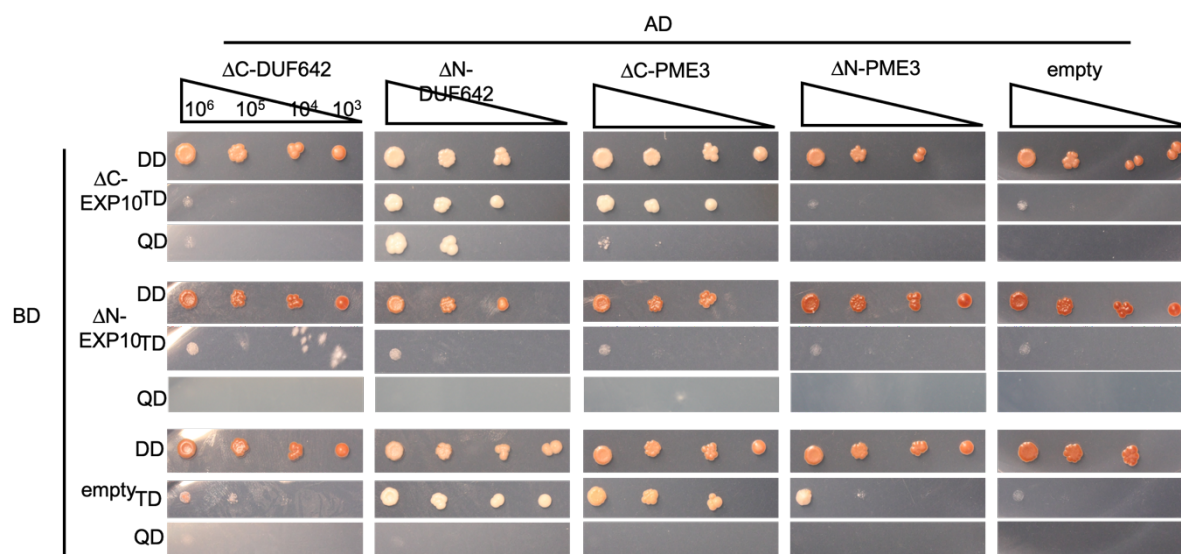
CPN60B1	3	1	3	0
PGIP1	2	2	2	0
TGG4	0	0	3	0
CSP41B	5	2	0	0
SCPL34	4	1	1	0
A1	5	0	1	0
PGIP2	3	2	2	0
EXPA3	1	1	3	1
AGP31	4	1	0	0
LHCB1	3	2	1	0
SBT11	4	0	0	0
SBT4	3	0	1	0
MSD2	1	1	0	3
NDPK3	1	3	0	5
GAPA1	5	0	0	0
BGLU23	2	1	2	1
SCPL24	2	0	2	0
GLU1	3	0	0	1
GLO1	2	1	0	0
BXL7	3	0	0	1
CTIMC	3	0	0	1
BGAL10	1	2	0	0
GLP10	1	2	2	0
MDAR5	4	0	0	0
RCA	3	0	0	0
PRS4	1	0	0	3
CMT3	0	1	2	0
MED37C	2	3	0	0
PME2	4	1	2	0
PGK1	2	1	0	0
AGT1	2	0	0	0
SCPL25	0	1	2	0
RPSaA	2	0	1	0
LTP2	2	0	1	0
PLAT2	2	0	1	0
GSTF5	2	0	0	0
FBA3	2	0	0	0

7.2 การศึกษาการทำงานร่วมกันระหว่างโปรตีนในผนังเซลล์ด้วยวิธี Yeast-two hybrid

เพื่อยืนยันการทำงานร่วมระหว่างโปรตีน DUF642 และโปรตีนอื่น ๆ ที่ตรวจพบจากการทดลอง co-immunoprecipitation จึงโคลนยีนสำหรับโปรตีนที่ตรวจพบและนำมาทดสอบการทำงานร่วมด้วยวิธี yeast two hybrid จากการทดสอบขั้นต้นไม่พบว่าการเข้าไจระหว่างคู่โปรตีนใดเลย อย่างไรก็ตาม คาดการณ์ว่าสภาวะการทำงานของโปรตีนในผนังเซลล์โดยปกติขึ้นอยู่กับสภาวะที่เป็นกรด แตกต่างจากสภาวะในนิวเคลียสที่เป็นกลางถึงด่าง ซึ่งเป็นสภาวะที่ใช้ทดสอบในเซลล์ยีสต์ ดังนั้นจึงได้ทดสอบแยกโดเมนของโปรตีน DUF642, Expansin และ PME3 ออกเป็น 2 ส่วนและทดสอบการทำงานแยกกัน จากการทดสอบเบื้องต้นพบว่าโปรตีน DUF642 ส่วน C-terminus สามารถเข้าจับกับ N-terminus ของ Expansin ได้ (ภาพที่ 30) และเมื่อนำไปทดสอบแบบละเอียดพบว่าโปรตีนทั้งสองสามารถเข้าจับกันได้ในสภาวะภายในนิวเคลียสของยีสต์ได้จริง (ภาพที่ 31) อย่างไรก็ตามการทดลองนี้ตรวจไม่พบการเข้าจับระหว่าง DUF642 กับ PME3 ตามการรายงานก่อนหน้านี้



ภาพที่ 30 การทดสอบ Yeast two hybrid เพื่อบ่งชี้การเข้าจับระหว่างโปรตีน DUF642 และโปรตีนในผนังเซลล์ ภาพแสดงยีสต์ลูกผสมบนอาหาร -Leu, -Trp, -His และ -Ade AD คือเวกเตอร์สำหรับ Activation domain และ BD คือเวกเตอร์สำหรับ binding domain



ภาพที่ 31 การทดสอบ Yeast two hybrid ระหว่างโปรตีน DUF642, Expansin และ PME3 โดยทดสอบแยกโดเมน N-terminus และ C-terminus (DD: double dropout, TD: triple dropout และ QD: quadruple dropout)

วิจารณ์ผลการทดลอง

การจัดกลุ่มยีนโดยใช้ลำดับกรดอะมิโนรอบครีวโปรตีน DUF642 และสร้างต้นไม้วิวัฒนาการสามารถแบ่งออกเป็น 5 กลุ่ม โดยผลจากต้นไม้วิวัฒนาการของงานวิจัยนี้สอดคล้องกับงานวิจัยของ (Vazquez-Lobo et al., 2012) ซึ่งได้รายงานผลการจัดกลุ่มครีวโปรตีน DUF642 ก่อนหน้านี้ ผลต้นไม้วิวัฒนาการทั้งสองมีรูปแบบที่ใกล้เคียงกัน และมีการจัดกลุ่มโปรตีนที่เหมือนกัน แบ่งเป็น 3 กลุ่มใหญ่ ๆ ที่มีความเหมือนกัน ซึ่งสังเกตจากค่า distant มีค่าน้อยโดยแบ่งออกเป็นกลุ่มที่ 1 มีสมาชิกโปรตีน DUF642-1, DUF642-3 และกลุ่มที่ 2 มีสมาชิกโปรตีน DUF642-8, DUF642-10, DUF642-7, DUF642-6, DUF642-4, DUF642-5 นอกจากนี้กลุ่มที่ 3 มีสมาชิกโปรตีน DUF642-9 ไม่สามารถจัดกลุ่มได้ เพราะลำดับมีความต่างจากสมาชิกทั้งหมด โดยการจัดกลุ่มต้นไม้วิวัฒนาการส่งผลต่อการวิเคราะห์หน้าที่การทำงานของโปรตีน โดยสมมุติฐานว่า โปรตีนที่จัดอยู่ในกลุ่มเดียวอาจมีหน้าที่การทำงานที่เหมือนกันหรือใกล้เคียงกัน

จากการผลิตโปรตีนรีคอมบิแนนท์ด้วย *E. coli* โดยใช้เวกเตอร์ pGEX และมีการตัดส่วนลำดับเบส 50 เบสแรกที่แปลรหัสเป็นเปปไทด์ส่งสัญญาณออก พบว่าสามารถสร้างรีคอมบิแนนท์โปรตีนได้ทั้งหมด โดยรีคอมบิแนนท์โปรตีนที่ได้มีโปรตีน GST สำหรับช่วยในการสกัดให้บริสุทธิ์ ทั้งนี้สามารถสกัดโปรตีนบริสุทธิ์ได้ปริมาณ 200 ไมโครกรัมต่ออาหารเลี้ยงเชื้อ 200 มิลลิลิตร โดยงานวิจัยของ (Vazquez-Lobo et al., 2012) การโคลนยีน DUF642 ใช้ยีน DUF642-6 ที่ปราศจากส่วน signal peptide ที่ช่วงต้นสอดคล้องกับงานวิจัยนี้ เพราะว่าครีวโปรตีน DUF642 ทำหน้าที่ภายในผนังเซลล์ โดยถ้าโปรตีน DUF642 ที่มีส่วน signal peptide จะหน้าที่ส่งโปร DUF642 ไปยังส่วนนอกเซลล์ หรือ apoplasmic space และในงานวิจัยของ (Vazquez-Lobo et al., 2012) ใช้เวกเตอร์ pMAL-c2 ในการผลิตรีคอมบิแนนท์โปรตีน MBP-DUF642-6-V5-6xHis โดยมีโดเมนโปรตีน MBP สำหรับช่วยในการสกัดแยกให้บริสุทธิ์

เมื่อเปรียบเทียบผลการทดสอบการจับกับโพลีแซคคาไรด์ก่อนและหลังการตัดด้วยเอนไซม์ทรอมบินพบว่าหลังการตัดสัญญาณที่ได้เพิ่มมากขึ้น บ่งชี้ว่าโปรตีน GST น่าจะส่งผลต่อการทำงานของรีคอมบิแนนต์โปรตีน DUF642 โดยโปรตีน GST ซึ่งมีขนาดโมเลกุลใหญ่ ประมาณ 26 กิโลเดอลตันอาจบดบังหรือรบกวนการจับระหว่างรีคอมบิแนนท์โปรตีน DUF642 กับโพลีแซคคาไรด์ เมื่อเปรียบเทียบกับงานวิจัยของ Vazquez-Lobo et al. (2012) ที่ใช้รีคอมบิแนนท์โปรตีน MBP-DUF642-6-V5-6xHis และทดสอบปฏิกิริยาการจับกับโพลีแซคคาไรด์กลุ่มเซลลูโลส, กลุ่มเฮมิเซลลูโลส และ กลุ่มเพคติน และตรวจสอบด้วย Anti-MBP และปฏิกิริยาของเอนไซม์ alkaline phosphatase พบว่าผลมีสัญญาณจุดบนแผ่นไนโตรเซลลูโลสแมมเบรนเข้มข้นกว่าชุดควบคุมเพียงเล็กน้อย สัญญาณการจับที่ต่ำนี้อาจเพราะโปรตีน MBP มีขนาดโมเลกุล 42.5 กิโลเดอลตัน ส่งผลรบกวนการทำงานของโปรตีน DUF642-6

งานวิจัยของ Vazquez-Lobo et al. (2012) ใช้กระบวนการตรวจสอบ colorimetric โดยทำการทดลองการจับระหว่าง รีคอมบิแนนท์โปรตีน MBP- DUF642-6-V5-6xHis และกลุ่มโพลีแซคคาไรด์ ด้วยวิธี ด็อทบลอต โดยตรวจสอบด้วย แอนติบอดี anti-MBP และปฏิกิริยาของเอนไซม์ alkaline phosphatase โดยใช้โปรตีน MBP- DUF642-6-V5-6xHis ความเข้มข้น 10 ไมโครกรัมต่อมิลลิลิตร และ 20 ไมโครกรัมต่อมิลลิลิตร ซึ่งใช้กับแผ่นไนโตรเซลลูโลส ประกอบด้วยโพลีแซคคาไรด์ 3 กลุ่มหลัก ทำ 2 ซ้ำ ซึ่งผลแสดงสัญญาณจุดบนแผ่นแมมเบรน มีสัญญาณจุดเข้มพอประมาณ แต่สัญญาณยังไม่ค่อยชัดเจน ซึ่งเมื่อเปรียบเทียบกับงานวิจัยนี้ใช้โปรตีนที่ปริมาณ 100 ไมโครกรัม ทดสอบปฏิกิริยาการจับระหว่างรีคอมบิแนนท์โปรตีน DUF642-HA กับโพลีแซคคาไรด์ ด้วยวิธี ELISA พบว่ามีความไวกว่าวิธีด็อทบลอต แสดงผลการจับเป็นค่าดูดกลืนแสง สามารถนำข้อมูลมาทำเป็นกราฟเพื่อเปรียบเทียบระดับการจับระหว่างโปรตีนแต่ละตัวได้

ก่อนหน้านี้งานวิจัยของ Vazquez-Lobo et al. (2012) ทดสอบการจับระหว่างโปรตีน DUF642-6 และโพลีแซคคาไรด์กลุ่ม เซลลูโลส กลุ่มเฮมิเซลลูโลส และ กลุ่มเพคติน ที่สกัดได้จากอะราบิโดบซิส สรุปว่าโปรตีน DUF642-6 มีความสามารถจับกับโพลีแซคคาไรด์กลุ่มเซลลูโลส และ กลุ่มเฮมิเซลลูโลส และไม่จับกับเพคติน เมื่อเปรียบเทียบกับผลทดลองงานโครงการนี้ด้วย ELISA ของโปรตีน DUF642-6-HA กลับไม่พบสัญญาณการจับกับเซลลูโลสชนิดไม่ละลายน้ำ แต่พบสัญญาณการจับโพลีแซคคาไรด์ชนิด Carboxy methyl cellulose ซึ่งเป็นเซลลูโลสที่มีการเติมหมู่เมธิลเพื่อเพิ่มความสามารถในการละลายน้ำ และ พบสัญญาณการจับกับกลุ่มเฮมิเซลลูโลส ได้แก่ tamarind xyloglucan, galacturonan, arabinoxylan และ กลุ่มเพคติน ได้แก่ PGA RG-I from soybean, pectin citrus, pectin apple

ผลการทดลอง DUF642-6-HA น่าจะเป็นไปในแนวทางเดียวกับรายงานก่อนหน้านี้ คือ DUF642-6สามารถเข้าจับเซลลูโลสชนิดที่ละลายน้ำได้ (amorphous) ซึ่งพบได้ในเซลลูโลสที่สกัดจากผนังเซลล์ของอะราบิโดบซิส Vazquez-Lobo et al. (2012) และเป็นคุณสมบัติของ CMC และ DUF642-6ไม่สามารถจับกับเซลลูโลสชนิดไม่ละลายน้ำ (crystalline) เนื่องจากไม่สามารถจับกับ fibrous cellulose

โปรตีน DUF642-6-HA สามารถกับกลุ่มเฮมิเซลลูโลส ได้แก่ tamarind xyloglucan, galactomannan และ arabinoxylan โดยสัญญาณการจับที่ค่อนข้างต่ำ ซึ่งตรงกับงานวิจัยของ Vazquez-Lobo et al. (2012) เพราะโปรตีน MBP-DUF642-6-V5-6xHis มีการจับกับเฮมิเซลลูโลส

โปรตีน DUF642-6-HA สามารถเข้าจับกับกลุ่มเพคติน ได้แก่ PGA RG-I from soybean, pectin citrus, pectin apple ซึ่งไม่ตรงกับรายงานของ Vazquez-Lobo et al. (2012) เพราะโปรตีน MBP-DUF642-6-V5-6xHis เพราะไม่มีสัญญาณการจับกับกลุ่มนี้ แต่ในงานวิจัยของ Sanchez et al. (2014)

พบว่าโปรตีน DUF642-7 มีความเกี่ยวข้องกับเอนไซม์ pectin methyl esterase โดยคาดว่าโปรตีน At4g32460 ทำหน้าที่จับกับเอนไซม์ Pectin methyl esterase และพาไปตัดหมู่ methyl ของเพคตินและควบคุมความแข็งแรงของโครงสร้างเพคติน ดังนั้นผลงานวิจัยดังกล่าวอาจสอดคล้องกับผลทดลองงานวิจัยนี้ โดยครอบครัวโปรตีน DUF642 ทั้ง 10 ตัว มีความสามารถในการจับกับโพลีแซคคาไรด์กลุ่มเพคติน

จากงานวิจัยของ Sanchez et al. (2014) สรุปว่าโปรตีนแฟมิลี DUF642 มีโปรตีนโดเมน carbohydrate binding module (CBM) ซึ่งมีผลต่อการจับโพลีแซคคาไรด์ โดยการทำงานของโมดูลนี้สัมพันธ์กับโพลีเมอร์เป้าหมายอย่างจำเพาะเจาะจงในพืช นอกจากนี้โมดูลนี้ยังปรากฏในเอนไซม์ของแบคทีเรีย ซึ่งเป็นเอนไซม์ ทำหน้าที่ไฮโดรไลซ์ โพลีแซคคาไรด์เฮมิเซลลูโลส และ โพลีแซคคาไรด์เพคตินในผนังเซลล์

จากผลการทดสอบปฏิกิริยาการจับระหว่างรีคอมบิแนนต์โปรตีน DUF642-HA กับโพลีแซคคาไรด์ มีแนวโน้มสอดคล้องกับผลจากการจัดกลุ่มต้นไม้วิวัฒนาการ โดยโปรตีน DUF642 ถ้าอยู่ในกลุ่มวิวัฒนาการเดียวกันนั้นจะมีผลการจับกับโพลีแซคคาไรด์ชนิดเดียวกัน หรือคล้ายกันมาก เช่น กลุ่มที่ 1 จากต้นไม้วิวัฒนาการ พบว่ามีโปรตีน DUF642-8-HA, DUF642-10-HA, DUF642-7-HA ผลการวิเคราะห์จากกลุ่มที่ 1 โดยโปรตีน DUF642-8-HA และโปรตีน DUF642-10-HA มีสายวิวัฒนาการใกล้เคียงกันมากกว่า พบว่าโปรตีน 2 ชนิดนี้มีการจับโพลีแซคคาไรด์ชนิดเดียวกัน คือ กลุ่มเพคติน ได้แก่ PGA โดยที่โปรตีน DUF642-7-HA มีสายวิวัฒนาการห่างมากที่สุดกับโปรตีนในกลุ่มนี้ มีจับกับ กลุ่มเซลลูโลส ได้แก่ carboxy methyl cellulose และกลุ่มเพคติน ได้แก่ PGA, RG-I from soybean, pectin citrus, pectin apple

กลุ่มที่ 2 จากต้นไม้วิวัฒนาการ คือโปรตีน DUF642-6-HA, DUF642-4-HA, DUF642-5-HA วิเคราะห์จากกลุ่มที่ 2 โดยโปรตีน DUF642-4-HA และโปรตีน DUF642-5-HA มีสายวิวัฒนาการใกล้เคียงกันมากกว่า พบว่าโปรตีน 2 ชนิดนี้มีการจับโพลีแซคคาไรด์ชนิดเดียวกัน คือ กลุ่มเซลลูโลส ได้แก่ carboxy methyl cellulose และกลุ่มเฮมิเซลลูโลส ได้แก่ wheat arabinoxylan และกลุ่มเพคติน ได้แก่ PGA, RG-I from soybean, pectin citrus, pectin apple ส่วนโปรตีน DUF642-6-HA มีสายวิวัฒนาการห่างมากที่สุดในกลุ่มนี้ และรีคอมบิแนนต์โปรตีน DUF642-6-HA จับกับเซลลูโลส ได้แก่ carboxy methyl cellulose และกลุ่มเพคติน ได้แก่ PGA, RG-I from soybean, pectin citrus, pectin apple

กลุ่มที่ 3 คือโปรตีน DUF642-2-HA มีโปรตีนเพียงชนิดเดียวในกลุ่ม มีผลการจับโพลีแซคคาไรด์ใกล้เคียงกับ กลุ่มที่ 1 และ กลุ่มที่ 2 เพราะจากต้นไม้วิวัฒนาการมีความใกล้ชิดกับสองกลุ่มแรก คือจับกับกลุ่มเซลลูโลส ได้แก่ carboxy methyl cellulose และกลุ่มเพคติน ได้แก่ PGA, RG-I from soybean, pectin citrus, pectin apple

กลุ่มที่ 4 คือโปรตีน DUF642-1-HA และ DUF642-3-HA พบว่าโปรตีน 2 ชนิดนี้มีการจับโพลีแซคคาไรด์ชนิดเดียวกัน คือ กลุ่มเซลลูโลส ได้แก่ carboxy methyl cellulose และกลุ่มเพคติน ได้แก่ PGA, RG-I from soybean, pectin citrus, pectin apple และสุดท้ายอย่างกลุ่มที่ 5 คือโปรตีน DUF642-9-HA ซึ่งมีลำดับกรดอะมิโนแตกต่างกับโปรตีน 9 ชนิดมากที่สุด และมีการจับโพลีแซคคาไรด์แตกต่างจากโปรตีนอื่นๆอีก 9 ตัว โดยจับกับโพลีแซคคาไรด์ทุกชนิด ทั้งในกลุ่มเซลลูโลส, เฮมิเซลลูโลส และ เพคติน

จากรายงานของ Sanchez et al., (2012) อธิบายว่า ครอบครัวโปรตีน DUF642 มีโปรตีนส่งสัญญาณ (signal peptide) และนอกจากนี้ มี 2 โดเมนหลัก ที่ปลาย N-terminus มี signal peptide เชื่อมต่อกับโดเมนที่ 1 คือ Unknown function domain และเชื่อมต่อกับโดเมนที่ 2 คือ carbohydrate binding domain ตามลำดับ เมื่อวิเคราะห์ผลต้นไม้วิวัฒนาการ โดยหลักการถาลำดับกรดอะมิโนมีความเหมือนกันจะจัดอยู่ในกลุ่มเดียวกัน ดังนั้นถ้าโปรตีนถูกจัดในกลุ่มเดียวกัน เพราะโปรตีนกลุ่มนี้มีลำดับกรดอะมิโนเหมือนกันในส่วน Unknown function domain เท่านั้น จะส่งผลให้โปรตีนกลุ่มนี้ไม่จับการจับโพลีแซคคาไรด์ชนิดเดียวกัน แต่ถ้าโปรตีนถูกจัดในกลุ่มเดียวกัน และโปรตีนกลุ่มนี้มีลำดับกรดอะมิโนเหมือนกันในส่วน carbohydrate binding domain จะส่งผลให้โปรตีนกลุ่มนี้มีความเหมือนในการจับการจับโพลีแซคคาไรด์ชนิดเดียวกันหรือใกล้เคียงกัน (ข้อมูลในการวิเคราะห์จากปฏิกิริยาการจับระหว่างรีคอมบิแนนต์โปรตีน DUF642-HA กับโพลีแซคคาไรด์ ด้วยกระบวนการ ELISA assays) เมื่อเปรียบเทียบพบว่า ผลจากการทดลองปฏิกิริยาการจับระหว่างรีคอมบิแนนต์โปรตีน DUF642-HA กับโพลีแซคคาไรด์ สอดคล้องกับ ผลจากต้นไม้วิวัฒนาการ

จากผลการตรวจสอบการสูญเสียหน้าที่ของยีน *DUF642* จำนวน 1 ตำแหน่ง จำนวน 9 ยีน ในต้นอะราบิโดพซิส ไม่พบความเปลี่ยนแปลงทางฟีโนไทป์อย่างชัดเจน แสดงให้เห็นถึงการทำงานแทนที่กันของยีนในแฟมิลี นอกจากนี้เมื่อวิเคราะห์ผลของการแสดงออกของยีนด้วย Promoter-GUS fusion ก็พบว่ามากกว่า 5 ยีนที่มีการแสดงออกสูงตลอดตั้งแต่ระยะต้นอ่อนจนถึงระยะการพัฒนา ทำให้เมื่อศึกษาต้นพันธุ์กลายที่มีการสูญเสียหน้าที่ยีนเพียง 1 ตำแหน่ง จึงไม่มีผลต่อการเจริญและพัฒนาของอะราบิโดพซิส

จากการศึกษาผลของการสูญเสียหน้าที่ของยีนพบว่า ต้นพันธุ์กลายที่มีการสูญเสียหน้าที่ยีนตั้งแต่ 1 ตำแหน่ง จนถึง 7 ตำแหน่ง รวมจำนวนพันธุ์กลายทั้งหมด 70 สายพันธุ์ เมื่อตรวจสอบลักษณะทางฟีโนไทป์ด้วยเทคนิคต่าง ๆ ได้แก่ การตรวจสอบการเจริญเบื้องต้นโดยการเพาะเลี้ยงเมล็ดในระยะต้นอ่อน และการตรวจสอบไฮโปคอทิล การทดสอบสออาหารที่เติม NaCl และ sorbitol อาหารที่เติมสารยับยั้งการทำงานของผนังเซลล์ ได้แก่ isoxaben, oryzalin และ $AlCl_3$ การตรวจสอบองค์ประกอบน้ำตาลโมเลกุลเดียวในผนังเซลล์ และการตรวจสอบการย่อยสลายของผนังเซลล์ ทั้งหมด ไม่สามารถบ่งชี้ความเปลี่ยนแปลงของต้นพันธุ์กลายได้ ยกเว้นเพียงการเจริญเติบโตที่ลดลงเมื่อทดสอบกับอาหารที่เติม $AlCl_3$ ซึ่งมีคุณสมบัติยับยั้งการทำงานของเพกติน อย่างไรก็ตามข้อมูลที่ได้นี้ยังไม่เพียงพอต่อการสรุปผลการทำงานของ DUF642 ต่อเพกติน เนื่องจากสายพันธุ์กลาย 7 ตำแหน่ง ยังคงมียีน *DUF642* ที่ยังทำหน้าที่ได้อยู่อีก 3 ยีน ดังนั้นหากต้องการทดสอบการทำงานของ DUF642 ในพืชทั้งหมด อาจต้องสร้างสายพันธุ์ที่มีการสูญเสียหน้าที่ของยีน *DUF642* ทั้ง 10 ยีนต่อไป

จาก 70 สายพันธุ์ มีเพียงสายพันธุ์กลาย 2 ตำแหน่งพร้อมกัน *duf642-1 duf642-4* เท่านั้นที่แสดงลักษณะการเจริญที่ลดลง มีขนาดเล็ก และส่งผลกับการพัฒนาส่วนอื่น ๆ ของพืชด้วย เช่น การเจริญของไฮโปคอทิล อย่างไรก็ตามการทดสอบ complementation แสดงให้เห็นว่าลักษณะที่เกิดขึ้นน่าจะเกิดจากยีนกลายในตำแหน่งอื่น ๆ ที่มีอยู่ในต้นพันธุ์กลาย พันธุ์กลายดังกล่าวขณะนี้กำลังตรวจหายีนกลายที่ส่งให้เกิดฟีโนไทป์ดังกล่าว

จากผลการเข้าจับของโปรตีน DUF642 กับโปรตีนและเอนไซม์อื่น ๆ ในผนังเซลล์ของอะราบิโดพซิสด้วยวิธี immunoprecipitation และ LC/MS/MS แสดงให้เห็นว่าโปรตีน DUF642 นั้นสามารถเข้าจับได้ทั้ง โพลีแซคคาไรด์ และ เอนไซม์ cell wall modifying enzymes จากผลดังกล่าวทำให้สามารถ

ตั้งสมมติฐานได้ว่าโปรตีน DUF642 น่าจะเป็นตัวกลางในการนำพาเอนไซม์ต่าง ๆ ในผนังเซลล์ไปทำงานที่ตำแหน่งโพลีแซคคาไรด์เป้าหมายได้ อย่างไรก็ตามเนื่องจากผลการทดลองได้รายชื่อของโปรตีนและเอนไซม์จำนวนมาก ทำให้ไม่สามารถบอกได้แน่ชัดว่าโปรตีนชนิดใดที่ทำงานได้โดยตรงกับโปรตีน DUF642 ถึงแม้ว่ามีตัวควบคุมที่มีโปรตีนอื่น ๆ ปะปนมาจำนวนน้อย ดังนั้นจึงต้องมีการทดสอบยืนยันการเข้าจับระหว่าง DUF642 กับโปรตีนที่ตรวจพบต่อไปด้วยวิธีที่จำเพาะมากกว่านี้

จากการทดสอบด้วยวิธี Yeast two hybrid พบว่าโปรตีนที่ตรวจพบทั้งหมดจากการทดลอง immunoprecipitation พบว่ามีเพียงโปรตีน Expansin เท่านั้นที่สามารถเข้าจับได้ในนิวเคลียสของยีสต์เมื่อทดสอบยืนยันอย่างละเอียดโดยการทดสอบกับอาหารคัดเลือก 3 ชนิด พร้อมตัวควบคุมและการเจือจางเซลล์ พบว่า โปรตีน DUF642 ส่วน C-terminus สามารถเข้าจับกับ N-terminus ของ Expansin ได้อย่างจำเพาะ และไม่พบว่าโปรตีน DUF642 เข้าจับกับ เอนไซม์ PME3 ตามที่มีรายงานไว้ใน Sanchez et al. (2014) สาเหตุหนึ่งที่น่าคิดว่าทำให้มีการตรวจพบโปรตีนหลายชนิดจากการทดลอง co-immunoprecipitation ด้วยโปรตีน DUF642 คือ การที่โปรตีนหลายชนิดในผนังเซลล์นั้นสามารถจับกับโพลีแซคคาไรด์ได้ เช่นเดียวกับ โปรตีน DUF642 ดังนั้นเมื่อใช้โปรตีน DUF642 ในการคัดแยกโปรตีนจากผนังเซลล์ จึงอาจเป็นการดึงโพลีแซคคาไรด์ขนาดเล็กที่มีโปรตีนอื่น ๆ จับเกาะอยู่ตามออกมาด้วย

สรุปผลการดำเนินงาน

งานที่เกี่ยวข้องโดยตรงกับทุนนี้ที่มีการตีพิมพ์แล้วคือ เรื่อง การวิเคราะห์การย่อยของผนังเซลล์ของข้าวพันธุ์ไทย ซึ่งได้รับการตอบรับให้ลงตีพิมพ์แล้ว อยู่ในระหว่างการออกเลขหน้า และ เรื่อง การพัฒนาเทคนิคการวิเคราะห์ขนาดของโพลีแซคคาไรด์ของผนังเซลล์ด้วยวิธี GPC-ELISA ซึ่งขณะนี้อยู่ในระหว่างการตรวจสอบต้นฉบับ (under review ในวารสาร The Plant Journal)

อย่างไรก็ตามเนื้อหาเกี่ยวกับการวิเคราะห์การทำงานของโปรตีน DUF642 กำลังอยู่ในระหว่างการทำข้อมูลเพิ่มเติม ทั้งนี้เนื่องจากงานดังกล่าวมีการดำเนินงานร่วมกับ Prof. Simon McQueen-Mason, University of York แห่ง สหราชอาณาจักร และ Prof. George Bassel, University of Warwick แห่ง สหราชอาณาจักร ซึ่งให้ได้คำแนะนำให้ทำการทดลองเพิ่มเติม เพื่อให้ได้ผลที่น่าเชื่อถือ และมีข้อมูลที่น่าสนใจสำหรับการตีพิมพ์ที่มี impact ที่สูงขึ้น เช่น การยืนยันผลการเข้าจับระหว่าง DUF642 และ Expansin ด้วยเทคนิคอื่น ๆ ในระดับ *in vivo* คือ BiFC และ Co-immunoprecipitation แบบเฉพาะ 2 โปรตีน และการสร้างสายพันธุ์กลายให้ครบ 10 ยีน เพื่อให้ได้ข้อมูลหน้าที่ของโปรตีนที่ชัดเจนก่อนการลงตีพิมพ์

นอกจากนี้ทุนวิจัยนักวิจัยรุ่นกลางได้นำไปใช้สนับสนุนดำเนินการวิจัยจนได้ผลงานตีพิมพ์อื่น ๆ ตามผลงานตีพิมพ์ที่แนบมาด้วยแล้ว

เอกสารอ้างอิงของโครงการวิจัย

1. Baker, B., Zambryski, P., Staskawicz, B., & Dinesh-Kumar, S. P. (1997). Signaling in plant-microbe interactions. *Science*, 276(5313), 726-733.
2. Bayer, E. M., Bottrill, A. R., Walshaw, J., Vigouroux, M., Naldrett, M. J., Thomas, C. L., & Maule, A. J. (2006). Arabidopsis cell wall proteome defined using multidimensional protein identification technology. *Proteomics*, 6(1), 301-311.
3. Bischoff, V., Nita, S., Neumetzler, L., Schindelasch, D., Urbain, A., Eshed, R., ... & Scheible, W. R. (2010). TRICHOME BIREFRINGENCE and its homolog AT5G01360 encode plant-specific DUF231 proteins required for cellulose biosynthesis in Arabidopsis. *Plant physiology*, 153(2), 590-602.
4. Carpita, N. C., & Gibeaut, D. M. (1993). Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. *The Plant Journal*, 3(1), 1-30.
5. Carpita, N. and McCann, M., 2000. Biochemistry and molecular biology of plants. *Buchanan, BB*, pp.52-108.
6. Chen, S., Songkumarn, P., Liu, J. and Wang, G.L., 2009. A versatile zero background T-vector system for gene cloning and functional genomics. *Plant physiology*, 150(3), pp.1111-1121.
7. Cosgrove, D. J. (1999). Enzymes and other agents that enhance cell wall extensibility. *Annual review of plant biology*, 50(1), 391-417.
8. Cosgrove, D. J. (2005). Growth of the plant cell wall. *Nature reviews molecular cell biology*, 6(11), 850-861.
9. Delmer, D. P. (1999). Cellulose biosynthesis: exciting times for a difficult field of study. *Annual review of plant biology*, 50(1), 245-276.
10. Epstein, L., & Lampion, D. T. (1984). An intramolecular linkage involving isodityrosine in extensin. *Phytochemistry*, 23(6), 1241-1246.
11. Fincher, G.B., Stone, B.A. and Clarke, A.E., 1983. Arabinogalactan-proteins: structure, biosynthesis, and function. *Annual Review of Plant Physiology*, 34(1), pp.47-70.
12. Fry, S. C. (1995). Polysaccharide-modifying enzymes in the plant cell wall. *Annual review of plant biology*, 46(1), 497-520.
13. Fry, S. C., Smith, R. C., Renwick, K. F., Martin, D. J., Hodge, S. K., & Matthews, K. J. (1992). Xyloglucan endotransglycosylase, a new wall-loosening enzyme activity from plants. *Biochem. j*, 282, 821-828.
14. Galperin, M. Y., & Koonin, E. V. (2010). From complete genome sequence to 'complete' understanding?. *Trends in biotechnology*, 28(8), 398-406.
15. Gao, Y., Badejo, A. A., Sawa, Y., & Ishikawa, T. (2012). Analysis of two l-galactono-1, 4-lactone-responsive genes with complementary expression during the development of Arabidopsis thaliana. *Plant and Cell Physiology*, 53(3), 592-601.
16. Gao, Y., Nishikawa, H., Badejo, A. A., Shibata, H., Sawa, Y., Nakagawa, T., ... & Ishikawa, T. (2011). Expression of aspartyl protease and C3HC4-type RING zinc finger genes are

- responsive to ascorbic acid in *Arabidopsis thaliana*. *Journal of experimental botany*, 62(10), 3647-3657.
17. Gomez, L. D., Steele-King, C. G., & McQueen-Mason, S. J. (2008). Sustainable liquid biofuels from biomass: the writing's on the walls. *New Phytologist*, 178(3), 473-485.
 18. Hansen, S. F., Harholt, J., Oikawa, A., & Scheller, H. V. (2012). Plant glycosyltransferases beyond CAZy: a perspective on DUF families. *Frontiers in plant science*, 3.
 19. Irshad, M., Canut, H., Borderies, G., Pont-Lezica, R., & Jamet, E. (2008). A new picture of cell wall protein dynamics in elongating cells of *Arabidopsis thaliana*: confirmed actors and newcomers. *BMC Plant Biology*, 8(1), 94.
 20. Keegstra, K., Talmadge, K. W., Bauer, W. D., & Albersheim, P. (1973). The structure of plant cell walls III. A model of the walls of suspension-cultured sycamore cells based on the interconnections of the macromolecular components. *Plant Physiology*, 51(1), 188-197.
 21. Kieliszewski, M. J., & Lampley, D. T. (1994). Extensin: repetitive motifs, functional sites, post-translational codes, and phylogeny. *The Plant Journal*, 5(2), 157-172.
 22. Lee, C., Teng, Q., Zhong, R., Yuan, Y., Haghighat, M., & Ye, Z. H. (2012). Three *Arabidopsis* DUF579 domain-containing GXM proteins are methyltransferases catalyzing 4-O-methylation of glucuronic acid on xylan. *Plant and Cell Physiology*, 53(11), 1934-1949.
 23. McCann, M. C., Wells, B., & Roberts, K. (1990). Direct visualization of cross-links in the primary plant cell wall.
 24. McCann, M. C., & Carpita, N. C. (2008). Designing the deconstruction of plant cell walls. *Current opinion in plant biology*, 11(3), 314-320.
 25. McQueen-Mason, S., Durachko, D. M., & Cosgrove, D. J. (1992). Two endogenous proteins that induce cell wall extension in plants. *The Plant Cell*, 4(11), 1425-1433.
 26. Mewalal, R., Mizrahi, E., Mansfield, S. D., & Myburg, A. A. (2014). Cell wall-related proteins of unknown function: Missing links in plant cell wall development. *Plant and Cell Physiology*, pcu050.
 27. Popper, Z. A., & Fry, S. C. (2005). Widespread occurrence of a covalent linkage between xyloglucan and acidic polysaccharides in suspension-cultured angiosperm cells. *Annals of Botany*, 96(1), 91-99.
 28. Punta, M., Coghill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., ... & Finn, R. D. (2011). The Pfam protein families database. *Nucleic acids research*, gkr1065.
 29. Ridley, B. L., O'Neill, M. A., & Mohnen, D. (2001). Pectins: structure, biosynthesis, and oligogalacturonide-related signaling. *Phytochemistry*, 57(6), 929-967.
 30. Zúñiga-Sánchez, E., Soriano, D., Martínez-Barajas, E., Orozco-Segovia, A. and Gamboa-deBuen, A., 2014. BILD1, the At4g32460 DUF642 gene, is involved in pectin methyl esterase regulation during *Arabidopsis thaliana* seed germination and plant development. *BMC plant biology*, 14(1), p.338.
 31. Showalter, A. M. (1993). Structure and function of plant cell wall proteins. *The Plant Cell*, 5(1), 9.

32. Showalter, A. M., Keppler, B. D., Lichtenberg, J., Gu, D., & Welch, L. R. (2010). A bioinformatics approach to the identification, classification, and analysis of hydroxyproline-rich glycoproteins. *Plant Physiology*, pp-110.
33. Spalding, E. P., & Cosgrove, D. J. (1992). Mechanism of blue-light-induced plasma-membrane depolarization in etiolated cucumber hypocotyls. *Planta*, 188(2), 199-205.
34. Urbanowicz, B. R., Peña, M. J., Ratnaparkhe, S., Avci, U., Backe, J., Steet, H. F., ... & York, W. S. (2012). 4-O-methylation of glucuronic acid in Arabidopsis glucuronoxylan is catalyzed by a domain of unknown function family 579 protein. *Proceedings of the National Academy of Sciences*, 109(35), 14253-14258.
35. Van Sandt, V. S., Suslov, D., Verbelen, J. P., & Vissenberg, K. (2007). Xyloglucan endotransglucosylase activity loosens a plant cell wall. *Annals of Botany*, 100(7), 1467-1473.
36. Vázquez-Lobo, A., Roujol, D., Zúñiga-Sánchez, E., Albenne, C., Piñero, D., de Buen, A. G., & Jamet, E. (2012). The highly conserved spermatophyte cell wall DUF642 protein family: phylogeny and first evidence of interaction with cell wall polysaccharides in vitro. *Molecular phylogenetics and evolution*, 63(2), 510-520.
37. Willats, W. G., Steele-King, C. G., Marcus, S. E., & Knox, J. P. (1999). Side chains of pectic polysaccharides are regulated in relation to cell proliferation and cell differentiation. *The Plant Journal*, 20(6), 619-628.
38. Wolf, S., Hématy, K., & Höfte, H. (2012). Growth control and cell wall signaling in plants. *Annual review of plant biology*, 63, 381-407.
39. Wu, H.Y., Liu, K.H., Wang, Y.C., Wu, J.F., Chiu, W.L., Chen, C.Y., Wu, S.H., Sheen, J. and Lai, E.M., 2014. AGROBEST: an efficient Agrobacterium-mediated transient expression method for versatile gene function analyses in Arabidopsis seedlings. *Plant methods*, 10(1), p.19.
40. Xu, J., Tan, L., Lamport, D. T., Showalter, A. M., & Kieliszewski, M. J. (2008). The O-Hyp glycosylation code in tobacco and Arabidopsis and a proposed role of Hyp-glycans in secretion. *Phytochemistry*, 69(8), 1631-1640.
41. Zhou, Y., Li, S., Qian, Q., Zeng, D., Zhang, M., Guo, L., ... & Li, J. (2009). BC10, a DUF266-containing and Golgi-located type II membrane protein, is required for cell-wall biosynthesis in rice (*Oryza sativa* L.). *The Plant Journal*, 57(3), 446-462.
42. Zúñiga-Sánchez, E., & Gamboa-de Buen, A. (2012). *The two DUF642 At5g11420 and At4g32460-encoded proteins interact in vitro with the AtPME3 catalytic domain*. INTECH Open Access Publisher.

Output จากโครงการวิจัยที่ได้รับทุนจาก สกว.

1. ผลงานตีพิมพ์ในวารสารวิชาการนานาชาติ (ระบุชื่อผู้แต่ง ชื่อเรื่อง ชื่อวารสาร ปี เล่มที่ เลขที่ และหน้า) พร้อมแจ้งสถานะของการตีพิมพ์ เช่น submitted, accepted, in press, published

1. Sakulsathaporn, A., Wonnapijit, P., **Vuttipongchaikij, S.** and Apisitwanich, S., 2017. The complete chloroplast genome sequence of Asian Palmyra palm (*Borassus flabellifer*). *BMC research notes*, 10(1), p.740. (Published)
2. Pipatchartlearnwong, K., Swatdipong, A., **Vuttipongchaikij, S.** and Apisitwanich, S., 2017. Cross-genera transferability of microsatellite loci for Asian Palmyra Palm (*Borassus flabellifer* L.). *HortScience*, 52(9), pp.1164-1167. (Published)
3. Suttangkakul, A., Sirikhachornkit, A., Juntawong, P., Puangtame, W., Chomtong, T., Srifa, S., Sathitnaitham, S., Dumrongthawatchai, W., Jariyachawalid, K. and **Vuttipongchaikij, S.**, 2019. Evaluation of strategies for improving the transgene expression in an oleaginous microalga *Scenedesmus acutus*. *BMC biotechnology*, 19(1), p.4. (Published)
4. Pipatchartlearnwong, K., Juntawong, P., Wonnapijit, P., Apisitwanich, S. and **Vuttipongchaikij, S.**, 2019. Towards sex identification of Asian Palmyra palm (*Borassus flabellifer* L.) by DNA fingerprinting, suppression subtractive hybridization and de novo transcriptome sequencing. *PeerJ*, 7, p.e7268. (Published)
5. Sakulsathaporn, A., Wonnapijit, P., Suttangkakul, A., Apisitwanich, S. and **Vuttipongchaikij, S.**, 2019. RNA editing in the chloroplast of Asian Palmyra palm (*Borassus flabellifer*). *Genetics and Molecular Biology*, 42(4). (Published)
6. Kophimai, Y., Cheenacharoen, S., Simister, R., Gomez, L., McQueen-Mason, S., **Vuttipomngchaikij, S.**, 2020. Straw digestibility of Thai rice accessions. *Agriculture and Natural Resources*. (Accepted, in press)
7. Sathitnaitham, S., Suttangkakul, A., Passorn Wonnapijit, P., McQueen-Mason, S., **Vuttipongchaikij, S.** GPC-ELISA method for systematic mass distribution profiling of plant cell wall matrix polysaccharides. (Submitted, Under review)

2. การนำผลงานวิจัยไปใช้ประโยชน์ เชิงวิชาการ สร้างนักวิจัยใหม่ โดยกำลังจะจบการศึกษาในปีการศึกษา 2563

1. นางสาวสุจิตา สติตโนธรรม ระดับปริญญาโท
2. นางสาวพิชชาพร วรรณนิธิกุล ระดับปริญญาเอก

3. การเสนอผลงานในที่ประชุมวิชาการ

1. Sakulkoo, J., Chomtong, T., Suttangkakul, A., **Vuttipongchaikij, S.** (2017) Functional characterization of DUF642 cell wall proteins in *Arabidopsis thaliana*. Cell wall Biology Meeting, Dalian, China (Poster presentation).
2. Sakulkoo, J., Suttangkakul, A., **Vuttipongchaikij, S.** (2019) Functional characterisation of DUF642 cell wall protein family in Arabidopsis. XV Cell wall meeting, Cambridge, UK. (Poster presentation).
3. Sathitnaitham, S., Suttangkakul, A., **Vuttipongchaikij, S.** (2019) GPC-ELISA based method for profiling the molecular mass distribution of plant cell wall polysaccharides. XV Cell wall meeting, Cambridge, UK. (Oral presentation).

ภาคผนวก

RESEARCH NOTE

Open Access



The complete chloroplast genome sequence of Asian Palmyra palm (*Borassus flabellifer*)

Arpakorn Sakulsathaporn^{1,2}, Passorn Wonnapijit^{3,4}, Supachai Vuttipongchaikij^{3,4,5*} and Somsak Apisitwanich^{1,2,3,4,5,6*}

Abstract

Objective: *Borassus flabellifer* or Asian Palmyra palm is widely distributed in South and Southeast Asia and is horticultural and economic importance for its fruit and palm sugar production. However, its population is in rapid decline, and only a few genetic data are available. We sequenced the complete chloroplast (cp) genome of *B. flabellifer* to provide its genetic data for further utilization.

Results: The cp genome was obtained by Illumina sequencing and manual gap fillings providing 160,021 bp in length containing a pair of inverted repeats (IRs) with 27,256 bp. These IRs divide the genome into a large single copy region 87,444 bp and a small single copy region 18,065 bp. In total, 113 unique genes, 134 SSRs and 47 large repeats were identified. This is the first complete cp genome reported in the genus *Borassus*. A comparative analysis among members of the Borasseae tribe revealed that the *B. flabellifer* cp genome is, so far, the largest and the cp genomes of this tribe have a similar structure, gene number and gene arrangement. A phylogenetic tree reconstructed based on 74 protein-coding genes from 70 monocots demonstrates short branch lengths indicating slow evolutionary rates of cp genomes in family Arecaceae.

Keywords: Arecaceae, Borasseae, Commelinids, Phylogeny, Plastid, Tandem repeats

Introduction

Borassus flabellifer or Asian palmyra palm (family Arecaceae, subfamily Coryphoideae, Borasseae tribe) is a massive dioecious monocot plant with its single stem reaching 30 m in height and large fan-shaped leaves spanning 1–3 m in diameter [1]. Six species are present in the Borasseae tribe including *B. aethiopum* [2], *B. akeassii* [3], *B. sambiranensis* [4] and *B. madagascariensis* [5], which are distributed in Africa, *B. heineanus* [6] found in New Guinea and *B. flabellifer*, which is solely found in Asia [1]. *B. flabellifer* is widespread in the South and Southeast Asia and is of horticultural and economic importance. The fruit is widely consumed, and the

flower sap has been used for palm sugar production for hundreds of years [7]. *B. flabellifer* is currently in rapid decline due to following reasons. First, it grows extremely slow requiring 12–20 years to reach maturity and produce its first inflorescence [8]. Second, urbanization and agricultural development has eliminated a large number of the wild population [9]. Third, it reproduces via cross pollination, but there is currently no reliable mean for sex determination prior its first flowers [10]. Fourth, a clonal propagation method for this species is not well established. With these reasons, conservation and breeding programs of *B. flabellifer* is urgently needed, and genetic data are required for supporting the programs.

To date, genetic data of *B. flabellifer* are limited. A number of DNA markers including RAPD [11], ISSR [12], EST-SR and gSSR [13, 14] have been developed for studying the population in south and southeast Asia and demonstrated its low genetic diversity. However, more

*Correspondence: fsciscv@ku.ac.th; fscissa@ku.ac.th

³ Department of Genetics, Faculty of Science, Kasetsart University, 50 Ngarm Wong Wan Road, Chattuchak, Bangkok 10900, Thailand
Full list of author information is available at the end of the article

sequence data are still needed for detailed studies on genetic diversity and evolution. In particular, the chloroplast (cp) genome sequence would provide both species specific and population specific makers for studying *B. flabellifer*. Here, we report the complete cp genome sequence of *B. flabellifer* obtained by using both next-generation sequencing and manual gap fillings. The cp genome structure, characteristic and gene organization are described. Repetitive sequences were identified. Comparative genome analysis was performed to understand the evolutionary relationship among the Borasseae tribe.

Main text

The complete cp genome sequence of *B. flabellifer*

Because *B. flabellifer* leaf materials are very hard and a direct isolation of cpDNA with high purity is often difficult to obtain, chloroplast was firstly isolated using a modified protocol from Triboush et al. [15] and purified using a modified sucrose gradient method from Sandbrink et al. [16] (Additional file 1: Figure S1). The third leaf from the top (a fully expanded leaf with dark green and no more than 6-month-old) was collected and stored at 4 °C for 7 days to reduce accumulated starch before use. CpDNA was then isolated from the purified chloroplast using DNeasy Plant Mini Kit (Qiagen), and *EcoRI* restriction digests were used for verifying the purity of the cpDNA. Illumina Hiseq 2000 system generated 7,695,267 pair-end reads with an approximately 100 bp average read length. After filtering and eliminating low quality reads and contaminants using FastQC [17] and Trimmomatic [18], a total of 1,539,053,400 bp was obtained. A sliding window size of 4 with an average of Phred score ≥ 20 and removal of 5' and 3' ends with Phred score ≤ 3 were used as the trimming criteria. By mapping to the cp genome of *C. nucifera* (NC_022417) [19] using SOAPEC v2.03, the reads provided an average of 100 \times sequencing depth coverage, and eight contigs covering 92% of the entire cp genome was obtained. Specific PCR amplification and sequencing were performed to fill the missing gaps. The genome map was then drawn by GenomeVx [20].

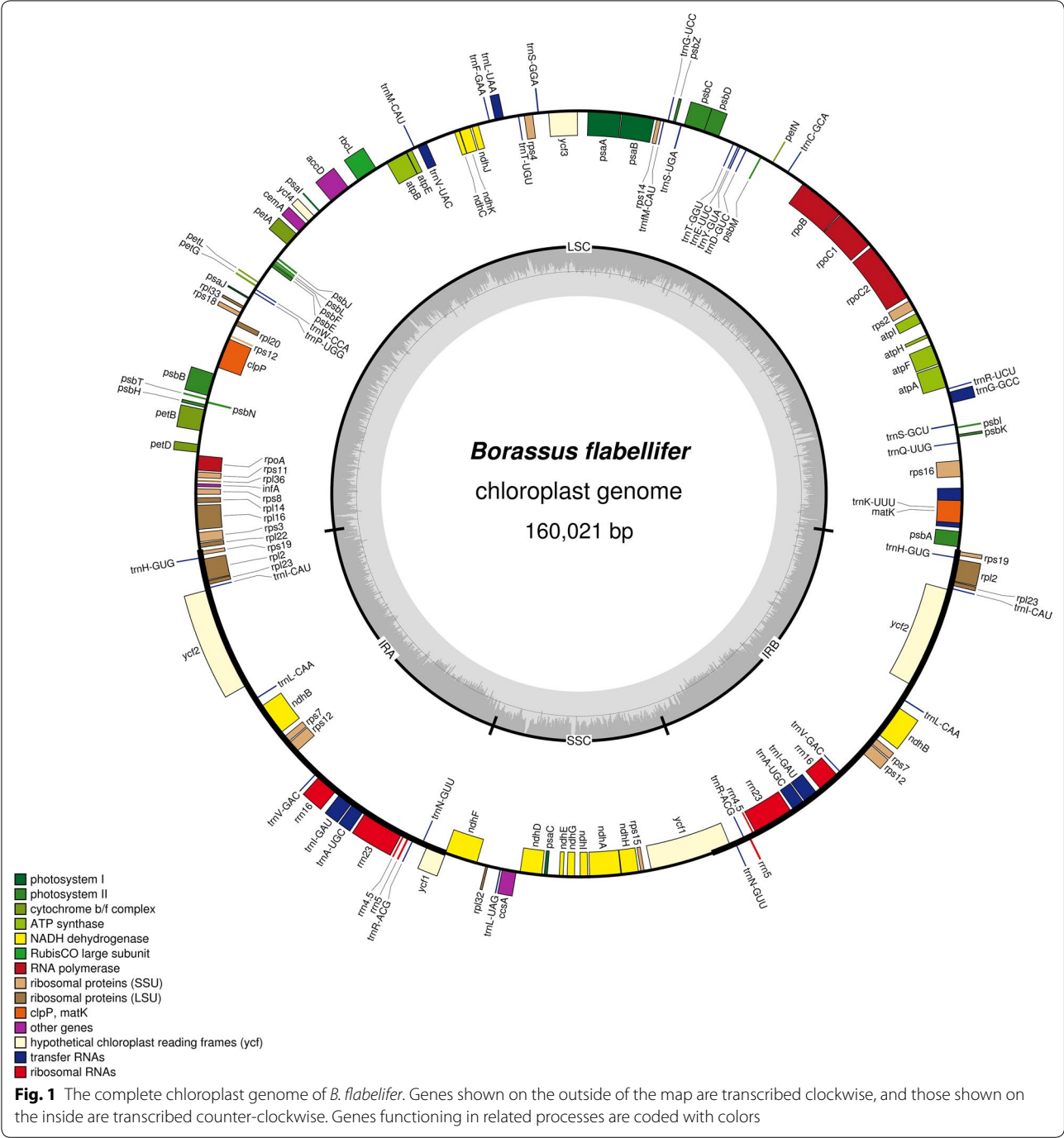
The circular double-stranded DNA of the complete *B. flabellifer* cp genome is 160,021 bp in length (Fig. 1, GenBank Accession Number: KP_901247). It has a typical quadripartite structure composing of a pair of inverted repeat (IR) regions (27,256 bp each), a large single-copy (LCS) region (87,444 bp) and a small single-copy (SSC) region (18,065 bp). The overall GC content is 37.23%. Genome annotation using DOGMA [21] and CpGAVAS [22] with *Phoenix dactylifera* [23] as a reference and tRNAs prediction using tRNA-ScanSE [24] provided that the cp genome contains 113 unique genes: 79

protein-coding genes, 30 tRNA genes, and four rRNA genes (Additional file 2: Table S1). All of the four rRNA genes (*rrn4.5*, *rrn5*, *rrn16* and *rrn23*), seven protein-coding genes (*rps19*, *rpl2*, *rpl23*, *ndhB*, *rps7*, *ycf1* and *ycf2*), two pseudogenes (*ycf15*, *ycf68*) and eight tRNA genes (*trnH*-GUG, *trnI*-CAU, *trnL*-CAA, *trnV*-GAC, *trnI*-GAU, *trnA*-UGC, *trnR*-ACG, *trnN*-GUU) are located within the IR regions. The LSC region contains 82 protein-coding genes and 21 tRNA genes, while the SSC region contains 13 protein-coding genes and one tRNA gene. The rRNA, tRNA and protein-coding genes cover 9040 bp (5.65%), 2873 bp (1.80%) and 79,368 bp (49.6%), respectively, of the complete genome.

Among 113 unique genes, there are 18 intron-containing genes (Additional file 2: Table S1): 16 genes with a single intron and two genes with two introns. Among these, *trnK*-UUU (3216 bp) contains the largest intron, in which *matK* gene (1551 bp) is located. Four pairs of overlapped genes with different ranges of overlapped bases were observed including *atpE* and *atpB* (four overlapped bases), *ndhK* and *ndhC* (10 overlapped bases), *psbD* and *psbC* (53 overlapped bases) and *ndhF* and pseudo-*ycf1* (60 overlapped bases). The frequency of codons in this cp genome was calculated from the exons of protein-coding genes (pseudogenes were omitted) using Maga 6 (Additional file 2: Table S2). The observed initiation codons are AUG, GUG and ACG. The GUG initiation codon was found to be specific for *rps19* and *ndhD*, while the ACG initiation codon was found only for *rpl2*.

Simple sequence repeats (SSRs) and repetitive sequences

Identifications of SSRs and repetitive sequences using by REPuter program (under a cut off $n \geq 10$ with 100% sequence identities) [25] and GMATo v1.2 [26] showed that the cp genome contains, in total, 134 SSR loci and 47 large repeat loci (Additional file 2: Tables S3 and S4). Among the 134 SSRs, 98 and 20 loci are homopolymers and dipolymers, respectively. And, 108 loci are located in intergenic spacer (IGS) regions, while 26 loci are located in the protein-coding genes including *cemA*, *matK*, *ndhD*, *ndhF*, *ndhH*, *rpoC2*, *rps14*, *rps19*, *rps4* and *ycf1*. Neither pentapolymer nor hexapolymer was observed in the protein-coding regions. All 47 large repeat sequences contain four non-tandem direct repeats, six inverted repeats and 37 tandem repeats. The sizes of the repeating unit were in the range of 11–39 bp (Additional file 2: Table S4). Noting that most of the large repeats are located in the IGSs inside the single-copy regions, especially in the large single-copy region. Only eight repeats including one direct repeat and seven tandem repeats are located in the coding sequence of three protein-coding genes: *rpoC2*, *ycf1*, and *ycf2*.



Comparative analysis of the plastid genomes among the Borasseae tribe and phylogenetic analysis among monocots

The cp genomes of 4 species including *B. flabellifer*, *Bismarckia nobilis* (NC_020366.1) [27], *Borassodendron machadonis* (NC_029969.1) [27] and *Lodoicea maldivica* (NC_029960.1) [27], which are members of Borasseae tribe are in the range between 158,144 and 160,021 bp

(Additional file 2: Table S5). The differences in the cp genome sizes are due to the lengths of the LSC, SSC and IR regions. The cp genome of *B. flabellifer* is, so far, the largest among the Borasseae tribe with the longest LSC and SSC regions. These long LSC and SSC regions contain the same number of genes as in the other three cp genomes. Comparative analysis using mVISTA [28] showed that the four cp genomes are highly

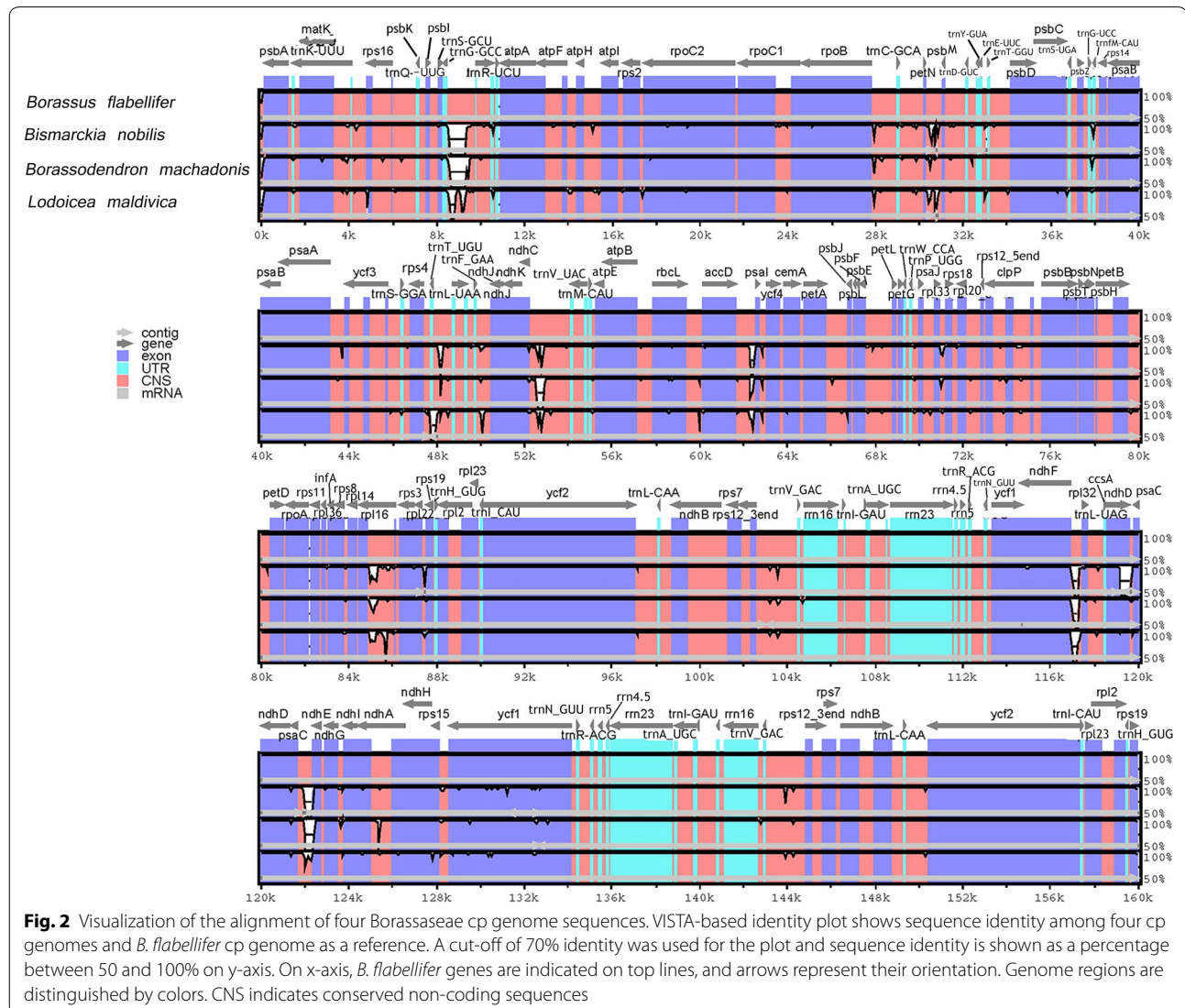
similar (Fig. 2). Among the four sequence regions, both IR regions are more conserved than the LSC and SSC, which contain several variable regions in the intergenic regions such as between *ndhF-rpl32*, *trnG-trnR* and *rpl32-trnL*. Besides, there are 11 small variable regions inside the coding regions of *accD*, *ccsA*, *matK*, *ndhA*, *ndhD*, *ndhE*, *rbcl*, *rpl16*, *rpoC2*, *rps16* and *ycf1*. Noting that *ycf1* and *rpoC2* also carry both SSRs and large repeats.

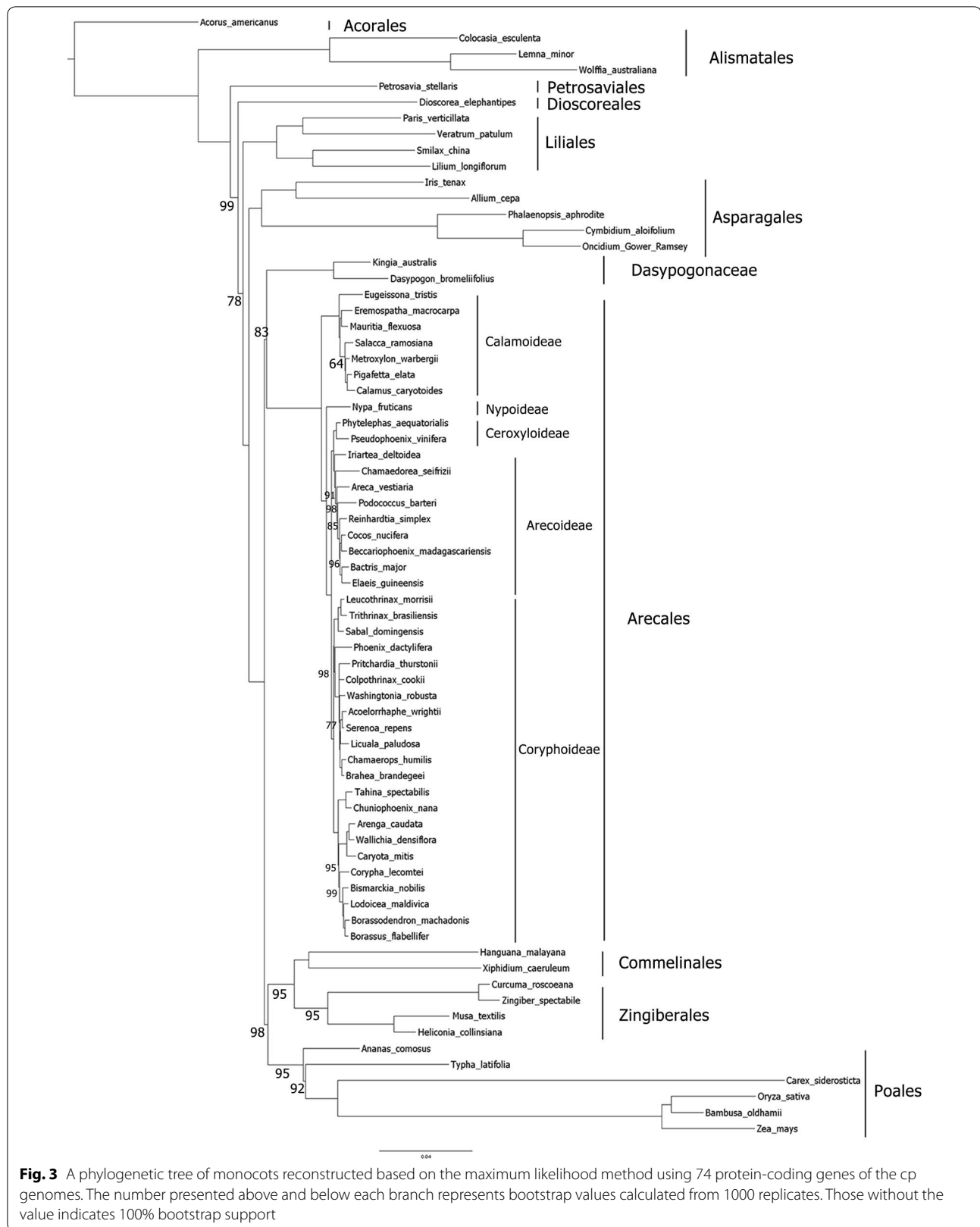
A phylogenetic tree based on the maximum likelihood method were reconstructed with raxmlGUI [29] using 74 protein-coding genes from 70 monocot species. The evolutionary relationship among monocots is presented with high bootstrap supports (Fig. 3). Previously, phylogenetic relationship among subfamilies, families and orders of the Commelinid clade using cp genomes has been described [27, 30, 31], and our result is consistent with

these reports. Tribes within subfamily Coryphoideae was previously divided into two major clades: [(Phoenixae, Livistoneae)(Sabaleae, Cryosophileae)) and (Chuniophoeniceae (Caryota (Coryphoideae, Borasseae)))] [27], and, here, we provide a confirmation for this clustering with 100% bootstrap supports. The phylogenetic tree showed that *B. machadoris* is closely related to *B. flabellifer* as supported by 100% bootstrap replicates. Furthermore, our phylogenetic tree showed that the branch lengths of all members of the family Arecaceae are short, suggesting slow evolutionary rates of the cp genomes in this family.

Limitations

The complete cp genome of *B. flabellifer* reported here provides a valuable resource for genetic analysis of this and related palm species. A number of SSRs, repetitive





sequences and highly variable regions identified here would provide useful markers for studying the genetic diversity and microevolution of this species, although these have to be priority verified in a number of *B. flabellifer* populations. Indeed, further verification of these markers would provide an insight for establishing breeding and conservation programs for this palm species. Because the complete genome sequences of other plants in genus *Borassus* are not yet available, we were able to describe the evolutionary relationship among the members of Borasseae tribe, but not within the genus. Further analysis at the genus level will provide insight into recent evolutionary of palm species.

Additional files

Additional file 1. A schematic procedure for isolation and purification of the chloroplast from *B. flabellifer* (a). The procedure is divided into three main steps: leaf sample preparation and disruption, chloroplast isolation and chloroplast purification. A step sucrose gradient for chloroplast purification, before and after ultra-centrifugation (b). A quality assessment of purified cpDNA using *EcoRI* digestion and agarose gel electrophoresis (c). The left panel represents cpDNA and *EcoRI* treated cpDNA isolated from chloroplast pellets without sucrose gradient purification, while the right panel represents those that isolated from chloroplast pellets with sucrose gradient purification.

Additional file 2. Table S1. Gene annotation of the *B. flabellifer* cp genome. **Table S2.** Codon usages of the *B. flabellifer* cp genome. **Table S3.** Distribution of SSRs in the *B. flabellifer* cp genome. **Table S4.** Large repeat sequences in the *B. flabellifer* cp genome. **Table S5.** Comparison of the sequence sizes in four cp genomes of the Borasseae tribe.

Abbreviations

CP: chloroplast; IGS: intergenic spacer; IR: inverted repeat; ISSR: inter simple sequence repeat; LCS: large single-copy; RAPD: random amplified polymorphic DNA; SSC: small single-copy; SSR: short sequence repeat.

Authors' contributions

SA and SV conceived and designed the experiments. AS, PW and SV performed the experiments, analyzed the data and wrote the manuscript. AS prepared the figures and tables. SA and SV corrected and proofread the manuscript. All authors read and approved the final manuscript.

Author details

¹ Center for Agricultural Biotechnology, Kasetsart University, Kamphaeng Saen Campus, Nakhon Pathom 73140, Thailand. ² Center of Excellence on Agricultural Biotechnology: (AG-BIO/PERDO-CHE), Kasetsart University, 50 Ngarm Wong Wan Road, Chattuchak, Bangkok 10900, Thailand. ³ Department of Genetics, Faculty of Science, Kasetsart University, 50 Ngarm Wong Wan Road, Chattuchak, Bangkok 10900, Thailand. ⁴ Center of Advanced Studies for Tropical Natural Resources, Kasetsart University, 50 Ngarm Wong Wan, Chattuchak, Bangkok 10900, Thailand. ⁵ Special Research Unit in Microalgal Molecular Genetics and Functional Genomics (MMGFG), Department of Genetics, Faculty of Science, Kasetsart University, 50 Ngarm Wong Wan Road, Chattuchak, Bangkok 10900, Thailand. ⁶ School of Science, Mae Fah Luang University, Chiang-Rai 57100, Thailand.

Acknowledgements

We thank Dr. Anongpat Suttangkakul for discussion and proofreading.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The *B. flabellifer* complete cp genome sequence is available in GenBank (Accession Number: KP_901247).

Consent to publish

Not applicable.

Ethics approval and consent to participate

Not applicable.

Funding

This work was supported by Kasetsart University Research and Development Institute (KURDI), Faculty of Science Research Fund (ScRF), the National Research Council of Thailand and Thailand Research Fund (TRF-RSA6080031). Arpakorn Sakulsathaporn is supported by the Center of Excellence on Agricultural Biotechnology, Science and Technology Postgraduate Education and Research Development Office, Office of Higher Education Commission, Ministry of Education. (AG-BIO/PERDO-CHE).

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 3 April 2017 Accepted: 8 December 2017

Published online: 16 December 2017

References

- Morton JF. Notes on distribution, propagation and products of *Borassus* palms (Arecaceae). *Econ Bot.* 1988;2:420–41.
- Haynes J, McLaughlin J. Edible palms and their uses. Gainesville: University of Florida; 2000.
- Bayton RP, Ouédraogo A, Guinko S. The genus *Borassus* (Arecaceae) in West Africa, with a description of a new species from Burkina Faso. *Bot J Linn Soc.* 2006;150:419–27.
- Dransfield J, Beentje H. The palms of Madagascar. The Royal Botanic Gardens, Kew and the International Palm Society; 1995.
- Bayton RP, Obunyal C, Ranaivojaona R. A re-examination of *Borassus* in Madagascar. *Palms.* 2003;47:206–19.
- Eagleton GE. Persistent pioneers; *Borassus* L. and *Corypha* L. in Malesia. *Biodiversitas.* 2016;17(2):716–32.
- Hazarika T, Marak S, Mandal D, Upadhyaya K, Nautiyal B, Shukla A. Underutilized and unexploited fruits of Indo-Burma hot spot, Meghalaya, north-east India: ethno-medicinal evaluation, socio-economic importance and conservation strategies. *Genet Resour Crop Evol.* 2016;63:289–304.
- Davis TA, Johnson DV. Current utilization and further development of the palmyra palm (*Borassus flabellifer* L., Arecaceae) in Tamil Nadu State, India. *Econ Bot.* 1987;41(2):247–66.
- Ambastha K, Hussain SA, Badola R. Resource dependence and attitudes of local people toward conservation of Kabartal wetland: a case study from the Indo-Gangetic plains. *Wetl Ecol and Manag.* 2007;15:287–302.
- George J, Karun A, Manimekalai R, Rajesh M, Remya P. Identification of RAPD markers linked to sex determination in palmyrah (*Borassus flabellifer* L.). *Curr Sci.* 2007;93:1075–7.
- Raju DC, Reji J. Genetic diversity analysis in palmyrah palms using RAPD markers. *Int J Pharma Bio Sci.* 2015;6:244–50.
- Ponnuswami V. Genetic diversity in palmyrah genotypes using morphological and molecular markers. *Electron J Plant Breed.* 2010;1:556–67.
- Pipatchartlearnwong K, Swatdipong A, Vuttipongchaikij S, Apisitwanich S. Cross-genera transferability of microsatellite loci for Asian Palmyra Palm (*Borassus flabellifer* L.). *HortScience.* 2017;52(9):1164–7.
- Pipatchartlearnwong K, Swatdipong A, Vuttipongchaikij S, Apisitwanich S. Genetic evidence of multiple invasions and a small number of founders of Asian Palmyra palm (*Borassus flabellifer*) in Thailand. *BMC Genet.* 2017;18(1):88.
- Triboush SO, Danilenko NG, Davydenko OG. A method for isolation of chloroplast DNA and mitochondrial DNA from sunflower. *Plant Mol Biol Report.* 1998;16(2):183.

16. Sandbrink JM, Vellekoop P, Van Ham RJHC, Van Brederode J. A method for evolutionary studies on RFLP of chloroplast DNA, applicable to a range of plant species. *Biochem Syst Ecol*. 1989;17(1):45–9.
17. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. Accessed 27 Dec 2013.
18. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–20.
19. Huang YY, Matzke AJ, Matzke M. Complete sequence and comparative analysis of the chloroplast genome of coconut palm (*Cocos nucifera*). *PLoS ONE*. 2013;8:e74736.
20. Conant GC, Wolfe KH. GenomeVx: simple web-based creation of editable circular chromosome maps. *Bioinformatics*. 2008;24(6):861–2.
21. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*. 2004;20(17):3252–5.
22. Liu C, Shi L, Zhu Y, Chen H, Zhang J, Lin X, Guan X. CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. *BMC Genomics*. 2012;13(1):715.
23. Yang M, Zhang X, Liu G, Yin Y, Chen K, Yun Q, Zhao D, Al-Mssallem IS, Yu J. The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.). *PLoS ONE*. 2010;5:e12762.
24. Schattner P, Brooks AN, Lowe TM. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res*. 2005;33(suppl 2):686–9.
25. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res*. 2001;29(22):4633–42.
26. Wang X, Lu P, Luo Z. GMATo: a novel tool for the identification and analysis of microsatellites in large genomes. *Bioinformatics*. 2013;9(10):541–4.
27. Barrett CF, Baker WJ, Comer JR, Conran JG, Lahmeyer SC, Leebens-Mack JH, Li J, Lim GS, Mayfield-Jones DR, Perez L. Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. *New Phytol*. 2016;209:855–70.
28. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. *Nucleic Acids Res*. 2004;32(suppl 2):273–9.
29. Silvestro D, Michalak I. raxmlGUI: a graphical front-end for RAxML. *Org Divers Evol*. 2012;12(4):335–7.
30. Givnish TJ, Ames M, McNeal JR, McKain MR, Steele PR, Graham SW, Pires JC, Stevenson DW, Zomlefer WB, Briggs BG, Duvall MR. Assembling the tree of the monocotyledons: plastome sequence phylogeny and evolution of poales. *Ann Mo Bot Gard*. 2010;97(4):584–616.
31. Barrett CF, Davis JI, Leebens-Mack J, Conran JG, Stevenson DW. Plastid genomes and deep relationships among the commelinid monocot angiosperms. *Cladistics*. 2013;29(1):65–87.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit



Cross-genera Transferability of Microsatellite Loci for Asian Palmyra Palm (*Borassus flabellifer* L.)

Kwanjai Pipatchartlearnwong, Akarapong Swatdipong, and Supachai Vuttipongchaikij¹

Department of Genetics, Faculty of Science, Kasetsart University, 50 Ngam Wong Wan Road, Chatuchak, Bangkok 10900, Thailand; and Center of Advanced Studies for Tropical Natural Resources, Kasetsart University, Ngam Wong Wan Road, Chatuchak, Bangkok 10900, Thailand

Somsak Apisitwanich¹

Department of Genetics, Faculty of Science, Kasetsart University, 50 Ngam Wong Wan Road, Chatuchak, Bangkok 10900, Thailand; Center of Advanced Studies for Tropical Natural Resources, Kasetsart University, Ngam Wong Wan Road, Chatuchak, Bangkok 10900, Thailand; and School of Science, Mae Fah Luang University, Chiang-Rai 57100, Thailand

Additional index words. *Arecaceae*, expressed sequence tag-simple sequence repeat (EST-SSR), genomic simple sequence repeat (gSSR), polymorphism information content (PIC)

Abstract. Asian Palmyra palm, found throughout south and southeast Asia, is important for local economies, especially for sugar palm production. Unlike its related species, such as oil palm and coconut, only a few genetic markers are available for Asian Palmyra palm. In this study, we tested the transferability of molecular markers derived from oil palm, and a set of selected markers were used for evaluating the diversity of Asian Palmyra palm growing in Thailand. From 545 primer pairs of expressed sequence tag-simple sequence repeat (EST-SSR) and genomic simple sequence repeat (gSSR) markers, 317 (58.17%) primer pairs were able to amplify the Asian Palmyra palm DNA, and 19 (5.99%) pairs were polymorphic. After extensively genotyping 164 samples from 12 populations, we obtained 25 loci with the polymorphism information content (PIC) average of 0.37 and allele numbers ranging from one to five. The observed and expected heterozygosity ranged from 0 to 1 and 0 to 0.76, respectively. A dendrogram showed separation of the palm populations into two clades, between north-eastern and southern-central regions. This study provides a set of microsatellite markers for use in further genetic studies of Asian Palmyra palm.

Asian Palmyra palm ($2n = 36$), found widespread in the Indian subcontinent and Southeast Asia, is a monocotyledonous dioecious woody perennial tree in the *Arecaceae* family. This palm tree is important for local agriculture and economies as its inflorescence sap is used for palm sugar production and its fruits are widely consumed (Lim, 2012; Morton, 1988). The Asian Palmyra palm has a very slow growth rate and requires 12–20 years to produce its first inflorescence flowers, only then the sex is revealed (Davis and Johnson, 1987). It is widely believed that Asian Palmyra palm

originated in Africa and was introduced to India and then into the southeast Asia more than a thousand years ago (Kovoor, 1983).

With such a long juvenile stage and without any direct sex determination, most growers are reluctant to cultivate new crops, and, with the expansion of farmlands and urbanization, Asian Palmyra palm population is in rapid decline. Conservation plans and improvement for use are operating in South and Southeast Asia to conserve this plant species (Barfod et al., 2015; Davis and Johnson, 1987; Sirajuddin et al., 2016). However, genetic data of Asian Palmyra palm are currently limited, and only a few molecular markers including RAPDs and ISSRs have been reported (George et al., 2016; Vinayagam et al., 2009). Markers with higher information such as microsatellites are needed to evaluate the genetic groups and diversity of the Asian Palmyra palm before establishing effective conservation plans.

New microsatellite markers require a high developing cost. Nonetheless, many microsatellite markers have been shown to be transferable across plant species and genera, providing an economical way for marker

development (Karaca et al., 2013; Whankaew et al., 2011; Zehdi et al., 2012). Oil palm (*Elaeis guineensis*), a closely related species to Asian Palmyra palm, has available genome sequences and a wide range of developed microsatellite markers including genomic SSR (gSSR) and EST-SSR (Billotte et al., 2005; Singh et al., 2013; Ukoskit et al., 2014). Thus, it is feasible to apply these markers on the Asian Palmyra palm. In this study, we evaluated the cross-genera transferability of the gSSRs and EST-SSRs derived from oil palm and determined the diversity of some Asian Palmyra palm populations in Thailand.

Materials and Methods

Plant material, DNA isolation, and marker analysis. Leaf samples were collected from 164 accessions located in 12 provinces in three geographical regions in Thailand (Table 1). Oil palm was used as an out-group. Because the original CTAB (Cetyltrimethylammonium bromide) method (Gawel and Jarret, 1991) gave low yields when isolating DNA from the leaf samples of Asian Palmyra palm. A medication was made by adding 10% polyvinylpyrrolidone in the extraction buffer to help improving the DNA yield. DNA samples from one oil palm and 10 randomly selected Asian Palmyra palm accessions were used for an initial screening using markers previously developed from the oil palm: 289 EST-SSR (Ukoskit et al., 2014) and 256 gSSRs (Billotte et al., 2005). The polymerase chain reaction (PCR) reactions were performed in a 20 μ L volume containing 20 ng of total DNA, 20 μ M of each dNTP, 0.25 μ M of each primer, 10X PCR buffer (with 1.5 mM $MgCl_2$), and 0.5 U *Taq* DNA polymerase (Vivantis Technologies, Selangor Darul Ehsan, Malaysia). Thermocycling parameters included an initial denaturation step of 5 min at 94 $^{\circ}C$, followed by 35 cycles of 30 s at 94 $^{\circ}C$, 1.30 min annealing temperature following Temp in Table 2, 30 s extension step at 72 $^{\circ}C$, and 8 min at 72 $^{\circ}C$ for the final extension step. The PCR products were separated and resolved by 6% polyacrylamide gel electrophoresis with silver staining. Primer pairs, which generated polymorphic markers, were then used for analyzing the 164 accessions. Fragments were scored based on the Low Molecular Weight DNA Ladder (New England Biolabs, Ipswich, MA) and converted into binary data.

Data analysis. Linkage disequilibrium (LD) was calculated using PowerMarker version 3.2.5 (Liu and Muse, 2005), and sequential Bonferroni correction (Holm, 1979) was performed according to the multiple comparisons of the LD tests. Some loci were excluded after observing significant LD. The Hardy–Weinberg equilibrium (HWE) was tested using POPGENE version 1.31 (Yeh et al., 1999). Polymorphic EST-SSR and gSSR loci were used for calculating the number of alleles per locus (N_A), observed heterozygosity (H_o), expected heterozygosity (H_e), and polymorphism information

Received for publication 2 June 2017. Accepted for publication 29 June 2017.

This work was supported by Kasetsart University Research and Development Institute (KURDI), Kasetsart University Doctoral Degree Scholarships and Faculty of Science Research Fund (ScRF) and Thailand Research Fund (TRF-RSA6080031).

We thank Passorn Wonnapijit and Anongpat Suttangkakul for proofreading this article.

¹Corresponding authors. E-mail: fsciscv@ku.ac.th or fscissa@ku.ac.th.

Table 1. Location and the number of Asian Palmyra palm samples used in this study.

Part of Thailand	Province	Code	Sample numbers	Location
North-Eastern	Burirum	BU	11	14°36'53.9"N 102°47'17.1"E
	Sisaket	SS	9	14°50'23.5"N 104°11'37.2"E
	Nakhon-Ratchasima	NR	10	15°09'04.3"N 102°10'54.6"E
	Kalasin	KB	6	16°20'09.9"N 103°38'09.3"E
Central	Phetchaburi	PH	18	13°03'33.8"N 99°54'07.8"E
	Nakhon Pathom	NP	11	13°52'13.4"N 100°10'18.9"E
	Kanchanaburi	KA	22	14°03'04.9"N 99°32'46.6"E
	Phachinburi	PC	18	14°07'30.7"N 101°31'20.3"E
	Chainat	CN	14	15°06'22.6"N 100°14'37.7"E
	Nakhon Sawan	NS	14	15°52'11.7"N 100°15'35.2"E
Southern	Songkhla	SK	18	7°27'40.8"N 100°26'29.6"E
	Surat Thani	ST	13	9°24'56.8"N 99°12'11.3"E

Table 2. Characteristics of 19 polymorphic microsatellite markers tested in the Asian Palmyra palm.

Primer name	Repeat motif	Temp (°C)	Allele size (bp)	N _a ^z	H _o ^y	H _e ^x	PIC ^w
ESSR75	(AAG) ₅	60	270–285	2.17	0.4	0.5	0.38
			175–190	2.33	0.1	0.38	0.48
			350–365	3.33	0.97	0.62	0.54
ESSR76	(AGC) ₅	60	190–205	2.75	0.23	0.51	0.55
			200–215	2	0.54	0.35	0.39
			230–242	2.67	0.77	0.51	0.42
ESSR82	(GCT) ₅	60	220–239	2.08	0.87	0.51	0.38
ESSR332	(AT) ₆	60	240–254	2.83	0.4	0.44	0.48
ESSR553	(A) ₁₉	60	125–139	2.92	0.35	0.52	0.6
ESSR566	(AG) ₇	60	150–164	2.58	0.55	0.41	0.38
ESSR609	(GA) ₇	59	200–228	1.92	0.24	0.33	0.34
ESSR650	(AG) ₁₄	59	150–168	2	0.26	0.34	0.32
ESSR652	(GAG) ₆	60	150–168	1.83	0.23	0.19	0.18
ESSR673	(GGC) ₈	59	75–113	2	0.3	0.37	0.34
ESSR681	(AAAT) ₅	59	350–370	2.42	0.46	0.43	0.47
mEgCIR2332	(GA) ₁₄	55	250–278	3.25	0.49	0.43	0.38
mEgCIR3295	(GT) ₇ (GA) ₂₃	55	300–360	4.17	0.33	0.61	0.65
			250–310	2.17	0.14	0.17	0.17
mEgCIR3311	(GA) ₁₅	55	180–210	2.25	0.51	0.36	0.34
mEgCIR3413	(GA) ₁₈	55	250–286	2.5	0.8	0.52	0.43
mEgCIR3477	(GA) ₂₂	55	170–214	2.17	0.11	0.15	0.16
mEgCIR3592	(GA) ₂₀	55	150–190	2.58	0.37	0.34	0.34
mEgCIR3755	(GA) ₁₅	55	380–410	1.17	0	0.02	0.02
			140–170	1.67	0.12	0.1	0.12
mEgCIR3788	(GA) ₁₈	55	180–216	2.58	0.73	0.49	0.41
				2.41	0.41	0.38	0.37

^zNumber of alleles.^yObserved heterozygosity.^xExpected heterozygosity.^wPolymorphic information content.

content (PIC) value (Botstein et al., 1980) using POPGENE version 1.31 (Yeh et al., 1999). A dendrogram was constructed using the EST-SSR and gSSR loci based on Nei's standard genetic distance (Nei, 1972) and the UPGMA (Unweighted Pair Group Method with Arithmetic Mean) clustering method in Populations software version 1.2.32 (Langella, 1999).

Results and Discussion

An initial marker screening using 289 EST-SSR and 256 gSSR primers derived from oil palm provided 154 (53.3%) and 163 (63.7%) amplifiable markers, respectively. Polymorphic bands were observed from 11 EST-SSR and eight gSSR primers, and these were subsequently used for evaluating the polymorphism of the palm populations. Analysis of 164 accessions using these 19 primers provided 36 polymorphic loci (21 EST-SSR and 15 gSSR loci), but only 15 EST-SSR and 10 gSSR loci were not significant

in the LD and HWE tests after the sequential Bonferroni correction ($P > 0.05$) (Table 2). The average PIC value of combined EST-SSRs and gSSRs was 0.37. An average 2.41 alleles per locus was observed. The means of observed and expected heterozygosity were 0.41 and 0.38, respectively.

A dendrogram of 12 populations showed that the populations were clustered into two main clades, clearly separated from the oil palm outgroup (Fig. 1). The north-eastern clade was separated from those of the southern and central clade. The two clades were geographically separated on the map of Thailand and this likely resulted from geographical barriers (Pyšek et al., 2008). Phayayen hill and the Phetchabun mountain ranges located between the two areas had interrupted human movement in the past and, thereby, limited the anthropogenic spread of the Asian Palmyra palm. By contrast, there is no geographical barrier between the southern and central regions resulting in a considerable widespread of the species.

Transferability of SSRs between species or genera has been reported in many plant species including cereals (Castillo et al., 2010; Ince et al., 2010; Sim et al., 2009; Tang et al., 2006) and woody species (Gasic et al., 2009; Park et al., 2010; Yu et al., 2011). The cross-genera transferability of EST-SSRs and gSSRs shown in this study ranged between 53.3% and 63.7%, which are within the 30% to 65% reported for the transferability in other plant species including legumes and grasses (Choudhary et al., 2009; Karaca et al., 2013; Whankaew et al., 2011; Yu et al., 2013; Zeid et al., 2010; Zhou et al., 2013). The transferred EST-SSR and gSSR markers will be a valuable tool for a variety of genetic analyses for the Asian Palmyra palm. In addition, with the availability of genome sequence of date palm (*Phoenix dactylifera*) (Al-Dous et al., 2011; Al-Mssallem et al., 2013), a closely related species to Asian Palmyra palm, the numbers of transferable markers for the Asian Palmyra palm could be further expanded.

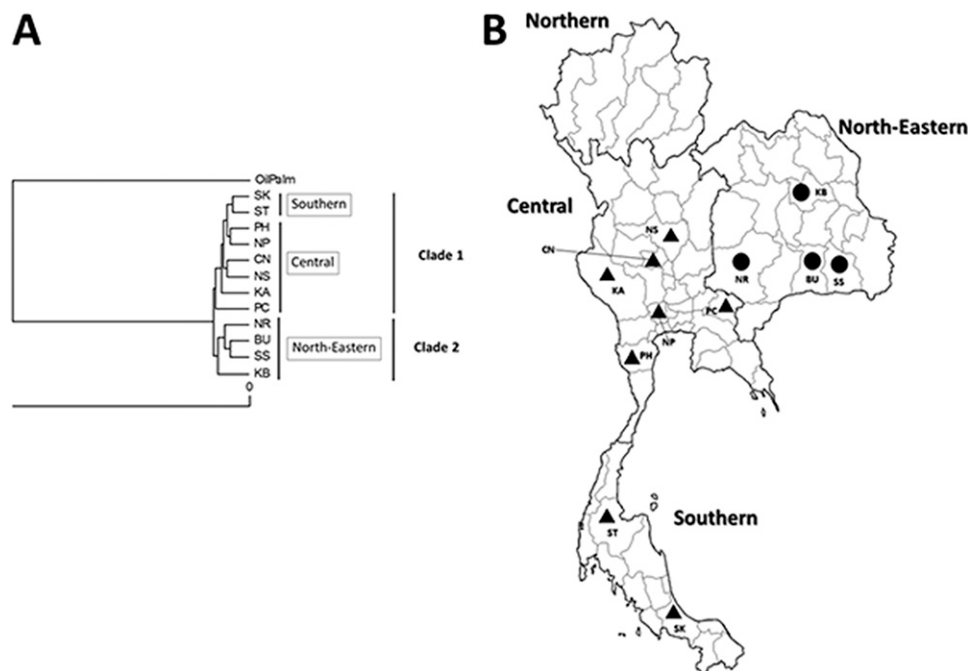


Fig. 1. (A) UPGMA dendrogram representing the relationship among Asian Palmyra palm populations in Thailand. Clade1 consists of southern (SK and ST) and central (PH, NP, CN, NS, KA, and PC) populations while clade2 is only north-eastern (NR, BU, SS, and KB) populations. (B) A map of Thailand indicating the distribution of two distinct genetic groups based on the UPGMA dendrogram.

The genetic study of the Asian Palmyra palm was, so far, limited to a few markers, for example, RAPDs developed for sex determination (George and Karun, 2011; George et al., 2007), and ISSRs used for diversity assessment in India (Vinayagam et al., 2009). These, however, were unable to provide a genetic resolution for the population structure of the Asian Palmyra palm because of their limitation as dominant markers. This study provides a set of codominant microsatellite EST-SSR and gSSR markers, which would be used to further study the genetics of Asian Palmyra palm. Furthermore, we found that the population diversity of Asian Palmyra palm in Thailand is very low as indicated by the genetic variation indices (PIC, N_a , H_o , and H_e). This observation is somewhat similar to previous studies in India using ISSRs that showed 0.84 average similarity coefficient and PIC ranging between 0 and 0.56 (Vinayagam et al., 2009) and RAPDs that had 0.76 average similarity coefficient (George et al., 2016; Ponnuswami, 2010; Ponnuswami et al., 2008; Raju and Reji, 2015). The small number of alleles per locus coincides with the notion that Asian Palmyra palm is an introduced species (Kovoor, 1983), and, with limited numbers of originally introduced plants, this eventually resulted in low genetic diversity.

In conclusion, the evaluation of marker transferability in this study provided a set of microsatellite markers for studying the genetic diversity and population structure of the Asian Palmyra palm in Thailand. This data could be useful for the management of the Asian Palmyra palm plantation and advance its breeding efforts.

Literature Cited

- Al-Dous, E.K., B. George, M.E. Al-Mahmoud, M.Y. Al-Jaber, H. Wang, Y.M. Salameh, E.K. Al-Azwani, S. Chaluvadi, A.C. Pontaroli, J. DeBarry, and V. Arondel. 2011. De novo genome sequencing and comparative genomics of date palm (*Phoenix dactylifera*). *Nat. Biotechnol.* 29(6):521–527.
- Al-Mssallem, I.S., S. Hu, X. Zhang, Q. Lin, W. Liu, J. Tan, X. Yu, J. Liu, L. Pan, T. Zhang, Y. Yin, C. Xin, H. Wu, G. Zhang, M.M. Ba Abdullah, D. Huang, Y. Fang, Y.O. Alnakhli, S. Jia, A. Yin, E.M. Alhuzimi, B.A. Alsaihati, S.A. Al-Owayyed, D. Zhao, S. Zhang, N.A. Al-Otaibi, G. Sun, M.A. Majrashi, F. Li, Tala, J. Wang, Q. Yun, N.A. Alnassar, L. Wang, M. Yang, R.F. Al-Jelaify, K. Liu, S. Goa, K. Chen, S.R. Alkhaldi, G. Liu, M. Zhang, H. Guo, and J. Yu. 2013. Genome sequence of the date palm *Phoenix dactylifera* L. *Nat. Commun.* 4:2274.
- Barfod, A.S., M. Balhara, J. Dransfield, and H. Balslev. 2015. SE Asian palms for agroforestry and home gardens. *Forests* 6(12):4607–4616.
- Billotte, N., N. Marseillac, A.M. Risterucci, B. Adon, P. Brottier, F.C. Baurens, R. Singh, A. Herrán, H. Asmady, C. Billot, P. Amblard, T. Durand-Gasselin, B. Courtois, D. Asmono, S.C. Cheah, W. Rohde, E. Ritter, and A. Charrier. 2005. Microsatellite-based high density linkage map in oil palm (*Elaeis guineensis* Jacq.). *Theor. Appl. Genet.* 110(4):754–765.
- Botstein, D., R.L. White, M. Skolnick, and R.W. Davis. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Amer. J. Hum. Genet.* 32(3):314–331.
- Castillo, A., H. Budak, A.C. Martin, G. Dorado, A. Borner, M. Roder, and P. Hernandez. 2010. Interspecies and intergenus transferability of barley and wheat D-genome microsatellite markers. *Ann. Appl. Biol.* 156:347–356.
- Choudhary, S., N.K. Sethy, B. Shokeen, and S. Bhatia. 2009. Development of chickpea EST-SSR markers and analysis of allelic variation across related species. *Theor. Appl. Genet.* 118:591–608.
- Davis, T.A. and D.V. Johnson. 1987. Current utilization and further development of the palmyra palm (*Borassus flabellifer* L., *Arecaceae*) in Tamil Nadu State, India. *Econ. Bot.* 41(2):247–266.
- Gasic, K., Y. Han, S. Kerbundit, V. Shulaev, A.F. Iezzoni, E.W. Stover, R.L. Bell, and M.E. Wisniewski. 2009. Characteristics and transferability of new apple EST-derived SSRs to other Rosaceae species. *Mol. Breed.* 23:397–411.
- Gawel, N.J. and R.L. Jarret. 1991. A modified CTAB DNA extraction procedure for Musa and Ipomoea. *Plant Mol. Biol. Rpt.* 9(3):262–266.
- George, J. and A. Karun. 2011. Marker assisted detection of seed sex ratio in palmyrah palm (*Borassus flabellifer* L.). *Curr. Sci.* 100(6):922–925.
- George, J., A. Karun, R. Manimekalai, M.K. Rajesh, and P. Remya. 2007. Identification of RAPD markers linked to sex determination in palmyrah (*Borassus flabellifer* L.). *Curr. Sci.* 93:1075–1077.
- George, J., K. Venkataramana, P. Nainar, M. Rajesh, and A. Karun. 2016. Evaluation of molecular diversity of ex situ conserved germplasm of palmyrah (*Borassus flabellifer* L.) accessions using RAPD markers. *J. Plant. Crops* 44(2):96–102.
- Holm, S. 1979. A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* 6(2):65–70.
- Ince, A.G., M. Karaca, and A.N. Onus. 2010. Polymorphic microsatellite markers transferable across *Capsicum* species. *Plant Mol. Biol. Rpt.* 28:285–291.
- Karaca, M., A.G. Ince, A. Aydin, and S.T. Ay. 2013. Cross-genera transferable microsatellite markers for 12 genera of the *Lamiaceae* family. *J. Sci. Food Agr.* 93(8):1869–1879.


- Kovoor, A. 1983. The palmyra palm: potential and perspectives FAO plant production and protection.
- Langella, O. 1999. Populations 1.2.28 (2002): A population genetic software. CNRS UPR9034.
- Lim, T.K. 2012. Edible medicinal and non-medicinal plants. Springer, New York, NY.
- Liu, K. and S.V. Muse. 2005. PowerMarker: An integrated analysis environment for genetic marker analysis. *Bioinformatics* 21:2128–2129.
- Morton, J. 1988. Notes on distribution, propagation, and products of *Borassus* palms (*Areaceae*). *Econ. Bot.* 42(3):420–441.
- Nei, M. 1972. Genetic distance between populations. *Amer. Nat.* 106:283–292.
- Park, Y.H., S.G. Ahn, Y.M. Choi, H.J. Oh, D.C. Ahn, J.G. Kim, J.S. Kang, Y.W. Choi, and B.R. Jeong. 2010. Rose (*Rosa hybrid* L.) EST-derived microsatellite markers and their transferability to strawberry (*Fragaria* spp.). *Sci. Hort.* 125:733–739.
- Ponnuswami, V. 2010. Genetic diversity in palmyrah genotypes using morphological and molecular markers. *Electron. J. Plant Breeding* 1(4):556–567.
- Ponnuswami, V., R. Jagadeesan, A. Ramesh Kumar, M. Prabhu, and S. Makesh. 2008. Genetic relationship and diversity in palmyrah accessions based on RAPD marker. *Eurasian J. Sustain. Agr.* 2:165–171.
- Pyšek, P., D.M. Richardson, J. Pergl, V. Jarošík, Z. Sixtová, and E. Weber. 2008. Geographical and taxonomic biases in invasion ecology. *Trends Ecol. Evol.* 23(5):237–244.
- Raju, D.C. and J.V. Reji. 2015. Genetic diversity analysis in palmyrah palms using RAPD markers. *Intl. J. Pharma Bio Sci.* 6:244–250.
- Sim, S.C., J.K. Yu, Y.K. Jo, M.E. Sorrells, and G. Jung. 2009. Transferability of cereal EST-SSR markers to ryegrass. *Genome* 52:431–437.
- Singh, R., M. Ong-Abdullah, E.T.L. Low, M.A.A. Manaf, R. Rosli, R. Nookiah, L.C.L. Ooi, S.E. Ooi, K.L. Chan, M.A. Halim, and N. Azizi. 2013. Oil palm genome sequence reveals divergence of interfertile species in old and new worlds. *Nature* 500(7462):335–339.
- Sirajuddin, S., M. Mulyadi, G.D. Dirawan, F. Amir, and N. Pertiwi. 2016. Conservation status of lontar palm trees (*Borassus flabellifer* Linn) in Jeneponto district, South Sulawesi, Indonesia. *J. Trop. Crop Sci.* 3(1):28–33.
- Tang, J., Y. Gao, and J. Jia. 2006. Homologous analysis of SSR-ESTs and transferability of wheat SSR-EST markers across barley, rice, and maize. *Euphytica* 151:87–93.
- Ukoskit, K., V. Chanroj, G. Bhusudsawang, K. Pipatchartlearnwong, S. Tangphatsomruang, and S. Tragoonrungs. 2014. Oil palm (*Elaeis guineensis* Jacq.) linkage map, and quantitative trait locus analysis for sex ratio and related traits. *Mol. Breed.* 33(2):415–424.
- Vinayagam, P., J. Dhandapani, P. Raman, R.K. Alagarsamy, P. Muthusamy, and S.S. Balaraman. 2009. Molecular characterization for genetic diversity of palmyrah (*Borassus flabellifer*) accessions using Inter Simple Sequence Repeat (ISSR) markers. *Asian Austral. J. Plant Sci. Biotechnol.* 3(1):11–15.
- Whankaew, S., S. Kanjanawattanawong, C. Phumichai, D.R. Smith, J. Narangajavana, and K. Triwitayakorn. 2011. Cross-genera transferability of (simple sequence repeat) SSR markers among cassava (*Manihot esculenta* Crantz), rubber tree (*Hevea brasiliensis* Muell. Arg.) and physic nut (*Jatropha curcas* L.). *Afr. J. Biotechnol.* 10(10):1768–1776.
- Yeh, F.C., R.C. Yang, and T. Boyle. 1999. POPGENE version 1.31: Microsoft Windows-based freeware for population genetic analysis, quick user guide. Center for International Forestry Research, University of Alberta, Edmonton, Alberta, Canada.
- Yu, F., B.H. Wang, S.P. Feng, J.Y. Wang, W.G. Li, and Y.T. Wu. 2011. Development, characterization, and cross-species/genera transferability of SSR markers for rubber tree (*Hevea brasiliensis*). *Plant Cell Rpt.* 30:335–344.
- Yu, J., H. Zhao, T. Zhu, L. Chen, and J. Peng. 2013. Transferability of rice SSR markers to *Miscanthus sinensis*, a potential biofuel crop. *Euphytica* 191(3):455–468.
- Zehdi, S., E. Cherif, S. Rhouma, S. Santoni, A.S. Hannachi, and J.C. Pintaud. 2012. Molecular polymorphism and genetic relationships in date palm (*Phoenix dactylifera* L.): The utility of nuclear microsatellite markers. *Sci. Hort.* 148:255–263.
- Zeid, M., J.K. Yu, I. Goldowitz, M.E. Denton, D.E. Costich, C.T. Jayasuriya, M. Saha, R. Elshire, D. Benscher, F. Brescghello, J. Munkvold, R.K. Varshney, G. Belay, and M.E. Sorrells. 2010. Cross-amplification of EST-derived markers among 16 grass species. *Field Crops Res.* 118(1):28–35.
- Zhou, Y., J. Li, S.S. Korban, and Y. Han. 2013. Apple SSRs present in coding and noncoding regions of expressed sequence tags show differences in transferability to other fruit species in *Rosaceae*. *Can. J. Plant Sci.* 93(2): 183–190.

RESEARCH ARTICLE

Open Access



Evaluation of strategies for improving the transgene expression in an oleaginous microalga *Scenedesmus acutus*

Anongpat Suttangkakul^{1,2}, Anchalee Sirikhachornkit^{1,2}, Piyada Juntawong^{1,2}, Wilasinee Puangtame¹, Thitikorn Chomtong¹, Suchada Srifa¹, Sukhita Sathitnaitham¹, Wasawat Dumrongthawatchai¹, Kanidtha Jariyachawalid³ and Supachai Vuttipongchaikij^{1,2*} 

Abstract

Background: Genetic transformation of microalgae has been hampered by inefficient transgene expression, limiting the progress of microalgal biotechnology. Many vector tools and strategies have been developed in recent years to improve transgene expression in the model microalga *Chlamydomonas*, but these were hardly applied to other microalgae. In this work, naturally-isolated oleaginous microalgae were accessed for genetic transformation, and various expression systems were evaluated in a selected microalga to circumvent inefficient transgene expression.

Results: Initially, a strain of *Scenedesmus acutus* was selected from the oleaginous microalgal collection based on its highest transformation rate and transgene stability. This strain, which had very low or no GFP reporter expression, was first tested to improve transgene expression by using intron-containing constructs and the transcript fusion using *ble::E2A*. The intron-containing constructs yielded 2.5–7.5% of transformants with 2–4-fold fluorescence signals, while the majority of the transformants of the transcript fusion had the fluorescence signals up to 10-fold. Subsequently, three UV-induced *S. acutus* mutants were isolated with moderate increases in the level and frequency of transgene expression (2–3-fold and 10–12%, respectively). Finally, a transcript fusion system was developed using *psy* white mutants with an expression vector containing *PSY::E2A* for complementation and light selection. Transformants with green colonies were selected under light exposure, and the transgene expression was detected at protein levels. Although the improvement using *PSY::E2A* was only minor (1–2-fold increase and ~ 7% of transformants), this system provides an alternative selectable marker that is compatible with large-scale culture.

Conclusions: Here, the overall improvement of transgene expression using the *Chlamydomonas* tools was moderate. The most effective tool so far is the transcript fusion using *ble::E2A* system. This work demonstrates that, so far, genetic engineering of non-model microalgae is still a challenging task. Further development of tools and strategies for transgene expression in microalgae are critically needed.

Keywords: *Agrobacterium*-mediated transformation, Biofuel, Fluorescence reporter, Introns, Lipid, Microalgae, Mutagenesis, Transcript-fusion

* Correspondence: fsciscv@ku.ac.th

¹Special Research Unit in Microalgal Molecular Genetics and Functional Genomics (MMGFG), Department of Genetics, Faculty of Science, Kasetsart University, 50 Ngam Wong Wan road, Chatuchak, Bangkok 10900, Thailand

²Center of Advanced studies for Tropical Natural Resources, Kasetsart University, 50 Ngam Wong Wan road, Chatuchak, Bangkok 10900, Thailand
Full list of author information is available at the end of the article



Background

Oleaginous microalgae emerged as a promising renewable feedstock for biofuel production [1–3]. These microalgae can be cultivated using wastewater, seawater or simple medium and, under certain conditions such as nutrient deprivation or stresses, they can accumulate high lipid contents [4]. For decades, microalgae have been isolated and screened from natural sources for properties beneficial to biofuel production including rapid cell growth, high lipid yields, tolerance to stresses, ease of harvest and ability to secrete lipid [5]. To date, however, these microalgae are unable to produce biofuels at competitive costs compared to those of fossil fuels. Genetic improvement is a major step for making algal biofuels economically feasible [6, 7].

Genetic transformation has not been widely used in microalgae, partly because of difficulties in nuclear gene transformation, as reported in the model microalga *Chlamydomonas reinhardtii* and other species for low transformation efficiencies [8], inefficient transgene expression [9], transgene cleavages [9, 10], transgene instability [11] and highly control transcriptional silencing [12, 13]. Genome editing was also proved to be very difficult as the homology-directed repair (HDR) is extremely inefficient [14, 15] and CRISPR/Cas9 system was shown to be toxic for the algal cell [16]. The only exception so far is a marine microalga *Nannochloropsis*, which was reported for an efficient nuclear transformation and its feasibility for genome editing by either HDR [17] and CRISPR/Cas9 [18]. Initially, many attempts were made for improving transgene expression in *Chlamydomonas* by codon optimization [19] and screening large numbers of transformants to minimize transgene positional effects [20], but little improvement was achieved. Recent developments have demonstrated marked improvements by a number of approaches including mutant isolations for reduced transcriptional silencing [21] and the incorporation of introns in the expression constructs [22] and transcript fusion system [23]. However, most of these tools were developed for *Chlamydomonas*, and only a few have been applied in other microalgae [24].

Various oleaginous microalgae have been tested for genetic transformation, and, indeed, they suffer from inefficient transgene expression as *Chlamydomonas* had in the past [25, 26]. Notably, most of the work reported so far relied on the gene or enzymatic activity to represent the transgene expression, without confirmation at protein levels [24, 27–29]. This shortcoming is problematic as reliable detection at protein levels is the key indicator for establishing stable transgene expression, before evaluating the effects of transgenes such as phenotypes or compound production. Along the line of improvements in *Chlamydomonas*, many developed tools and

strategies are now available, and it would be proven fruitful if these could be used for improving other oleaginous microalgae.

Here, ten oleaginous microalgae isolated from natural sources in Thailand were tested for genetic transformation. Initially, some of these microalgae were tested by electro-transformation, but very few or no transformant was obtained. In contrast, *Agrobacterium*-mediated transformation yielded much more transformants and was used to access the transformation efficiency of the ten microalgae. A strain of *S. acutus*, which has rapid growth and high lipid accumulation, was selected based on its high transformation rate and transgene stability. As expected, this strain has a low capability for transgene expression as indicated by very low levels of GFP reporter. This strain was then used for evaluating various expression tools, which were previously developed in *Chlamydomonas*, whether they could mitigate the low transgene expression problem. These included the use of constructs containing introns, constructs with transcript fusion and mutant isolation for reduced transcriptional gene silencing. Furthermore, a transgene expression system using novel selectable marker was developed by generating *psy* white mutants and complementation using *PSY* gene and light as the mean of selection. Transgene expression was analyzed for both expression levels and the frequency of expressors among the transformants.

Results

Agrobacterium-mediated transformation of wild-isolated oleaginous microalgal strains

Ten wild-isolated freshwater unicellular microalgae with high lipid accumulations were tested for *Agrobacterium*-mediated transformation using a method based on Kumar et al. [30]. Before transformation, axenic cultures of the ten strains were verified for hygromycin B sensitivity to estimate the selective concentrations (Additional file 1). After transformation using *Agrobacterium* harboring pCXS_N-GFP, hygromycin B resistant colonies were obtained from six out of ten strains with transformation rates approximately 10–200 CFU per 10⁶ cells (Table 1, see also Additional file 2 for an example of selection plates). Some background colonies conferring spontaneous resistant to hygromycin B were observed from TISTR8511, 8555, 8519, 8540 and 8447, accounting for 5–9% of total colonies counted in transformation plates. Two *S. acutus* strains with the highest transformation rates, TISTR8540 and 8447, were tested for assessing the transgene stability. After 10–12 rounds of subcultures on the non-selective medium before transferring onto the hygromycin B selective medium, the growth of the TISTR8447 transformants was more consistent on the selective medium than those of TISTR8540 transformants

Table 1 Transformation rates of ten microalgal strains by *Agrobacterium*-mediated transformation

Species	Algal media	Concentration of hygromycin B ($\mu\text{g ml}^{-1}$)	Selection period (days)	Transformation rate (CFU per 10^6 cells)	Background control (CFU per 10^6 cells)
<i>Coelastrum</i> sp. (TISTR 8511)	TAP	10	15	0 ($n = 44$)	0
<i>S. acutus</i> (TISTR 8555)	TAP	30	15	20 ± 5.5 ($n = 12$)	1.5 ± 1.3
<i>S. acutus</i> (TISTR 8540)	TAP	50	10	72 ± 41 ($n = 8$)	3.7 ± 3.5
<i>S. acuminatus</i> (TISTR 8519)	TAP	50	10	31 ± 23 ($n = 11$)	2.7 ± 2.5
<i>S. acutus</i> (TISTR 8447)	TAP	30	15	217.5 ± 75 ($n = 12$)	19.5 ± 6.8
<i>C. humicola</i> (TISTR 8434)	TAP	20	15	10.8 ± 32 ($n = 4$)	0
<i>M. braunii</i> (TISTR 8429)	BG-11	10	15	0 ($n = 8$)	0
<i>A. falcatus</i> (TISTR 8557)	BG-11	10	15	0 ($n = 12$)	0
<i>T. cumbriacus</i> (TISTR 8480)	TAP	50	18	9 ± 8.8 ($n = 5$)	0
<i>A. densus</i> (TISTR 8505)	BG-11	10	15	0 ($n = 8$)	0

(Additional file 3). This result suggests the higher stability of transgenes in TISTR8447. This *S. acutus* TISTR8447 was used for subsequent studies.

Evaluation of TISTR8447 as a platform for gene expression

TISTR8447 is a fast-growing strain, reaching its stationary phase within 5 days in TAP medium. It accumulates high lipid contents up to 3 ± 0.17 and $10 \pm 1.04\%$ dry weight in N-supplemented and N-deprived media, respectively, based on total lipid quantification using vanillin staining. The lipid content in this strain is somewhat in moderate levels for lipid production as compared to other potential feedstock microalgae [31]. Nile red staining also showed the abundance of lipid droplets in the cells (Additional file 4). TISTR8447 was further tested for transformation efficiency using four *Agrobacterium* strains: A41, EHA105, GV3101 and LBA4404. Despite no statistical difference among the *Agrobacterium* strains ($p > 0.05$), EHA105 and GV3101 appeared to provide the highest transformation rates (Additional file 5). The effect of Acetosyringone was tested, but no significant difference among the tested concentrations (0, 50, 100 and $200 \mu\text{M}$) was observed ($p > 0.05$, data not shown), indicating that Acetosyringone is not required.

Transformed TISTR8447 were analyzed for the presence and expression of the transgene. Among 15 randomly selected potential-transformants subjected to either PCR or RT-PCR analysis, the *aphIV* gene (hygromycin B resistant gene) and its transcript was detected in 12 and 13 transformants, respectively (Additional file 6). This analysis indicates that most of the hygromycin B resistant colonies contained the transgene. However, after analyzing more than 60 transformants by confocal microscopy, little or no GFP signal was detected above the autofluorescence background (Additional file 6). RT-PCR of the *GFP* transcript and immunoblot analysis using anti-GFP antibody also

failed to detect the presence of the transcript and protein (data not shown). These data demonstrate that TISTR8447 is a potential microalgal strain for biofuel production through its properties for growth, lipid accumulation and its ability to be efficiently transformed by *Agrobacterium*, but has a very low capacity for transgene expression.

Transgene expression in TISTR8447 using *Chlamydomonas* introns

To test the effect of introns in transgene expression, four expression constructs with *Chlamydomonas* intron-carrying *mRuby2*, *mCerulean3*, *mVenus* and *Clover* reporters, which were previously shown to improve transgene expression in *Chlamydomonas* [22], were tested in the TISTR8447 through electro-transformation. Our electro-transformation rate for TISTR8447 was only 4.95 ± 5.18 CFU per 5×10^6 cells ($n = 20$), and no background colony was observed when using paromomycin selection. Noting that electro-transformation of *S. obliquus* was previously reported as 494 ± 48 CFU per 10^6 cells [25]. Fluorescence microplate analysis showed that most of the transformants had very low fluorescence signal ratios for the four reporters, similar to the levels observed in non-transformed controls (Fig. 1). Only a few showed 2–4-fold signal ratios indicating the expression of the reporter proteins above the background signals. The frequency of the expressers could be estimated 1–3 among 40 transformants (2.5–7.5%), which, nonetheless, is an improvement from the transformation using the pCXS-GFP. This result indicates that the incorporation of *Chlamydomonas* introns can partly improve transgene expression in TISTR8447.

Improving transgene expression in TISTR8447 using *ble::E2A* transcript fusion

Previously, transcript fusion constructs using a self-cleavage 2A peptide from foot-and-mouth disease virus (FMDV) [32] with *ble* selectable marker [33] were

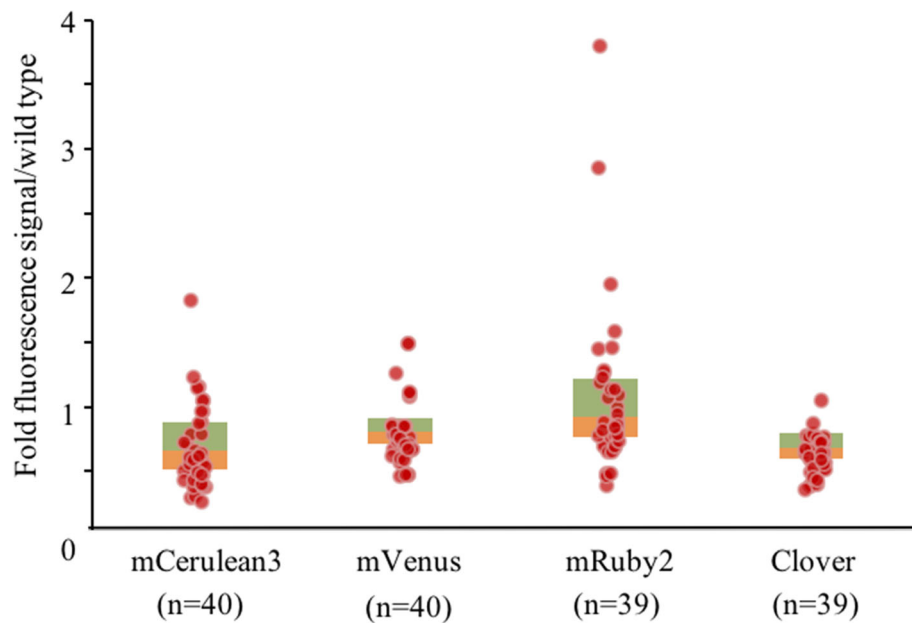


Fig. 1 Analysis of TISTR8447 transformants expressing four fluorescence reporter genes harboring *Chlamydomonas* introns using fluorescence microplate reader. Data are presented as box plots overlaid with scatter plots for fold fluorescence reads over the wild type values. n indicates the number of transformants reads for each reporter

shown to be a strategy of choices for highly efficient transgene expression in *Chlamydomonas* [23, 34–37]. To test whether this could improve transgene expression in TISTR8447, *ble::E2A* fusion vectors were constructed using with- or without-introns *ble* marker and *mCherry* reporter (pCreZ-*intble::E2A-mCherry* and pCreZ-*ble::E2A-mCherry*, Fig. 2 and Additional file 7). Transformants of these vectors were analyzed for the fluorescence reporter signal, and *mCherry* was readily observed under the confocal microscope (Fig. 2b). Fluorescence microplate analysis showed that most of the transformants from both constructs (19/24 or 79% for *ble* and 12/14 or 85% for *int-ble*) had more than a 2-fold increase of *mCherry* signal ratio to the wild type controls (Fig. 2c). The use of *RBSC2* promoter and introns in the *ble* gene (*int-ble*) showed higher signals compared to that using *PSAD* promoter and *ble* gene without introns (*ble*), though without significant difference by Kruskal-Wallis Test ($p > 0.05$). This result indicates that the *ble::E2A* transcript fusion system could provide an efficient transgene expression system for the TISTR8447 under selective pressure of zeocin.

Subsequently, the expression level of the *ble::E2A::mCherry* transcript fusion was tested whether it could be induced in a dose-dependent manner with elevated zeocin concentrations. Selected transformants with low, medium and high expression were cultured in at $10 \mu\text{g ml}^{-1}$ zeocin, before transferring into the media supplemented with 0– $50 \mu\text{g ml}^{-1}$ zeocin for 2 days. However, no significant increase of the fluorescence signals following zeocin

concentrations for both constructs was observed (Fig. 2d). At concentrations above $20 \mu\text{g ml}^{-1}$ zeocin, the transformants grew poorly with low cell mass and became chlorosis. This observation demonstrates that increases in zeocin selective pressure could not provide a further improvement on transgene expression in the TISTR8447.

Improving transgene expression in TISTR8447 by UV mutagenesis

To further improve the transgene expression, TISTR8447 mutants with enhanced transgene expression were generated using UV mutagenesis and a selection strategy using *CRY1-1* (conferring emetine resistant) as reported by Neupert et al. [21]. Transformation of the TISTR8447 using pCRE-*CRY1-1* was selected on the medium supplemented with $1 \mu\text{g ml}^{-1}$ emetine. Fifty selected transformants were tested on media supplemented with 1– $20 \mu\text{g ml}^{-1}$ emetine to confirm that they could grow on the medium supplemented with at most $1 \mu\text{g ml}^{-1}$ emetine. Three transformants (S14) were chosen for UV mutagenesis at 0.1% survival rate and subsequently selected using 2, 5 and $10 \mu\text{g ml}^{-1}$ emetine. Finally, three mutants (SUV1, 2 and 3), which were able to grow on $10 \mu\text{g ml}^{-1}$ emetine, were obtained (Fig. 3a). These mutants were then tested for any improvement on transgene expression by transformation using pOPT-*mCerulean3* followed by fluorescence microplate analysis (Fig. 3b). Although the highest levels of the fluorescence signal ratios were in the range of 2- to 3-fold, which were similar to transformants of control

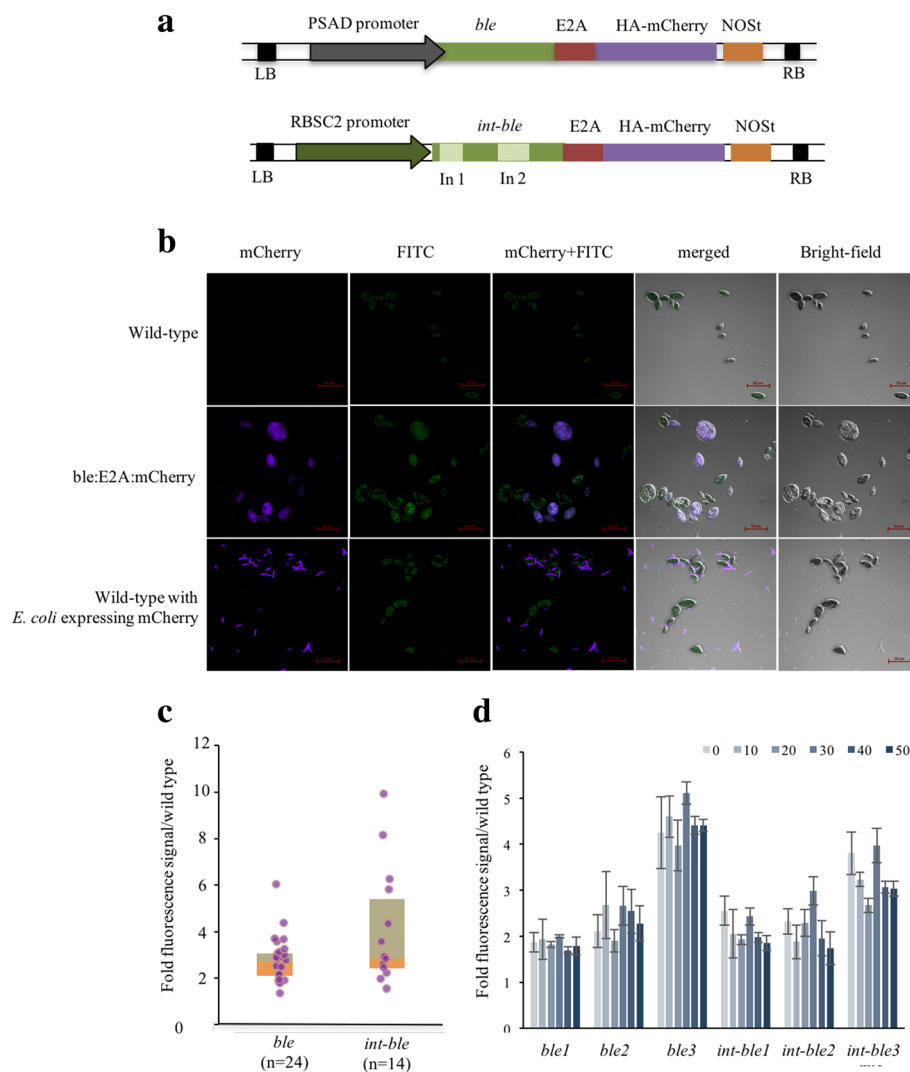


Fig. 2 Analysis of TISTR8447 transformants expressing transcript fusions. **a** pCreZ-*ble*::E2A::mCherry (*ble*) and pCreZ-*intble*::E2A::mCherry (*int-ble*) vectors are illustrated. **b** TISTR8447 transformants analyzed by confocal microscopy showing mCherry signal and FITC channel-autofluorescence (scale bar = 20 μ m). *E. coli* expressing mCherry was added to TISTR8447 wild type as a positive control. **c** mCherry signals of *ble* and *int-ble* vector transformants are presented as box plots overlaid with scatter plots. (n) indicates the number of transformants. **d** Fluorescence signals of low, medium and high mCherry expressors from each construct after subjected to media supplemented with 0–50 μ g ml⁻¹ zeocin for 2 days. Error bars represent standard deviations (n = 3)

strains (S14), the SUV1-3 provided overall higher frequencies of transformants with high fluorescence signal ratios (8/81 or 10%, 5/43 or 12% and 2/17 or 12% for SUV1-3, respectively). This result demonstrates that although the level of transgene expression is only moderately improved, these TISTR8447 mutants provide increases in the frequency of transgene expressed transformants.

Improving transgene expression by generating TISTR8447 white mutants and *PSY*::E2A transcript fusion

The transcript fusion system using *ble*::E2A have been shown so far to be the best improvement of transgene

expression in TISTR8447, but the use of zeocin as a selective pressure is impractical for large-scale culture and affects the algal growth. To further improve this, an alternative expression strategy was developed by using *Phytoene synthase* (*PSY*) mutants, *PSY* gene complementation and light as a selective pressure. *Chlamydomonas* with *PSY* mutations were previously shown to exhibit pale-green or whitish colony, perish under the light as low as 8 μ mol photons m⁻² s⁻¹ and to be complemented by transformation using *PSY* gene [38]. By adopting this idea, white mutants of the TISTR8447 were generated by UV mutagenesis and selection of pale-green colonies grown in the dark. From 50 selection plates, each with

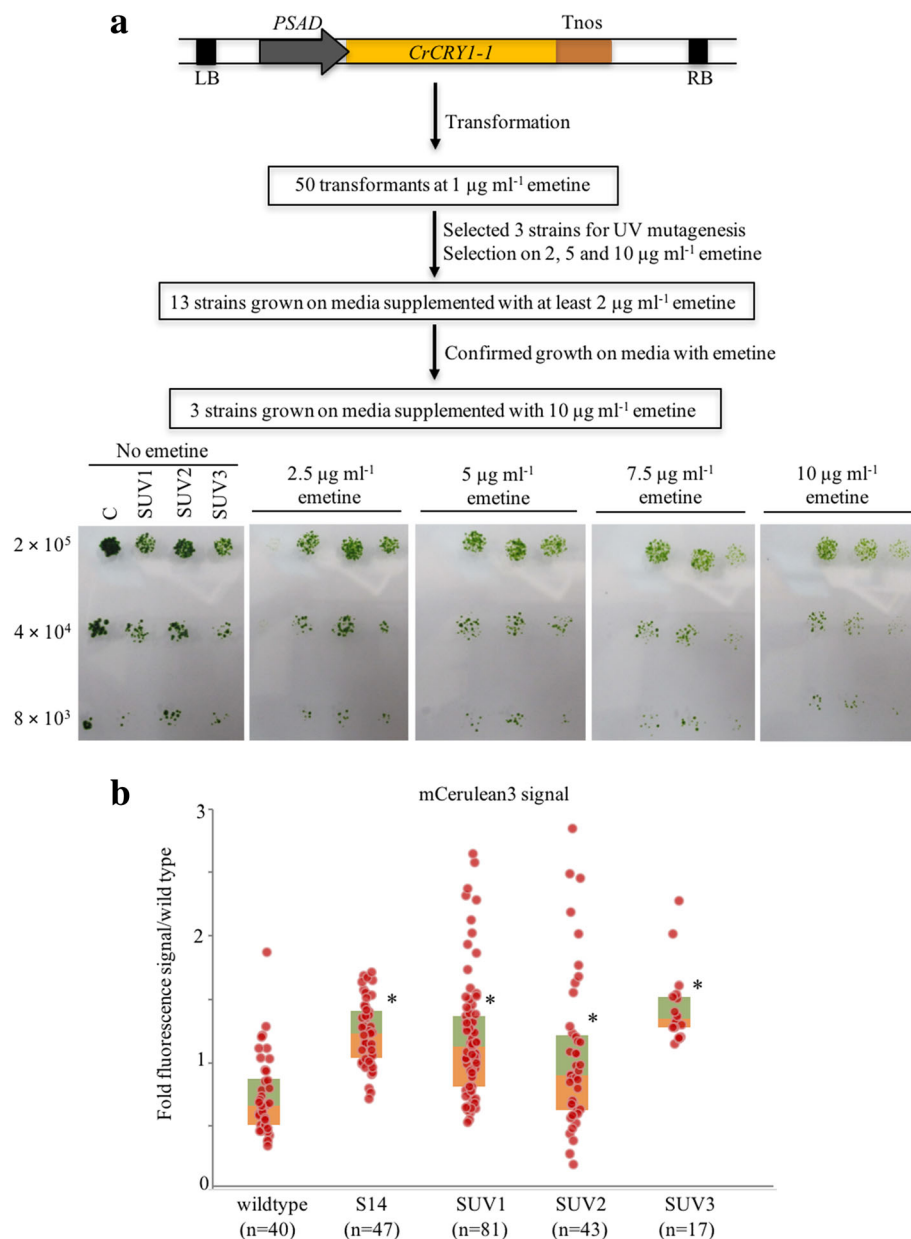


Fig. 3 UV mutagenesis and selection of SUV strains. **a** *CrCRY1-1* expression construct and the experimental procedure for selecting transformants. The lower panel shows three selected strains, SUV1-3, that grew on the medium supplemented with emetine up to 10 $\mu\text{g ml}^{-1}$ compared to the TISTR8447 wild type. **b** Box plots overlaid with scatter plots of mCerulean3 fluorescence signals of transformants generated from wild type, S14 (a *CrCRY1-1* expressing strain) and SUV1-3. (n) indicates the number of transformants for each strain. Asterisks indicate significant differences between signals from transformants generated from wild type and others by Kruskal-Wallis Test ($p < 0.05$)

10^6 cells, six white mutants, named G1-G6, were obtained after three rounds of subcultures in the dark (Fig. 4a). Four white mutants (G1, G2, G3 and G5) ceased growth under light at 50 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$, similar to the cc4113 and cc4109 *Chlamydomonas psy* white mutants [38], while G4 and G6 mutants were able to grow under the light. After sequencing the *PSY* gene in these mutants, point mutations were observed in the *PSY* coding sequence in G1, G2 and G5 mutants (Fig. 4b),

but not in G3. The G1, G2 and G5 *psy* mutants were then tested for complementation using a construct containing *Chlamydomonas PSY* coding sequence fused with *E2A* and *HA-mCherry*. Transformation of these three white mutants using the *Agrobacterium* method was unsuccessful, and this may be because the mutant grew very slow and could not out-compete *Agrobacterium* growth. Electro-transformation yielded complementation from G1 and G2 mutants with green colonies after

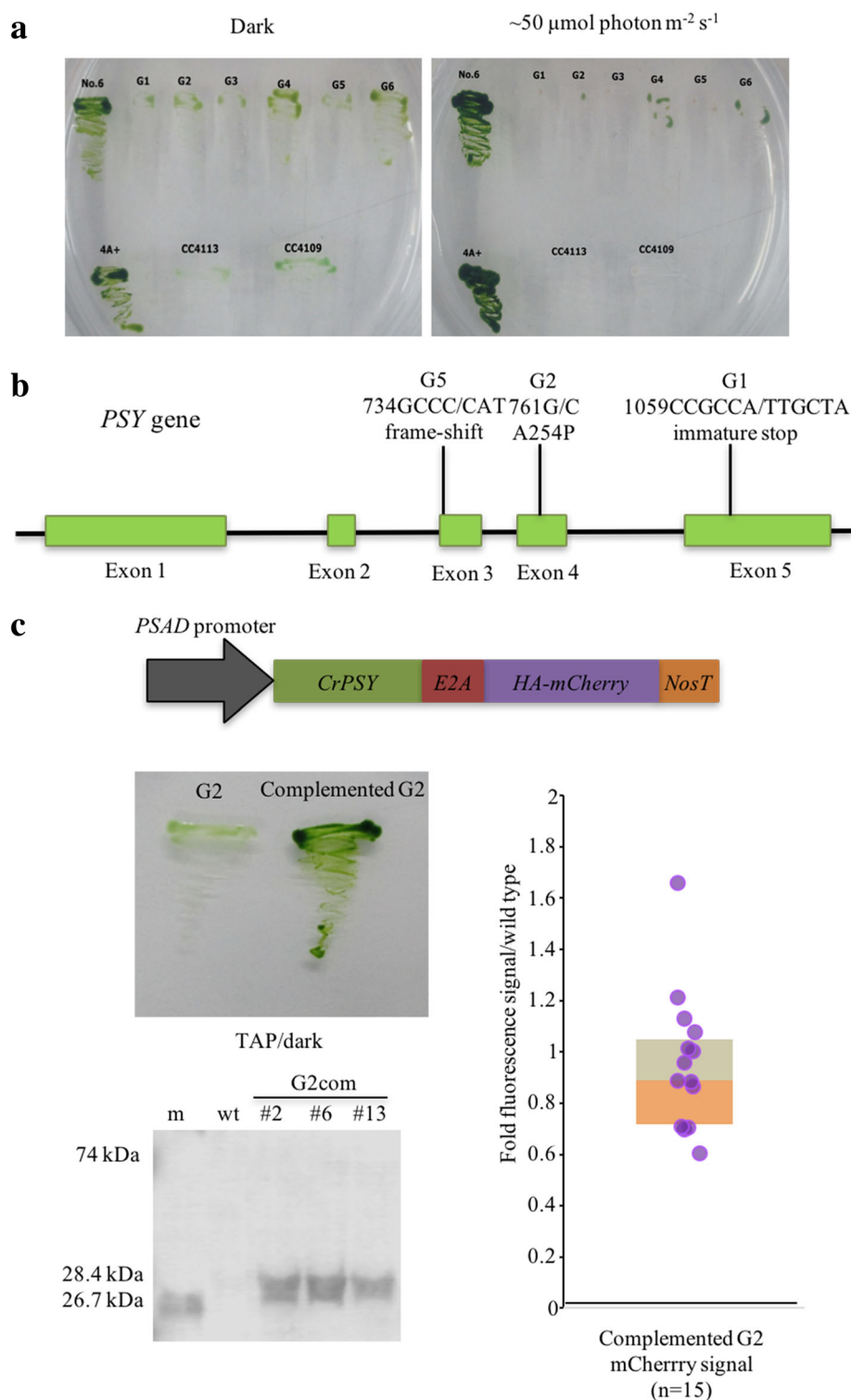


Fig. 4 Generation of *psy* white mutants and complementation using *CrPSY*. **a** Six isolated white mutants (G1-6) on TAP medium under light and dark conditions (No.6: TISTR8447, 4A+: *Chlamydomonas*, cc4113 and cc4109: *Chlamydomonas psy* white mutants). **b** Genotyping of *psy* mutants by direct PCR and sequencing. **c** Complementation of *psy* white mutants using *CrPSY* coding sequence fused in-frame with *E2A* and *mCherry*. *mCherry* was detected by both fluorescence microplate reader and immunoblot analysis (m: *mCherry* from *E. coli*; wt: TISTR8447). *mCherry* carrying HA tag in the algae is 28.4 kDa, while that in *E. coli* is 26.7 kDa. A faint band representing 74 kDa of *CrPSY*::*E2A*::*mCherry* fusion protein was observed (see Additional file 8 for longer exposures and Additional file 10 for the full length blot)

selection under light at $50 \mu\text{mol photons m}^{-2} \text{s}^{-1}$ for 7–10 days (Fig. 4c). However, the transformation rates were extremely low, approximately 1–4 CFU per 5×10^6 cells per plate.

The G2 complemented strains were analyzed by fluorescence microplate reader and showed that most of the transformants had low levels of the fluorescence signals for mCherry (Fig. 4c). The frequency of the transgene expressers (up to a 2-fold signal ratio) was one among 15 transformants (~7%), somewhat similar to that of intron-containing vectors. Unlike that of *ble::E2A* vectors with the zeocin selection, the expression of *PSY::E2A* was still low. This may be because the PSY functions as an enzyme in carotenoid biosynthesis pathways, while ble protein functions by binding to zeocin in a direct proportion [23]. Therefore, light selection could not drive the PSY expression to the levels achieved by zeocin selection. Spontaneous revertants were also monitored in the transformation controls (without plasmid) and, so far, there were only two revertants observed among 40 selection plates (5×10^6 cells per plate). To preclude the possibility of these complemented strains being derived from revertants, mCherry expression was confirmed by immunoblot analysis (Fig. 4c). The non-cleaved PSY::E2A::HA-mCherry fusion protein was observed at a very low level indicating that the E2A peptide was efficiently self-cleaved in the TISTR8447 (Additional file 8). This result demonstrates that *psy* white mutants of TISTR8447 can be used as host cells for transformation using PSY selectable marker and light selection. However, the expression levels of the *PSY::E2A* transgene were in the same range as that of constructs containing introns and far lesser than those using SUV mutants and *ble::E2A* with the zeocin selection, respectively.

Growth and total lipid contents of SUV mutants and complemented white mutants

To examine whether mutations for SUV and white mutants had any effect on growth or lipid production, the mutants were analyzed for growth curves and total lipid contents. While the growth of three complemented G2 strains and S14 was similar to that of the wild type, SUV1-3 grew slower (Fig. 5a). The cell growth of SUV1-3 was limited to approximately 10×10^6 cells ml^{-1} , whereas the wild type could grow up to 15×10^6 cells ml^{-1} . The lipid contents of S14 and G2 complemented strains in both N-supplemented and N-deprived media were similar to those of wild type, but those of SUV1-3 were significantly lower for both media ($p < 0.05$) (Fig. 5b). This result indicates that mutations, which provide improved transgene expression in SUV mutants, also affect physiological processes of the TISTR8447 resulting in slow growth and less accumulated lipids.

Discussion

This work shows that *Agrobacterium*-mediated transformation is a convenient method for simultaneously testing the transformation of many microalgae. In this work, ten strains were tested, and six of them were successfully transformed. The best transformation rate was obtained from *S. acutus* TISTR8447 at 217.5 ± 75 CFU per 10^6 cells, and this was selected as our best candidate for evaluating the transgene expression. *Coelastrum*, *Chlorococcum*, *Monoraphidium* and *Tetradasmus* are reported for their first trial here for the *Agrobacterium* method. Noting that although *M. braunii*, *A. falcatus*, *A. densus* and *Coelastrum* spp. were unable to be transformed in this work, this should not preclude their possibility to be transformed by other means. *Agrobacterium*-mediated transformation of microalgae was first introduced by Kumar et al. [30] for a successful transformation in *Chlamydomonas*, and various protocols have been developed since then to improve its efficiency [9, 39, 40]. It was then applied to many other microalgal species including *Haematococcus pluvialis* with the transformation rate at 153 ± 4 CFU per 10^6 cells [41], *Dunaliella bardawil* at 39–42 CFU per 10^6 cells [42], *Schizochytrium* at 50–170 CFU per 10^6 cells [43], *Scenedesmus almeriensis* at 90 CFU per 10^6 cells [26], *Dunaliella salina* at 40 ± 5 CFU per 10^6 cells [29] and *Tetraselmis chuii* at 150 ± 90 CFU per 10^6 cells [27]. And, with a different unit of transformation rate, *Nannochloropsis* sp. [44], *Chlorella vulgaris* [45] and *Ankistrodesmus* sp. [46] were shown to be transformed based on GUS expression at approximately 21, 30 and 3.5% efficiency, respectively. Taken together, this work provides support for the use of the *Agrobacterium*-mediated method for transforming other potential industrial microalgae yet to be tested.

Extremely inefficiently transgene expression is commonly observed in microalgae [9], and this appears to be the case for *Scenedesmus* as indicated by this and previous works. For example, transformation of *S. almeriensis* yielded little or very low GUS reporter expression even after screening many transformants [26], that of *S. bajacalifornicus* had only 2.96% of transformants with detectable GUS staining [46] and that of *S. obliquus* also showed only 1.5% of the transformants with GFP expression [25]. In our case, no GFP expression was observed at neither transcript nor protein level. Our attempt to improve the transgene expression in TISTR8447 by using tools based on *Chlamydomonas*, so far, shows that the transcript fusion using *ble::E2A* provided the highest improvement with up to 10-fold increases of mCherry expression and up to 85% of transformants with observable mCherry. However, other strategies including the uses of intron-containing constructs, SUV1-3 mutants and white mutants with *PSY::E2A* complementation provided only minor improvements with at most 2–4-fold increases of the reporter expression. The majority of the

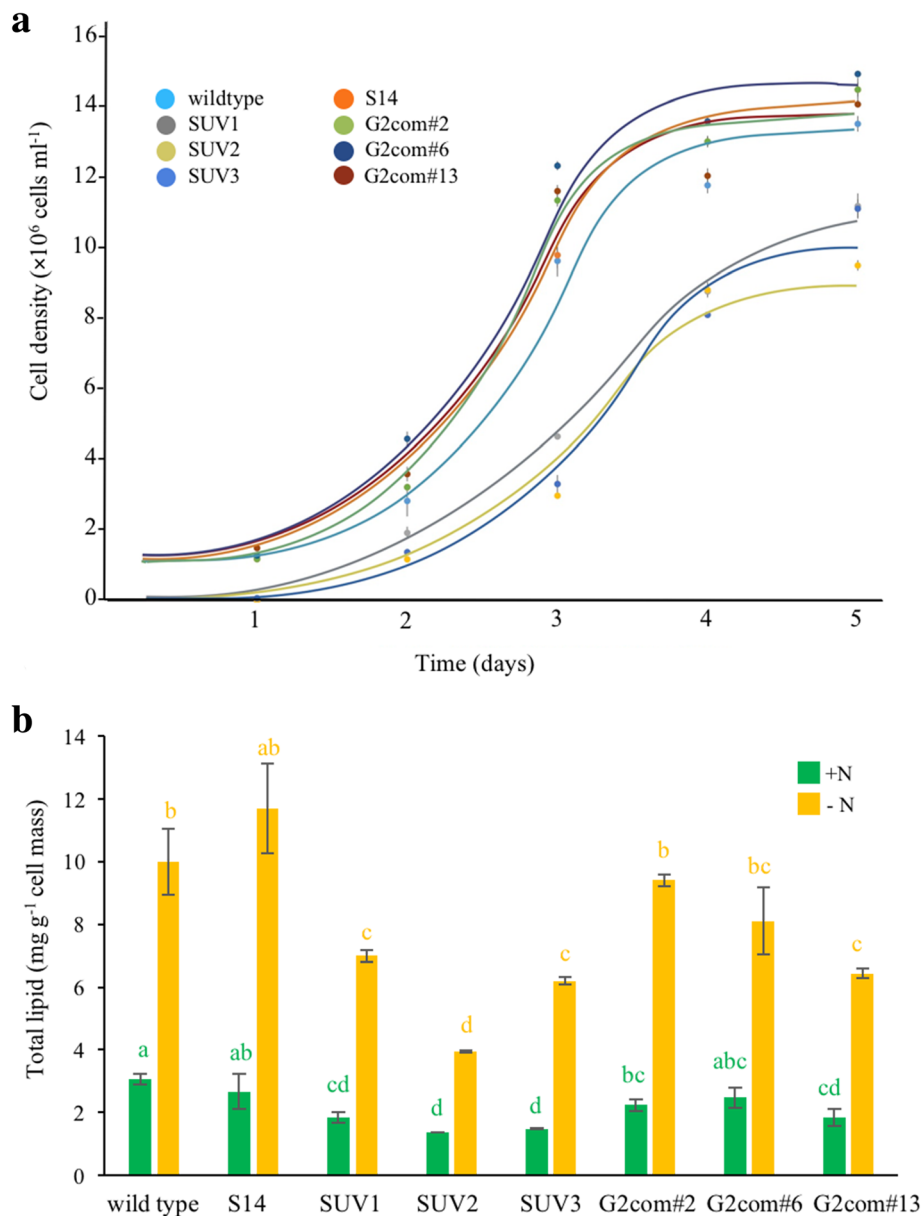


Fig. 5 a Growth curves of S14, SUV1-3 and complemented G2 strains compared to TISTR8447 wild type. **b** Total lipid content of the microalgae in N-supplemented and N-deprived media grown for 3 days. Error bars represent standard deviation ($n = 6$), and bars with the same letter are not significantly different (Tukey's test, $p < 0.05$)

transformants from the intron-containing constructs and the white mutants had very low transgene expression, while SUV mutants had increases in the frequency of transgene expressors.

Transcriptional gene silencing was proposed to be the key factor that causes the low transgene expression in *Chlamydomonas* [13, 19, 21, 36, 47] and most likely in many other microalgae [48–50]. Noting that a recent work by Mini et al. [9] reported that rearrangements of transgenes could also be another factor contributing to the inefficient expression in *Chlamydomonas*. Because gene

expression tools that could circumvent the gene silencing effects were successfully developed in *Chlamydomonas*, these tools could be applied to other microalgae, particularly in the TISTR8447. However, the improvements were far less than those obtained in *Chlamydomonas*. The development of the *ble::2A* systems provided more than a 100-fold improvement of transgene expression in *Chlamydomonas* [23, 34, 35, 37], while it was only a 10-fold increase in the TISTR8447. Insertions of an endogenous intron(s) into transgenes or the surrounding 5' and 3' untranslated regions led to improved transgene expression in

Chlamydomonas [51–53]. In contrast, these *Chlamydomonas* introns only provided small increases in the transgene expression in the TISTR8447. A little improvement in the TISTR8447 observed here by the use of the introns could be because of the difference in intron splicing between the species, or these introns could only occasionally help the transgene escaping the silencing controls in the TISTR8447. Perhaps, *Scenedesmus* might have a more robust gene silencing mechanism than that in *Chlamydomonas*.

Neupert et al. [21] first reported the strategic isolation of *Chlamydomonas* mutants with improved transgene expression using *CRY1-1* and elevated emetine concentrations for selection. The selected mutants, UVM4 and UVM11, had demonstrated up to 100-fold increases in emetine resistance. Less than expected, our selections of mutagenized TISTR8447 yielded up to a 10-fold increase in emetine resistance, and the expression levels based on the fluorescence ratios were improved only up to a 2-fold. Nonetheless, the SUV1-3 had higher frequencies of transgene expressors, in which transformants with at least a 2-fold signal could be accounted for ~10–12% compared to ~2.5% observed in wild type. This improvement is somewhat comparable to UVM4 and UVM11 that provided 4.5-fold increases in the frequency of transgene expressors compared to the wild type [47].

MET1 (DNA methyltransferase1) is one of the key enzymes for the DNA methylation transgene silencing along with other DNA-methylation independent gene silencing enzymes [13]. *Chlamydomonas met1* insertional mutants were shown to provide an improvement for transgene expression. Work from Kurniasih et al. [13], which performed UV mutagenesis using the *met1* mutant and selection using Neupert et al. [21] selection method, yielded mutants with even higher transgene expression. Because only moderate improvement was achieved in the SUV mutants, it is likely that there are still other transcriptional silencing genes present in the TISTR8447. Further mutations in the transcriptional silencing related genes may be required for further improvement.

Mutations that provide improved transgene expression in *Chlamydomonas* are known to be associated with epigenetic controls, and these may affect cell growth, as demonstrated in *Chlamydomonas* UVM mutants of the *met-1* insertional mutant [13]. Indeed, mutations of SUV1-3 are most likely related to epigenetic control machinery, and they affect both cell growth and lipid production, as none of these effects was observed in the S14 control strain. This result shows that while the SUV1-3 mutants have gained some improvement on the transgene expression, their growth and ability to accumulate lipids were weakened, compromising cell mass and lipid production. This result confirms the previous observation of the negative aspect of using mutants related to epigenetic controls.

Thus, cautions should be made before performing genetic screening for enhancing transgene expression, especially when the cell performances are the key to the application. Nonetheless, the white mutants, when complemented, could provide sufficiently intact growth and have only slight decreases in the lipid content. The result also suggests that while *CrPSY* can sufficiently complement the *SaPSY* knockout, yielding complemented strains with robust cell growth, it could not provide the lipid accumulation as efficient as the *SaPSY*.

Even though the *PSY::E2A* construct is not as effective as *ble::E2A* in forcing expression, this system provides an alternative selectable marker for microalgae. Current selectable markers used for microalgae transformation are based on antibiotic selection, for example, bleomycin/zeocin resistance gene [33] and paromomycin resistance gene [54]. Recently, the herbicide norflurazon was recently developed for transformation selection using *Phytoene desaturase* (*PDS*) with an amino substitution L505F in some microalgae [55–57]. The use of *PSY* gene as a selectable marker for microalgae, so far, has not been reported. This may be because of two main reasons: white mutants needed to be generated by mutagenesis and screened for each microalga, and white mutants are weak with slow growth and difficult to transform. For the work presented here, this is one of the most economical transgene selection approaches as only a light source is required. This selectable marker is compatible with an outdoor and large-scale culture as transgene constructs can be maintained by sunlight.

To this end, among the *Chlamydomonas* gene expression tools tested, the transcript fusion using *ble::E2A* appeared to be the most efficient and could instantaneously mitigate the inefficient transgene expression in microalgae, in particular, the TISTR8447. However, the exposure of zeocin antibiotic to the algal cells generally results in cell toxicity and affects cell growth and potentially the production of products and yields [58]. Furthermore, the use of zeocin is uneconomical for large-scale production, limiting the use of *ble::E2A* system. A consensus for inefficient transgene expression has been evidenced in many oleaginous microalgae and *Scenedesmus* species as reported here. The challenges in gaining high transgene expression in microalgae are the critical bottleneck for current development in Eukaryotic microalgal biotechnology. Further improvements and newly developed tools are critically needed to allow efficient use of microalgae for biotechnology application.

Conclusions

This work demonstrates that oleaginous microalgae isolated from natural resources can be screened for selecting strains with high transformation efficiency, but they

generally have a low capacity for transgene expression. A selected strain, *S. acutus* TISTR8447, was tested to improve the transgene expression using strategies developed in *Chlamydomonas*, but overall improvement was of moderate for both levels and frequency of expression. So far, the most effective tool giving the highest improvement is the transcript fusion using *ble::E2A* system. Furthermore, a new strategy for transgene expression has been developed using *psy* white mutants with an expression vector containing *PSY::E2A* for complementation and light selection. Finally, this work demonstrates that genetic engineering of non-model microalgae is still a challenging task. New tools and strategies are critically needed for transgene expression in promising industrial microalgae.

Methods

Microalgae and cultivation conditions

Microalgae were obtained from Thailand Institute of Science and Technological Research (TISTR) including *Coelastrum* sp. (TISTR8511), *Scenedesmus acutus* (TISTR8555), *S. acutus* (TISTR8540), *S. acuminatus* (TISTR8519), *S. acutus* (TISTR8447), *Chlorococcum humicola* (TISTR8434), *Monoraphidium braunii* (TISTR8429), *Ankistrodesmus falcatus* (TISTR8557), *Tetradesmus cumbricus* (TISTR8480), *Ankistrodesmus densus* (TISTR8505). These accessions were chosen based on their potential high accumulation of lipid droplets under the N deprivation condition and the variation of algal species. *Chlamydomonas* white mutants CC4113 (Its1-207 mt+) and CC4109 (Its1-203 mt+) were obtained from Chlamydomonas Resource Center. Microalgae were cultured in either Tris-acetate phosphate (TAP) or BG-11 media at 25 °C under continuous light (50 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$), excepting that those white mutants were incubated in the dark. The N-deprived medium (TAP-N) was prepared by substituting KNO_3 with KCl. For growth analysis, three cultures of 50 ml were started at the of 10^6 cells ml^{-1} cell density, and 1 ml sample was collected daily to monitor for cell density.

Plasmid constructions

pOPT vector series with reporter genes including *mRuby2*, *mCerulean3*, *mVenus* and *Clover* [22] were obtained from Chlamydomonas Resource Center. pCXSN-GFP was constructed by TA cloning using the *Agrobacterium* transformation vector pCXSN [59] and *GFP* coding sequence, which was PCR amplified using GFP-F and GFP-R primers (Additional file 9). Transcript fusion vectors were constructed using pPLV02 [60] as the vector backbone. The pPLV02 backbone was amplified using pplv-F and pplv-R primers. PSAD promoter, RBSC2 promoter-*ble* with introns and *ble* without intron fragments were amplified from pSL18 [61, 62] using psad-F and psad-R primers, pSP124S [51] using rbcS-F and Zeo-E2A-R primers and pICZ (Thermo Fisher)

using Zeo-F and Zeo-E2A-R primers, respectively. Each primer has an adaptor sequence for subsequent sequence and ligation independent cloning (SLIC) [63]. All PCR cloning was performed using hi-fidelity Q5 DNA polymerase (New England Biolabs, USA). pCreZ-E2A and pCreZint-E2A were constructed by fusing the pPLV02 PCR fragment with PSAD promoter and *ble* without intron, and RBSC2 promoter-*ble* with introns, respectively, by SLIC technique. *HA-mCherry* was subsequently inserted into both vectors at NcoI site by PCR amplification using HA-mcherry-F and mcherry-R primers and SLIC.

For *CrCRY1-1* expression construct, the coding region of *CrCRY1* was PCR amplified from *Chlamydomonas* cDNA using gene specific primers, before being re-amplified using a reverse primer with a modified base at the last codon from CTG to CCG (cry1-1-F and cry1-1-TtoC-R primers) to generate *CrCRY1-1* coding sequence. The *CrCRY1-1* fragment with adaptor sequences was then inserted into pCreZ-E2A at EcoRI and BamHI sites by SLIC, removing *ble-E2A*. For *CrPSY::E2A::mCherry* constructs, the *CrPSY* coding sequence was PCR amplified from *Chlamydomonas* cDNA and re-amplified using crpsy-F and crpsy-R primers, which contain adaptor sequences for SLIC. The *CrPSY* fragment was then ligated to a vector fragment, which was amplified from pCreZ-E2A by E2A-F and psad-crpsy-R primers, replacing *ble* coding sequence and in frame with *E2A*. Plasmids for *Agrobacterium*-mediated transformation were transferred into *Agrobacterium* cells by electroporation.

Transformation

The microalgal strains were tested for their sensitivity to hygromycin B before performing an *Agrobacterium*-mediated transformation. Briefly, 10^6 cells were spread onto TAP or BG-11 agar media supplemented with 10–100 $\mu\text{g ml}^{-1}$ of hygromycin B for 14 days. This sensitivity test was performed in triplicate. For *Agro*-transformation, microalgae were cultured in liquid media for 3–4 days, and 10^6 cells were grown on solid media for 2–3 days until forming a lawn. *A. tumefaciens* EHA105 harboring transformation plasmids were grown in LB media supplemented with 25 $\mu\text{g ml}^{-1}$ of rifampicin and 50 $\mu\text{g ml}^{-1}$ of kanamycin under continuous shaking at 28 °C for 2 days. For evaluation of *Agrobacterium* strains, A41, GV3101 and LBA4404 strains were used in replacement of EHA105. The algal lawn was mixed with 100 μl of *Agrobacterium* ($\text{OD}_{600} = 1$) using a spreader and co-cultivated in the dark for 2 days before transferring onto solid TAP or BG-11 selective media supplemented with 250 $\mu\text{g ml}^{-1}$ cefotaxime and hygromycin B at specific concentrations. Colonies obtained from selection plates were subcultured twice on the selection media

before being designated as transformants. Transformation rates were calculated based on the numbers of resistant colonies per 10^6 starting cells. *Agrobacterium* cells without expression plasmid were used for transformation as negative controls.

For electro-transformation, log phase cells at no more than 2×10^6 cells ml^{-1} were harvested by brief centrifugation at $400 \times g$ and resuspended in TAP supplemented with 40 mM sucrose (TAP sucrose) at 2×10^8 cells ml^{-1} . Thirty-eight microlitres of the algal suspension were mixed with 2 μl of a plasmid vector (400 ng) before transferring to a 2 mm gap electroporation cuvette and pulsing using Bio-Rad Gene Pulser II with the following condition: 0.4 kV, 25 μF and 500 Ω . The algal cells were immediately transferred to 10 ml TAP sucrose and cultured at the normal condition with constant shaking at 100 rpm under low light for an overnight before spreading onto TAP agar supplemented with 30 $\mu\text{g ml}^{-1}$ of paromomycin or 10 $\mu\text{g ml}^{-1}$ of zeocin for 7–10 days.

For the transformation of white mutants, a white colony was cultured in TAP sucrose with constant shaking in the dark up to 10 days or until reaching $1\text{--}2 \times 10^6$ cells ml^{-1} . Cells were harvested and resuspended in 40 mM sucrose TAP at 2×10^8 cells ml^{-1} . Subsequently, 38 μl of the suspension was mixed with 2 μl of a plasmid before being electroporated using the same condition as above and cultured in 10 ml TAP sucrose for overnight with constant shaking in the dark. The cells were plated onto TAP agar and incubated in the dark for 5 days and then under light (50 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$) for 7–10 days.

UV mutagenesis and mutant screening

UV mutagenesis was performed by irradiating the TISTR8447 at a density of 5×10^6 cells ml^{-1} , which were previously kept in the dark for 2 h, in a petri dish using a UV transilluminator for 2 h to obtain a 0.1% survival rate. Cells were cultured overnight with shaking in the dark before plating onto TAP agar medium at 10^6 cells per plate. For SUV mutants, *CRY1-1* transformants were UV irradiated and plated onto TAP agar medium supplemented with emetine at 2, 5 and 10 $\mu\text{g ml}^{-1}$. For white mutants, irradiated cells were plated onto TAP agar medium and incubated in the dark for 14 days. Colonies with pale-green were subcultured in the dark.

Molecular characterization

For DNA and RNA isolation, the algal cells were subjected to a mechanical disruption in liquid nitrogen using a mortar and pestle. DNA was isolated using the CTAB method, whereas total RNA was isolated using GENEzol TriRNA Pure Kit (Geneaid, Taiwan). RNA was purified by using RNase-free DNase and RNA Cleanup Kit. cDNA was synthesized using one microgram of total

RNA by MMuLV reverse transcriptase. PCR and RT-PCR reactions were performed using *Taq* DNA polymerase and primers listed in Additional file 9. For sequencing of the *PSY* coding sequence of TISTR8447 (Genbank accession MF401544), *SaPSY* of the white mutants was amplified from total DNA using hi-fidelity Q5 DNA polymerase and sapsy-seq-F and sapsy-seq-R primers and sequenced.

For western analysis, a total protein was isolated by grinding the cells with liquid nitrogen and resuspended in a protein extraction buffer [20 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 10% (v/v) glycerol, 0.5% IGEPAL CA-360 and 1 mM PMSF]. The total protein was quantified by Bradford assay before subjecting to SDS-PAGE and blotting onto a PVDF membrane. mCherry was detected using rabbit anti-mCherry as a primary antibody at 1:1000 dilution and anti-rabbit-HRP as a second antibody at 1:10,000 dilution. The western signal was detected using WesternSure Premium Chemiluminescence Substrate (Li-COR, USA) and C-DiGit Blot Scanner (Li-COR, USA).

Fluorescence detection

Fluorescent reporter signals were analyzed using a microplate reader. The excitation and emission wavelengths are as follows; mRuby2 (ex = 558 nm, em = 605 nm), mCerulean3 (ex = 445 nm, em = 503 nm), mVenus (ex = 515 nm, em = 550 nm), Clover (ex = 477 nm, em = 515 nm), mCherry (ex = 562 nm, em = 607 nm). Chlorophyll autofluorescence was measured at 440 nm for excitation and 685 nm for emission. Standard curves for the conversion of fluorescence signal to protein contents were generated from recombinant GST-mCerulean and GST-mCherry purified from *E. coli* and quantified using Bradford assay. The protein concentrations used was 1.25–400 $\text{ng } \mu\text{l}^{-1}$ in the TISTR8447 algal culture at 10^6 cells ml^{-1} . The chlorophyll fluorescence was used for calculating the cell number based on a standard curve between the chlorophyll signals and cell numbers. Selected transformants with a high level of expression of transgenes were then assessed for fluorescence signal using Nikon C2 Si confocal microscopy.

Stability test

Transformants were maintained on media without hygromycin B for 2 weeks before subculturing onto media with and without hygromycin B to determine the retention of the selectable marker gene without the selection pressure. For the G2 complemented lines, the dark and light conditions were used instead of the antibiotic. These subcultures were repeated every 2 weeks.

Lipid quantification using vanillin method

This lipid quantification using vanillin was modified from Mishra et al. [64]. Cells were grown in a starting culture using liquid TAP medium for 2–3 days ($\sim 2 \times 10^6$ cells ml^{-1}

concentration) before dividing in half, and the medium of each half was replaced by liquid TAP or TAP-N at 2×10^6 cells ml^{-1} concentration before culturing for 3 days. Approximately 1.5×10^6 cells with known mass was mixed with 200 μl sulfuric acid, incubated at 95°C for 10 min before placing on ice for 5 min, added with 500 μl phospho-vanillin [0.12% (w/v) vanillin in 80% (v/v) phosphoric acid] and incubated at 37°C for 15 min. The solution was briefly centrifuged at 3000 rpm for 5 min, and the absorbance at 530 nm of the sample was assessed. Total lipid was quantified against the standard curve generated from canola oil.

Nile red staining

Microalgae were stained using Nile red ($10 \mu\text{g ml}^{-1}$ in DMSO) and Calcofluor (1%) for 30 min in the dark at 25°C . The samples were then fixed in 4% paraformaldehyde before washing twice with PBS solution and analyzed using Nikon C2 Si confocal microscopy (543 nm excitation, 525–605 nm emission).

Statistical analysis

The statistical tests used for analyzing the data varied based on the distribution of each data set. Kruskal-Wallis test was conducted to assess the difference in the transformation efficiency of different *Agrobacterium* strains and the levels of fluorescence signals among populations of microalgal strains. For datasets with a normal distribution, one-way analysis of variance (ANOVA) using an alpha of 0.05 and a multiple range test using Tukey's method was conducted.

Additional files

Additional file 1: Survival of ten microalgal strains on different concentrations of hygromycin B. (PDF 37 kb)

Additional file 2: Representative cultured and selective plates for *Agrobacterium*-mediated transformation of TISTR8540. (PDF 552 kb)

Additional file 3: A transgene stability test of TISTR8540 and TISTR8447. (PDF 1536 kb)

Additional file 4: Nile red staining of the TISTR8447 after 3 days of nitrogen starvation. (PDF 101 kb)

Additional file 5: Transformation rates of the TISTR8447 using four *Agrobacterium* strains. (PDF 39 kb)

Additional file 6: *Agrobacterium*-mediated transformation of *S. acutus* TISTR8447. (PDF 446 kb)

Additional file 7: The schematic of pCreZ vectors. (PDF 64 kb)

Additional file 8: Western analysis of CrPSY::E2A::mCherry fusion protein with extended exposure times. (PDF 127 kb)

Additional file 9: Oligonucleotide primer sequences. (PDF 111 kb)

Additional file 10: The full length immuno blot of G2 complemented strains. (PDF 170 kb)

Abbreviations

ANOVA: Analysis of variance; aphIV: Hygromycin phosphotransferase; BG11: Blue-green Medium; ble: Bleomycin resistant protein; CaMV: Cauliflower mosaic virus; CFU: Colony forming unit; CrPSY: *Chlamydomonas reinhardtii* phytoene synthase; CRY1-1: Cryptochrome

circadian clock 1-1; CTAB: Cetyltrimethylammonium bromide; DMSO: Dimethyl sulfoxide; EDTA: Ethylenediaminetetraacetic acid; GFP: Green fluorescence protein; GST: Glutathione S-transferase; GUS: Beta-glucuronidase; HA: Hemagglutinin; HRP: Horseradish peroxidase; MET1: DNA methyltransferase1; MMuLV: Reverse transcriptase: moloney murine leukemia virus reverse transcriptase; PDS: Phytoene desaturase; PMSF: Phenylmethylsulfonyl fluoride; PSY: Phytoene synthase; PVDF: Polyvinylidene fluoride; SaPSY: *Scenedesmus acutus* phytoene synthase; SDS-PAGE: Sodium dodecyl sulfate polyacrylamide gel electrophoresis; TAP: Tris-acetate phosphate buffer; TAP-N: Tris-acetate phosphate buffer without nitrogen; Taq polymerase: *Thermus aquaticus* polymerase; UV: Ultraviolet

Acknowledgments

We are grateful to the Thailand Institute of Scientific and Technological Research (TISTR) for providing the microalgae. We thank Pornthip Boonmahamongkol and Science Equipment center, Faculty of Science, Kasetsart University for equipment and technical support, particularly fluorescence microplate reader and confocal microscopy. We thank Dr. Thidaporn Supapakorn for support on statistical analysis. We thank Dr. Kunn Kangwansaichol, Chanakarn Chaikaedang, Jureeporn Konpiewlert and Chanarong Sartsanga for technical support.

Funding

This work was supported by Kasetsart University Research and Development Institute (KURDI), PPT Public Company Limited and Thailand Research Fund (TRF-RSA6080031). These funding bodies had no role in the design of this study or any role during the collection, analysis, and interpretation of the data, and in writing the manuscript.

Availability of data and materials

All data generated or analyzed during this study are included in this published article (and its Additional files) and are available from the corresponding author on reasonable request. Microalgae are stocked at the Thailand Institute of Science and Technological Research (TISTR). The PSY coding sequence of TISTR8447 is available at Genbank (accession MF401544).

Authors' contributions

ASu and SV conceived and designed the experiments. ASu, WP, TK, SSr, Ssa and WD performed experiments. ASu, ASi, PJ, KJ and SV analyzed the data. ASu and SV prepared the Figures and Tables and wrote the manuscript. ASi, PJ and SV corrected and proofread the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

This research was partly supported by PPT Public Company Limited (Thailand). Kanidtha Jariyachawalid is employed by the PTT public company limited. Patent applications (Thailand patent No.1501003942 and No.1601007906) have been filed for some of the technology disclosed in this publication. All other authors declare no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Special Research Unit in Microalgal Molecular Genetics and Functional Genomics (MMGFG), Department of Genetics, Faculty of Science, Kasetsart University, 50 Ngam Wong Wan road, Chatuchak, Bangkok 10900, Thailand.

²Center of Advanced studies for Tropical Natural Resources, Kasetsart University, 50 Ngam Wong Wan road, Chatuchak, Bangkok 10900, Thailand.

³PTT Research and Technology Institute, PTT Public Company Limited, Ayuthaya 13170, Thailand.

Received: 14 June 2018 Accepted: 25 December 2018

Published online: 10 January 2019

References

- Chisti Y. Biodiesel from microalgae beats bioethanol. *Trends Biotechnol.* 2008;26(3):126–31.
- Beer LL, Boyd ES, Peters JW, Posewitz MC. Engineering algae for biohydrogen and biofuel production. *Curr Opin Biotechnol.* 2009;20(3):264–71.
- Jones CS, Mayfield SP. Algae biofuels: versatility for the future of bioenergy. *Curr Opin Biotechnol.* 2012;23(3):346–51.
- Singh A, Olsen SI. A critical review of biochemical conversion, sustainability and life cycle assessment of algal biofuels. *Appl Energy.* 2011;88(10):3548–55.
- Wijffels RH, Barbosa MJ. An outlook on microalgal biofuels. *Science.* 2010;329(5993):796–9.
- Radakovits R, Jinkerson RE, Darzins A, Posewitz MC. Genetic engineering of algae for enhanced biofuel production. *Eukaryot Cell.* 2010;9(4):486–501.
- Georgianna DR, Mayfield SP. Exploiting diversity and synthetic biology for the production of algal biofuels. *Nature.* 2012;488(7411):329.
- Potvin G, Zhang Z. Strategies for high-level recombinant protein expression in transgenic microalgae: a review. *Biotechnol Adv.* 2010;28(6):910–8.
- Mini P, Demurtas OC, Valentini S, Pallara P, Aprea G, Ferrante P, Giuliano G. *Agrobacterium*-mediated and electroporation-mediated transformation of *Chlamydomonas reinhardtii*: a comparative study. *BMC Biotechnol.* 2018;18(1):11.
- Zhang R, Patena W, Armbruster U, Gang SS, Blum SR, Jonikas MC. High-throughput genotyping of green algal mutants reveals random distribution of mutagenic insertion sites and endonucleolytic cleavage of transforming DNA. *Plant Cell.* 2014;26(4):1398–409.
- Radakovits R, Jinkerson RE, Fuerstenberg SI, Tae H, Settlege RE, Boore JL, Posewitz MC. Draft genome sequence and genetic transformation of the oleaginous alga *Nannochloropsis gaditana*. *Nat Commun.* 2012;3:686.
- Cerutti H, Johnson AM, Gillham NW, Boynton JE. Epigenetic silencing of a foreign gene in nuclear transformants of *Chlamydomonas*. *Plant Cell.* 1997;9(6):925–45.
- Kurniasih SD, Yamasaki T, Kong F, Okada S, Widyaningrum D, Ohama T. UV-mediated *Chlamydomonas* mutants with enhanced nuclear transgene expression by disruption of DNA methylation-dependent and independent silencing systems. *Plant Mol Biol.* 2016;92(6):629–41.
- Sodeinde OA, Kindle KL. Homologous recombination in the nuclear genome of *Chlamydomonas reinhardtii*. *Proc Natl Acad Sci U S A.* 1993;90(19):9199–203.
- Gumpel NJ, Rochaix JD, Purton S. Studies on homologous recombination in the green alga *Chlamydomonas reinhardtii*. *Curr Genet.* 1994;26(5):438–42.
- Jiang W, Brueggeman AJ, Horken KM, Plucinak TM, Weeks DP. Successful transient expression of Cas9 and single guide RNA genes in *Chlamydomonas reinhardtii*. *Eukaryot Cell.* 2014;13(11):1465–9.
- Kilian O, Benemann CS, Niyogi KK, Vick B. High-efficiency homologous recombination in the oil-producing alga *Nannochloropsis* sp. *Proc Natl Acad Sci U S A.* 2011;108(52):21265–9.
- Wang Q, Lu Y, Xin Y, Wei L, Huang S, Xu J. Genome editing of model oleaginous microalgae *Nannochloropsis* spp. by CRISPR/Cas9. *Plant J.* 2016;88(6):1071–81.
- Fuhrmann M, Oertel W, Hegemann P. A synthetic gene coding for the green fluorescent protein (GFP) is a versatile reporter in *Chlamydomonas reinhardtii*. *Plant J.* 1999;19(3):353–61.
- Schroda M, Blöcker D, Beck CF. The HSP70A promoter as a tool for the improved expression of transgenes in *Chlamydomonas*. *Plant J.* 2000;21(2):121–31.
- Neupert J, Karcher D, Bock R. Generation of *Chlamydomonas* strains that efficiently express nuclear transgenes. *Plant J.* 2009;57(6):1140–50.
- Lauersen KJ, Kruse O, Mussgnug JH. Targeted expression of nuclear transgenes in *Chlamydomonas reinhardtii* with a versatile, modular vector toolkit. *Appl Microbiol Biotechnol.* 2015;99(8):3491–503.
- Rasala BA, Lee PA, Shen Z, Briggs SP, Mendez M, Mayfield SP. Robust expression and secretion of Xylanase1 in *Chlamydomonas reinhardtii* by fusion to a selection gene and processing with the FMDV 2A peptide. *PLoS One.* 2012;7(8):e43349.
- Zorin B, Grundman O, Khozin-Goldberg I, Leu S, Shapira M, Kaye Y, Tourasse N, Vallon O, Boussiba S. Development of a nuclear transformation system for oleaginous green alga *Lobosphaera* (*Parietochloris*) *incisa* and genetic complementation of a mutant strain, deficient in arachidonic acid biosynthesis. *PLoS One.* 2014;9(8):e105223.
- Guo SL, Zhao XQ, Tang Y, Wan C, Alam MA, Ho SH, Bai FW, Chang JS. Establishment of an efficient genetic transformation system in *Scenedesmus obliquus*. *J Biotechnol.* 2013;163(1):61–8.
- Dautor Y, Úbeda-Mínguez P, Chileh T, García-Maroto F, Alonso DL. Development of genetic transformation methodologies for an industrially-promising microalga: *Scenedesmus almeriensis*. *Trends Biotechnol.* 2014;36(12):2551–8.
- Úbeda-Mínguez P, Chileh T, Dautor Y, García-Maroto F, Alonso DL. Tools for microalgal biotechnology: development of an optimized transformation method for an industrially promising microalga—*Tetraselmis chuii*. *J Appl Phycol.* 2015;27(1):223–32.
- Chen CY, Kao AL, Tsai ZC, Chow TJ, Chang HY, Zhao XQ, Chen PT, Su HY, Chang JS. Expression of type 2 diacylglycerol acyltransferase gene *DGTT1* from *Chlamydomonas reinhardtii* enhances lipid production in *Scenedesmus obliquus*. *Biotechnol J.* 2016;11(3):336–44.
- Simon DP, Anila N, Gayathri K, Sarada R. Heterologous expression of β -carotene hydroxylase in *Dunaliella salina* by *Agrobacterium*-mediated genetic transformation. *Algal Res.* 2016;18:257–65.
- Kumar SV, Misquitta RW, Reddy VS, Rao BJ, Rajam MV. Genetic transformation of the green alga—*Chlamydomonas reinhardtii* by *Agrobacterium tumefaciens*. *Plant Sci.* 2004;166(3):731–8.
- Gim GH, Kim JK, Kim HS, Kathiravan MN, Yang H, Jeong SH, Kim SW. Comparison of biomass production and total lipid content of freshwater green microalgae cultivated under various culture conditions. *Bioprocess Biosyst Eng.* 2014;37(2):99–106.
- Ryan MD, King AM, Thomas GP. Cleavage of foot-and-mouth disease virus polypeptide is mediated by residues located within a 19 amino acid sequence. *J Gen Virol.* 1991;72(11):2727–32.
- Stevens DR, Purton S, Rochaix JD. The bacterial phleomycin resistance gene *ble* as a dominant selectable marker in *Chlamydomonas*. *Mol Gen Genet.* 1996;251(1):23–30.
- Rasala BA, Barrera DJ, Ng J, Plucinak TM, Rosenberg JN, Weeks DP, Oyler GA, Peterson TC, Haerizadeh F, Mayfield SP. Expanding the spectral palette of fluorescent proteins for the green microalga *Chlamydomonas reinhardtii*. *Plant J.* 2013;74(4):545–56.
- Rasala BA, Chao SS, Pier M, Barrera DJ, Mayfield SP. Enhanced genetic tools for engineering multigene traits into green algae. *PLoS One.* 2014;9(4):e94028.
- Kong F, Yamasaki T, Kurniasih SD, Hou L, Li X, Ivanova N, Okada S, Ohama T. Robust expression of heterologous genes by selection marker fusion system in improved *Chlamydomonas* strains. *J Biosci Bioeng.* 2015;120(3):239–45.
- Plucinak TM, Horken KM, Jiang W, Fostvedt J, Nguyen ST, Weeks DP. Improved and versatile viral 2A platforms for dependable and inducible high-level expression of dicistronic nuclear genes in *Chlamydomonas reinhardtii*. *Plant J.* 2015;82(4):717–29.
- McCarthy SS, Kobayashi MC, Niyogi KK. White mutants of *Chlamydomonas reinhardtii* are defective in phytoene synthase. *Genetics.* 2004;168(3):1249–57.
- Rajam MV, Kumar SV. Green alga (*Chlamydomonas reinhardtii*). In: Wang K, editor. *Methods in molecular biology: Agrobacterium protocols*, vol. 2. New Jersey: Human Press; 2006. p. 421–33. Part VIII.
- Pratheesh PT, Vineetha M, Kurup GM. An efficient protocol for the *Agrobacterium*-mediated genetic transformation of microalga *Chlamydomonas reinhardtii*. *Mol Biotechnol.* 2014;56(6):507–15.
- Kathiresan S, Chandrashekar A, Ravishankar GA, Sarada R. *Agrobacterium*-mediated transformation in the green alga *Haematococcus pluvialis* (Chlorophyceae, Volvocales). *J Phycol.* 2009;45(3):642–9.
- Anila N, Chandrashekar A, Ravishankar GA, Sarada R. Establishment of *Agrobacterium tumefaciens*-mediated genetic transformation in *Dunaliella bardawil*. *Eur J Phycol.* 2011;46(1):36–44.
- Cheng R, Ma R, Li K, Rong H, Lin X, Wang Z, Yang S, Ma Y. *Agrobacterium tumefaciens* mediated transformation of marine microalgae *Schizochytrium*. *Microbiol Res.* 2012;167(3):179–86.
- Cha TS, Chen CF, Yee W, Aziz A, Loh SH. Cinnamic acid, coumarin and vanillin: alternative phenolic compounds for efficient *Agrobacterium*-mediated transformation of the unicellular green alga, *Nannochloropsis* sp. *J Microbiol Methods.* 2011;84(3):430–4.
- San Cha T, Yee W, Aziz A. Assessment of factors affecting *Agrobacterium*-mediated genetic transformation of the unicellular green alga, *Chlorella vulgaris*. *World J Microbiol Biotechnol.* 2012;28(4):1771–9.

46. Sanitha M, Radha S, Fatima AA, Devi SG, Ramya M. *Agrobacterium*-mediated transformation of three freshwater microalgal strains. *Pol J Microbiol*. 2014; 63:387–92.
47. Kong F, Yamasaki T, Ohama T. Expression levels of domestic cDNA cassettes integrated in the nuclear genomes of various *Chlamydomonas reinhardtii* strains. *J Biosci Bioeng*. 2014;117(5):613–6.
48. Babinger P, Kobl I, Mages W, Schmitt R. A link between DNA methylation and epigenetic silencing in transgenic *Volvox carteri*. *Nucleic Acids Res*. 2001;29(6):1261–71.
49. Bird A. DNA methylation patterns and epigenetic memory. *Genes Dev*. 2002;16(1):6–21.
50. Fuks F, Hurd PJ, Wolf D, Nan X, Bird AP, Kouzarides T. The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. *J Biol Chem*. 2003;278(6):4035–40.
51. Lumbreras V, Stevens DR, Purton S. Efficient foreign gene expression in *Chlamydomonas reinhardtii* mediated by an endogenous intron. *Plant J*. 1998;14:441–7.
52. Eichler-Stahlberg A, Weisheit W, Ruecker O, Heitzer M. Strategies to facilitate transgene expression in *Chlamydomonas reinhardtii*. *Planta*. 2009;229:873–83.
53. Lauenstein KJ, Berger H, Mussnug JH, Kruse O. Efficient recombinant protein production and secretion from nuclear transgenes in *Chlamydomonas reinhardtii*. *J Biotechnol*. 2013;167(2):101–10.
54. Sizova I, Fuhrmann M, Hegemann PA. *Streptomyces rimosus* *phvIII* gene coding for a new type phosphotransferase provides stable antibiotic resistance to *Chlamydomonas reinhardtii*. *Gene*. 2001;277(1):221–9.
55. Steinbrenner J, Sandmann G. Transformation of the green alga *Haematococcus pluvialis* with a phytoene desaturase for accelerated astaxanthin biosynthesis. *Appl Environ Microbiol*. 2006;72(12):7477–84.
56. Liu J, Gerken H, Huang J, Chen F. Engineering of an endogenous phytoene desaturase gene as a dominant selectable marker for *Chlamydomonas reinhardtii* transformation and enhanced biosynthesis of carotenoids. *Process Biochem*. 2013;48(5):788–95.
57. Liu J, Sun Z, Gerken H, Huang J, Jiang Y, Chen F. Genetic engineering of the green alga *Chlorella zofingiensis*: a modified norflurazon-resistant phytoene desaturase gene as a dominant selectable marker. *Appl Microbiol Biotechnol*. 2014;98(11):5069–79.
58. Lin G, Wang Y, Guo L, Ding H, Hu Y, Liang S, Zhang Z, Yang G. Verification of mutagen function of Zeocin in *Nannochloropsis oceanica* through transcriptome analysis. *J Ocean Univ China*. 2017;16(3):501–8.
59. Chen S, Songkumarn P, Liu J, Wang GL. A versatile zero background T-vector system for gene cloning and functional genomics. *Plant Physiol*. 2009;150(3):1111–21.
60. De Rybel B, van den Berg W, Lokerse AS, Liao CY, van Mourik H, Möller B, Llavata-Peris CI, Weijers D. A versatile set of ligation-independent cloning vectors for functional studies in plants. *Plant Physiol*. 2011;156(3):1292–9.
61. Fischer N, Rochaix JD. The flanking regions of *PsaD* drive efficient gene expression in the nucleus of the green alga *Chlamydomonas reinhardtii*. *Mol Gen Genet*. 2001;265(5):888–94.
62. Pollock SV, Colombo SL, Prout DL, Godfrey AC, Moroney JV. Rubisco activase is required for optimal photosynthesis in the green alga *Chlamydomonas reinhardtii* in a low-CO₂ atmosphere. *Plant Physiol*. 2003; 133(4):1854–61.
63. Jeong JY, Yim HS, Ryu JY, Lee HS, Lee JH, Seen DS, Kang SG. One-step sequence-and ligation-independent cloning as a rapid and versatile cloning method for functional genomics studies. *Appl Environ Microbiol*. 2012; 78(15):5440–3.
64. Mishra SK, Suh WI, Farooq W, Moon M, Shrivastav A, Park MS, Yang JW. Rapid quantification of microalgal lipids in aqueous medium by a simple colorimetric method. *Bioresour Technol*. 2014;155:330–3.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions





Towards sex identification of Asian Palmyra palm (*Borassus flabellifer* L.) by DNA fingerprinting, suppression subtractive hybridization and *de novo* transcriptome sequencing

Kwanjai Pipatchartlearnwong¹, Piyada Juntawong^{1,2,3}, Passorn Wonnapijit^{1,2,3}, Somsak Apisitwanich^{1,2} and Supachai Vuttipongchaikij^{1,2,3}

¹ Department of Genetics, Faculty of Science, Kasetsart University, Bangkok, Thailand

² Center of Advanced studies for Tropical Natural Resources, Kasetsart University, Bangkok, Thailand

³ Omics Center for Agriculture, Bioresources, Food and Health, Kasetsart University (OmiKU), Bangkok, Thailand

ABSTRACT

Background. Asian Palmyra palm, the source of palm-sugar, is dioecious with a long juvenile period requiring at least 12 years to reach its maturity. To date, there is no reliable molecular marker for identifying sexes before the first bloom, limiting crop designs and utilization. We aimed to identify sex-linked markers for this palm using PCR-based DNA fingerprinting, suppression subtractive hybridization (SSH) and transcriptome sequencing.

Methods. DNA fingerprints were generated between males and females based on RAPD, AFLP, SCoT, modified SCoT, ILP, and SSR techniques. Large-scale cloning and screening of SSH libraries and *de novo* transcriptome sequencing of male and female cDNA from inflorescences were performed to identify sex-specific genes for developing sex-linked markers.

Results. Through extensive screening and re-testing of the DNA fingerprints (up to 1,204 primer pairs) and transcripts from SSH (>10,000 clones) and transcriptome data, however, no sex-linked marker was identified. Although *de novo* transcriptome sequencing of male and female inflorescences provided ~32 million reads and 187,083 assembled transcripts, PCR analysis of selected sex-highly represented transcripts did not yield any sex-linked marker. This result may suggest the complexity and small sex-determining region of the Asian Palmyra palm. To this end, we provide the first global transcripts of male and female inflorescences of Asian Palmyra palm. Interestingly, sequence annotation revealed a large proportion of transcripts related to sucrose metabolism, which corresponds to the sucrose-rich sap produced in the inflorescences, and these transcripts will be useful for further understanding of sucrose production in sugar crop plants. Provided lists of sex-specific and differential-expressed transcripts would be beneficial to the further study of sexual development and sex-linked markers in palms and related species.

Submitted 30 January 2019

Accepted 6 June 2019

Published 10 July 2019

Corresponding author
Supachai Vuttipongchaikij,
fsciscv@ku.ac.th

Academic editor
Gerard Lazo

Additional Information and
Declarations can be found on
page 19

DOI 10.7717/peerj.7268

© Copyright
2019 Pipatchartlearnwong et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

Subjects Bioinformatics, Genetics, Genomics, Molecular Biology, Plant Science

Keywords Sex-linked marker, Palm-sugar, Dioecious, DNA marker, Arecaceae

INTRODUCTION

Asian Palmyra palm (*Borassus flabellifer* L, $2n = 36$) is a dioecious and slow-growing perennial tree, requiring 12–15 years to reach its maturity and produce the first inflorescence (Kovoor, 1983). Once flowered, it vigorously and continuously produces flowers and fruits through its lifespan. This palm is found widespread in South and Southeast Asia and provides essential food and economic values throughout its parts including inflorescence sap for producing palm sugar and alcoholic beverages, fruits for consumption and the tree trunk for construction (Morton, 1988; Lim, 2012). Both sexes are morphologically identical except for the male and female inflorescences, and there are no reliable means of sex identification before the first bloom. Because of the long juvenile period and a preference for female plants for fruit production, growers hesitate to expand the plantation, and this limits the utilization of this palm. Being able to identify the sexes of Asian Palmyra palm at seedling stages will provide better designs for the crop production through optimal male and female ratios, breeding programs, conservation and utilization (Davis & Johnson, 1987; Barfod et al., 2015; Sirajuddin et al., 2016). Molecular markers for sex identification in Asian Palmyra palm is needed.

Asian Palmyra palm belongs to family Arecaceae, which includes many palm species that are among the world commercially important crops, including oil palm (*Elaeis oleifera*), coconut (*Cocos nucifera*) and date palm (*Phoenix dactylifera*) (Beck & Balick, 1990; Dransfield et al., 2005). Historically, it was placed among these valued crops as for producing palm sugar, alcohol products and its nutritious fruits dating back at least 2,500 years (Fox, 1977; Ferguson, 1888), before sugarcane gradually replaced its status. Historical and genetic studies suggest that Asian Palmyra palm was originated from tropical Africa, brought along the spice route to the east and settled in the Indian subcontinent, where it is widely propagated throughout India and Sri Lanka, before introduced to Southeast Asia more than 1,500 years ago (Pipatchartlearnwong et al., 2017a). Given its extended benefits since the historical time and because of its vigorous growth and continuous supply of inflorescence sap and fruits with the ability to withstand severe climate, arid conditions, pests and diseases, Asian Palmyra palm should be recognized as a potential recalcitrant food crop to the climate change for the tropical and subtropical regions.

Molecular data of Asian Palmyra palm are currently limited. The genome sequence is not available, and only its chloroplast genome was recently reported (Sakulsathaporn et al., 2017). A few sex-linked markers have been developed based on DNA fingerprinting, but only one RAPD-based marker was shown to be able to identify the sexes among the populations in India (George et al., 2007). However, our preliminary experiment showed that this marker was unable to identify the sexes in the population in Thailand. Previously, we showed that Asian Palmyra palm populations in Thailand are descended from a small number of seedlings brought in at least 1,500 years ago and represent a very narrow genetic diversity (Pipatchartlearnwong et al., 2017a; Pipatchartlearnwong et al., 2017b).

In this work, we aimed to develop sex-linked markers of Asian Palmyra palm through three approaches: DNA fingerprinting, direct cloning of subtraction subtractive hybridization (SSH) of cDNA from male and female inflorescences and *de novo*

transcriptome sequencing of male and female inflorescences. Extensive sets of PCR primers belonging to different DNA fingerprinting techniques including RAPD, AFLP, SCoT, ILP, TEs, EST-SSR and gSSR and those that specific to transcripts identified from SSH and transcriptome analysis were exhaustively tested to identify sex-linked markers. Although none of the sex-linked markers was obtained from this study, we have narrowed the path towards the sex identification of Asian Palmyra palm. Because identifying sex-linked markers for this species appeared to be extremely difficult as opposed to works in other dioecious plants ([Heikrujam et al., 2015](#)), we discussed the nature of this work in conjunction with others successfully identified markers with future directions. Furthermore, this work provided the first *de novo* transcriptome sequencing of Asian Palmyra palm. Lists of candidate transcripts that are specific to sexes and developmental stages of male and female inflorescences are presented, and these will be useful for further study on sex determination, sexual development and floral development of Asian Palmyra palm.

MATERIALS & METHODS

Plant materials

For DNA isolation, young leaves were collected from mature palm plants with known sexes from various locations in three regions of Thailand: the southern region (Song-Khla and Surat-Thani provinces), the central region (Phachinburi, Phetchaburi, Pathum-Thani, Nakon-Pathom, Kanchanaburi, Nakhon-Sawan and Chainat provinces) and the northeastern region (Nakhon-Ratchasima, Burirum, Ubon-Ratchathani, Kalasin and Amnat-Charoen provinces). For RNA isolation, male and female inflorescences were collected from the southern region (Song Khla province) and the central region (Nakon-Pathom province).

Nucleic acid isolation

Total DNA was isolated from young leaves using a modified method based on the CTAB method as described previously in [Pipatchartlearnwong et al. \(2017a\)](#) and [Pipatchartlearnwong et al. \(2017b\)](#). Total RNA was isolated from young inflorescences (see [Fig. 1](#)) using the modified CTAB method. Briefly, the sample was pulverized into fine powder in liquid N₂ by using a mortar and pestle and mixed with CTAB extraction buffer [2% (w/v) CTAB, 100 mM Tris-HCl pH.8, 20 mM EDTA and 1.4 M NaCl, 2% (w/v) polyvinylpyrrolidone-90 and 2% (v/v) β-mercaptoethanol]. RNA was then precipitated using 1/3 volume of 10 M LiCl at −20 °C for overnight and then centrifuged at 11,750× g at 4 °C for 30 min. Total RNA was treated with DNase I (New England BioLabs® Inc., Ipswich, MA, USA) at 37 °C for 30 min followed by phenol: chloroform extraction and ethanol precipitation. RNA quality and quantity were analyzed by agarose gel electrophoresis and NanoDrop (Thermo Scientific, Waltham, MA, USA).

DNA fingerprinting

The RAPD-based OPA-06 marker was conducted according to [George et al. \(2007\)](#) using total DNA from 20 male and 20 female samples. AFLP analysis was performed using 26

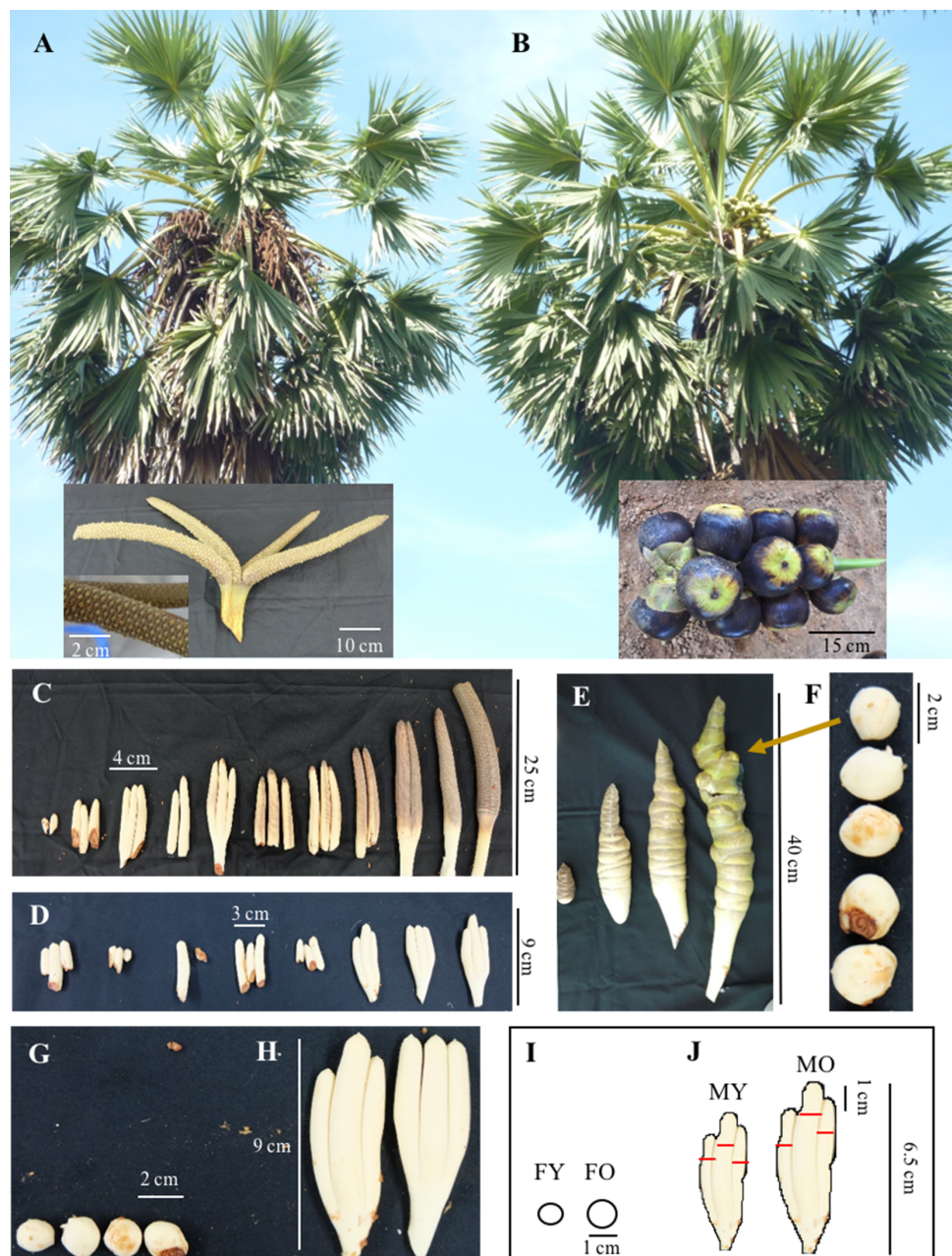


Figure 1 Asian Palmyra palm samples from male and female plants. (A) A mature male plant with a fully developed male inflorescence and (B) a mature female plant with a female inflorescence bearing fruits. (C) and (D) Representatives of developing male inflorescences from young to mature stages. (E) Representatives of developing female inflorescences, and (F) buds extracted from female inflorescences. (G–J) Representative female and male inflorescence samples used for the SSH experiment (G and H, respectively) and *de novo* transcriptome sequencing analysis (I and J, respectively). Scale bars are indicated within.

Full-size [DOI: 10.7717/peerj.7268/fig-1](https://doi.org/10.7717/peerj.7268/fig-1)

male and 46 female samples by, firstly, restriction digestion of 250 ng total DNA using EcoRI and MseI (Thermo Fisher Scientific, USA) and adaptor ligation. And, secondly, pre-selective amplification was performed in a 25 µl reaction volume containing 2 µl of the digested DNA, 5 mM dNTPs, 40 mM MgCl₂, 5 µM of each MseI adaptor+C and EcoRI adaptor+A primers and 1 unit of *Taq* DNA polymerase (Vivantis, Malaysia) using the conditions as follows: 20 cycles of 94 °C for 30 s, 56 °C for 1 min and 72 °C for 1 min, with a final extension step at 72 °C of 5 min. The reaction was diluted 20-fold using dH₂O before being used in a selective amplification reaction: 25 µl total volume containing 5 µl of the diluted DNA, 4 mM dNTPs, 40 mM MgCl₂, 5 µM MseI+3 and 5 µM of EcoRI+3 primers (Table S1) and 1 unit of *Taq* DNA polymerase using a touch-down condition (12 cycles of 94 °C for 30 s, 65 °C (−0.7 °C/cycle) for 30 s and 72 °C for 1 min, followed by 23 cycles using the annealing temperature at 56 °C). SCoT was performed using cDNA from four male and four female samples. SCoT primers were based on Collard & Mackill (2009) (Table S2). The PCR reaction was performed in a 20 µl volume, which included 50 ng of DNA, 4.8 mM dNTPs, 30 mM MgCl₂, 20 µM SCoT primer and 1 unit of *Taq* polymerase (Vivantis, Malaysia) using 35 cycles of 94 °C for 1 min, 50 °C for 1 min and 72 °C for 2 min, with a final extension for 5 min. The modified SCoT method was performed by mixing SCoT primers with primers from polyA, EST-SSRs of oil palm and transposon element (TE) markers (Tables S3 and S4) (Bureau & Wessler, 1994; Takata, Kishima & Sano, 2005). ILP marker based on Ukoskit & U-thoomporn (2016) (Table S5) was performed using total DNA from three male and three female samples. Each PCR reaction were performed in a 20 µl volume containing 20 ng of DNA, 4 mM dNTPs, 75 mM MgCl₂, 3.5 µM for each primer and 0.5 unit of *Taq* polymerase (Vivantis, Selangor, Malaysia) using 35 cycles of 94 °C for 30 s, 56 °C for 1.30 min and 72 °C for 30 s, with a final extension for 5 min. EST-SSR and gSSR microsatellite markers (Table S6 and S7, based on Arabnezhad et al., 2012; Billotte et al., 2004; Elmeer et al., 2011; Pipatchartlearnwong et al., 2017a; Pipatchartlearnwong et al., 2017b) were performed using total DNA from four males and four female samples using the same PCR condition as above. PCR products for RAPD, SCoT and ILP were resolved in 2% (w/v) agarose gel electrophoresis and visualized under ultraviolet (UV) light after staining with ethidium bromide. Those for AFLP, SCoT+EST-SSR, SCoT+TEs, EST-SSR and gSSR were resolved in 6% (w/v) polyacrylamide gel electrophoresis and visualized by silver staining.

Suppression subtractive hybridization (SSH)

Total RNA from one male and one female inflorescence samples from Song Khla province were used in the study (Figs. 1G–1H). The first-strand and double strand cDNA was synthesized from 1 µg of total RNA samples using the SMART cDNA Library Construction Kit (Clontech Laboratories Inc., Palo Alto, CA, USA) following to the manufacturer protocol. Double-strand cDNA of male and female was labeled with biotin-16 dUTP using nick translation: 100 µl total volume containing 400 ng of double-strand cDNA, 3.96 mM of dGTP, dATP and dCTP, 1.6 mM of dTTP, 2.4 µl of biotin-16 dUTP (Sigma-Aldrich), 48 µM of 5'PCR PrimerII A (5'-AGCAGTGGTATACAACGCAGAGT-3'), 10x *Taq* buffer, and 5 unit of *Taq* polymerase (Vivantis, Selangor, Malaysia) using 40 cycles of 94 °C for

15 s, 65 °C for 30 s, and 68 °C for 6 min. The biotin-16 dUTP labeled cDNA was checked for its detection signal using dot blot hybridization at least at 10^{-4} dilution. First-strand cDNA of male and female cDNA was subtracted using labeled double-strand cDNA of its opposite sex at ratio 1:3 following a method from [Rebrikov et al. \(2004\)](#). Subtracted samples were purified by ethanol precipitation, and double-strand cDNA was synthesized by PCR reaction containing 2 µl of subtracted cDNA, 48 µM of 5' PCR PrimerII A, 10 mM dNTPs, 5x NEB polymerase buffer and 5 unit of Q5[®] High-Fidelity DNA polymerase (New England Biolabs) through 21 cycles of 94 °C for 15 s, 65 °C for 30s and 68 °C for 6 min. Subtracted double-strand cDNA was purified and ligated into pGEM-Teasy (Promega). Clones were selected by blue-white colony section and screened using colony PCR for > 500 bp inserted fragment, before re-selection using dot blot hybridization I and II using probes from opposite sexes. Selected clones were sequenced and analyzed against GenBank using BLAST.

Transcriptome analysis

Transcriptome sequencing was performed by Macrogen Inc. (Republic of Korea) using 10 µg of total RNA. Total RNA was obtained from two male (MY and MO) and two female (FY and FO) inflorescence samples from Nakon-Pathom province ([Figs. 1I–1J](#)). Briefly, cDNA libraries were constructed using the TruSeq[™] RNA sample preparation kit (Illumina, USA) and sequenced on a HiSeq 2000 (Illumina, USA) with paired-end 100 bp read lengths. Initial raw reads were trimmed and filtered with FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>) and Trimmomatic version 0.32 (<http://www.usadellab.org/cms/?page=trimmomatic>) software to eliminate low-quality reads (quality score lower than 20) and remove adapters. Reads were considered as high quality if more than 70% of the bases had Phred values more than Q20. Reads produced from this study were assembled by Trinity software (version r20140717) using default parameters ([Grabherr et al., 2011](#)). For similarity search, the assembled transcripts were blasted against the NCBI non-redundant protein sequence database and TAIR database using Blast2GO with the e-value cutoff <10⁻¹⁰. RSEM version 1.2.15 software was used to estimate transcript abundance ([Li & Dewey, 2011](#)). The assembled sequences were analyzed for Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG: [Kanehisa & Goto, 2000](#)) using Blast2GO. For differential gene expression analysis, Fastq files were aligned to the assembled transcriptome using TopHat2 alignment program ([Kim et al., 2013](#)). Transcriptome annotation file (GFF) was built using the Cufflinks program ([Trapnell et al., 2010](#)) by performing a combined assembly of four transcriptome datasets (FY, FO, MY, and MO). Transcript candidates for male and female were chosen using parameters as follows; male-highly represented transcripts [FPKM > 10 and > 0.5 for MY or MO and FPKM = 0 for both FY and FO] and [FPKM > 5 for both MY and MO and FPKM = 0 for both FY and FO] with length > 300 bp, and female-highly represented transcripts [FPKM > 10 and > 0.5 for FY or FO and FPKM = 0 for both MY and MO] and [FPKM > 5 for both FY and FO and FPKM = 0 for both MY and MO] with length > 300 bp. To identify shared transcripts among the four datasets, transcripts from each dataset with FPKM > 5 and length > 300 bp were analyzed using Venny version 2.1.0. Identified male

Table 1 Total markers tested for DNA fingerprints between male and female plants. Numbers of male and female samples used for each marker are indicated in brackets as M and F, respectively.

Marker	Total tested markers	Amplifiable	Polymorphic marker (loci)	Sex polymorphic loci	Sex-linked marker
RAPD	1	1 (F23:M20)	0	0	–
AFLP	36	36 (F46:M26)	13 (141 loci)	0	–
SCoTs	36	36 (F4:M4)	3	0	–
SCoTs/A	36	36 (F4:M4)	3	0	–
SCoTs (36)/EST-SSRs (3)	108	64 (F4:M4)	7 (7 loci)	2 (F8:M8)	–
SCoTs (36)/TEs (11)	396	163 (F4:M4)	21 (48 loci)	48 (F4:M4)	–
Oil palm ILPs	41	36 (F3:M3)	0	0	–
Date palm gSSRs (high PIC values)	5	2 (F4:M4)	0	0	–
Oil palm EST-SSRs	289	150 (F20:M20)	11 (17 loci)	0	–
Oil palm gSSRs	256	168 (F20:M20)	8 (12 loci)	0	–
Total	1,204	583	66 (231)	50	–

and female specific transcripts were subjected to GO enrichment analysis using Fisher's Exact Test with FDR cutoff = 0.01. Differentially expressed genes (DEGs) among the four transcript datasets were analyzed by Cuffdiff with FDR cutoff <0.05.

RESULTS

Extensive DNA fingerprinting analysis failed to identify any sex-linked marker for Asian Palmyra palm in Thailand

Previously, *George et al. (2007)* had developed a RAPD based male-specific marker (OPA-06₆₀₀) for the Asian Palmyra palm population in India. Initially, we tested this marker for sex identification in our population in Thailand using up to 20 male and 24 female individuals, but this marker was unable to confirm the sexes (*Fig. S1*). To identify sex-linked markers for Asian Palmyra palm in Thailand, we generated male and female DNA fingerprints using ten different DNA fingerprinting techniques as presented in *Table 1*. Although a number of potential sex-linked bands were obtained from SCoTs/EST-SSRs and SCoTs/TEs, after sequencing and re-testing these bands using specific primers, no sex-linked marker was obtained. Despite such extensive screening of DNA markers up to 1,204 primer pairs by the ten techniques, we did not obtain any sex-linked marker. This experiment showed that the DNA fingerprinting covered here is inadequate for identifying a sex-linked marker for Asian Palmyra palm. It also suggests that sex determination region in this species could be of small and very difficult to identify. Further attempts for this scheme in the Asian Palmyra palm should be aware of this limitation. Other means of sex identification for Asian Palmyra palm should be explored.

Identification of sex-related transcripts by SSH analysis using male and female inflorescence flowers

To identify genes related to sexes, we performed SSH using total RNA isolated from male and female inflorescences of Asian Palmyra palm. Because the floral development of this species is not well defined, we collected young inflorescence stems from male and female

Table 2 Clone selection of SSH between male and female.

Library	Female	Male
Direct cloning by blue-white colony selection	9,820	2,574
Fragment size > 500 bp by PCR screening	4,097	1,288
Dot blot hybridization I	498	112
Dot blot hybridization II	63	81
BlastX	29	60

plants as soon as they emerged from the dense leaf sheets and isolated young floral tissues for RNA isolation. Female flower buds (~2 cm in diameter) were removed from the inflorescence (Fig. 1G), while the male inflorescences (~9 cm in length) were used as a whole (Fig. 1H). Direct cloning of subtracted-cDNA yielded 2,574 and 9,820 clones for male and female, respectively, and, after colony-PCR screening for > 500 bp inserted-fragments, we obtained 1,288 and 4,097 clones for male and female libraries, respectively (Table 2). These selected clones were re-tested against their opposite sex using two rounds of dot blot hybridization (I and II), resulting in 81 and 63 clones for male- and female-specific libraries, respectively. These clones were subsequently sequenced and searched in GenBank using BLASTX, and 60 male and 29 female clones were found matching to non-redundant genes in the plant database (Table S8). These sequences have been deposited in GenBank (JZ977504–JZ977592) as ESTs for male or female inflorescence flowers of Asian Palmyra palm.

Among the total 99 identified clones, 91 and 98 clones share high similarities to sequences in the nuclear genome of date palm and oil palm, respectively. As oil palm has both male and female genome sequences available, we observed that all 99 sequences identified in Asian Palmyra palm are present in both male and female genomes of oil palm, suggesting that these sequences could not be used as sex-specific markers. Nonetheless, we tested 19 selected sequences on gDNA from male and female plants by PCR using specific primers (Table S9), and these failed to identify the sexes. Although this experiment was unable to provide a sex-linked marker, the list of expressed genes during male- and female-floral development of Asian Palmyra palm could be of use for future study.

***De novo* transcriptome sequencing of male and female inflorescences**

To further identify sex-related genes for sex identification, we performed *de novo* transcriptome sequencing using RNA from male and female inflorescences. Four cDNA libraries were constructed from two floral stages of male and female (FY-female young inflorescences, FO-female old inflorescences, MY-male young inflorescences and MO-male old inflorescences; Figs. 1I–1J), which were in earlier stages than those used in the SSH experiment. From Illumina HiSeq2000 sequencing, we obtained 47,194,682–90,305,610 reads for each sample with an average length of 98 bp, after trimming the adapter sequences and removing low-quality nucleotides (<Q20) and short sequences (<25 nt) (Fig. 2). Sequence assembly using Trinity yielded 187,083 transcripts with 705 bp average length. Most of the assembled transcripts were between 200 and 300 bp in length, and up to 70,785 transcripts were > 500 bp in length (Fig. 2B). The transcripts were

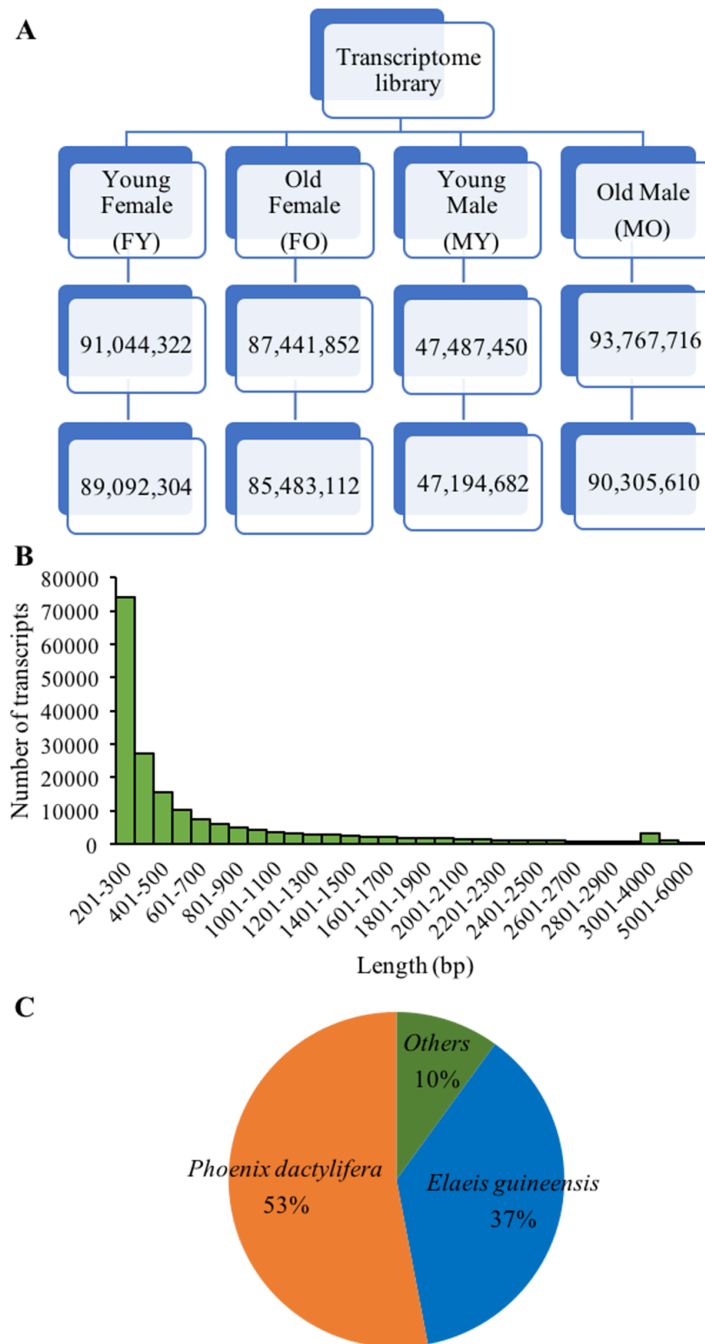


Figure 2 *De novo* transcriptome sequencing overview. (A) Total sequence reads (bp) after trimming and removing low quality read (<Q20) of four samples. (B) Distribution of number and length of assembled transcripts, and (C) sequence similarity of assembled transcripts to other plant species.

Full-size [DOI: 10.7717/peerj.7268/fig-2](https://doi.org/10.7717/peerj.7268/fig-2)

annotated by using BLASTX searches against the NCBI non-redundant protein database, and 77,578 transcripts (41.47% of initial transcripts) were identified as unique sequences with significant similarities ($E < 1e-10$) to known protein sequences from 5,331 different

species. The average alignment length (%) matched was 81.3% with the highest positive matched per alignment length being 90%, and the average percent identity was 81.24%. Most of these sequences were identified as similar sequences to those found in date palm (53%) and oil palm (37%) (Fig. 2C). The Transcriptome Shotgun Assembly are deposited at DDBJ/ENA/GenBank under the accession [GFYQ00000000](#).

GO classification and pathway assignment by KEGG

Gene ontology (GO) terms of the 77,578 annotated transcripts were assigned using Blast2GO program in three categories: biological process, cellular component and molecular function. The transcripts were assigned into 66 functional groups, and top GO terms with more than 1% assigned transcripts in each category are presented in Fig. S2. Binding and catalytic activity were dominant in the molecular function category, while the integral component of the membrane and nucleus dominated the cellular component category. For the biological process, oxidation–reduction process and protein phosphorylation were the most represented groups. Subsequently, function classification and pathway assignments based on Kyoto Encyclopedia of Genes and Genomes (KEGG) showed that, among the 77,578 annotated transcripts, 16,635 were annotated with enzyme code EC numbers and mapped into 139 KEGG pathways. Pathways with more than 1% matched transcripts are presented in Fig. S3. Purine (map00230) and Thymine (map00730) metabolisms were the most matched pathways with 4,913 transcripts (18.89%) and 2,581 transcripts (9.92%), followed by Biosynthesis of antibiotics (map01130) with 1,582 transcripts (6.08%). Interestingly, starch and sucrose metabolism (map00500) presented at the fourth rank with 1,187 transcripts (4.56%). Transcripts mapped into this pathway were mostly related to sugar metabolism for fructose and sucrose production, but less supported to starch biosynthesis (Fig. 3 and Table 3; see the transcript IDs in Table S10). Other related sugar metabolic pathways were also found among the list of top pathways including 1.54% of glycolysis/gluconeogenesis (map00010), 1.25% of pentose and glucuronate interconversions (map00040), 1.19% of Galactose metabolism (map00052), 1.10% of fructose and mannose metabolism (map00051), 0.87% of pentose phosphate pathway (map00030) and 0.85% of inositol phosphate metabolism (map00562). This observation coincides with the facts that both male and female inflorescences of Asian Palmyra palm produce sweet sap, which has been used for making palm sugar for centuries.

Differential expression of genes between male and female inflorescences

Among the initial 187,083 transcripts, we observed a number of transcripts that were highly represented in either male or female datasets, but none in their opposite sex (FPKM = 0); 33 and 11 transcripts were identified from female and male datasets, respectively (length > 300 bp, FPKM > 5 for both FY and FO or MY and MO, and FPKM > 10 and > 0.5 for FY or FO and MY or MO) (Table S11). Although annotations of these transcripts did not show any evidence related to genes for male- or female-specific development, we observed four cell wall-related transcripts in the female datasets, and this may reflect that the developing female flower undergoes rapid and extensive organ enlargement, whereas the male flower

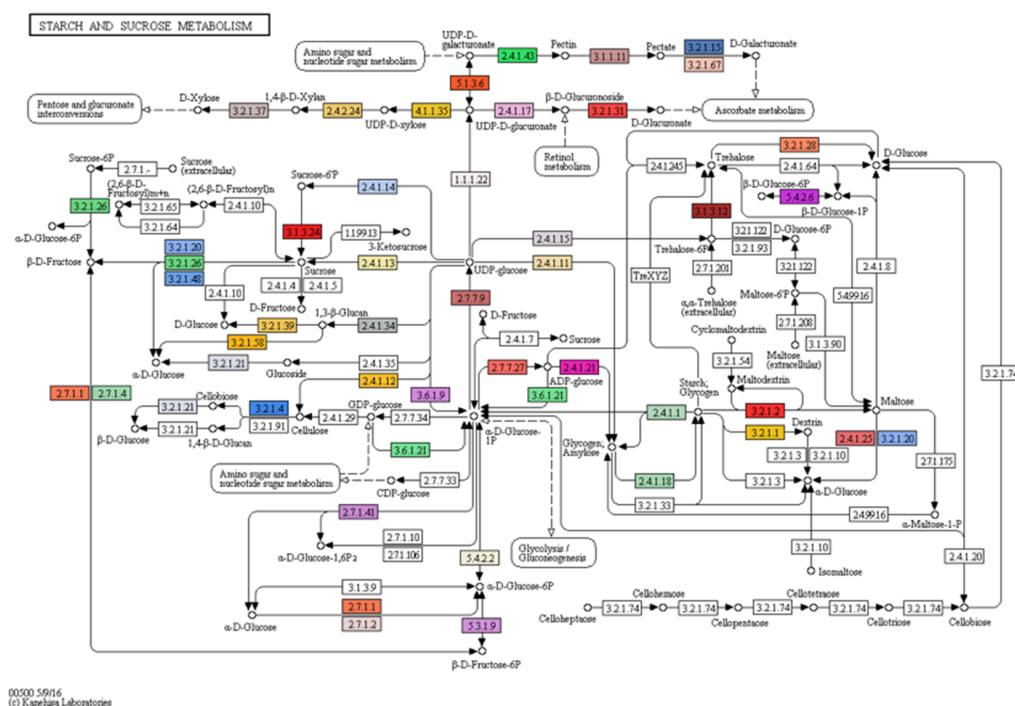


Figure 3 Mapping of assembled transcripts onto the starch and sucrose metabolism (map00500) taken from KEGG. Mapped enzymes are indicated by colors (see Table S10 for details). Kanehisa Laboratories 00500 5/9/16.

Full-size [DOI: 10.7717/peerj.7268/fig-3](https://doi.org/10.7717/peerj.7268/fig-3)

is limited to a small size. Because the genome sequence of Asian Palmyra palm is currently unavailable, we thus tested whether these transcripts could be used for sex identification by PCR using male and female gDNA and primers specific to these 44 transcript sequences (Table S12). However, these primers gave similar band patterns between male and female gDNA (Fig. S4), indicating that these primers cannot be used for identifying the sexes.

By comparing the transcripts among the four datasets (length >300 bp and FPKM > 5), we found that 17,231 transcripts were shared among all datasets, and 4,514 (14.4%) and 3,312 (10.5%) transcripts were identified as male- and female-specific transcripts, respectively (Fig. 4 and Table S13). These transcripts were further classified into specific MY (2,192 transcripts), MO (1,405 transcripts), FY (1,355 transcripts) and FO (1,165 transcripts) to identify transcripts that may relate to male or female floral development stages. Subsequently, we analyzed enriched GO terms for the male- and female-specific datasets (length > 300 bp and FPKM > 5), and top enriched GO terms (FDR <1E−02) are presented in Table 4 (see Table S14 for transcript IDs). All GO terms observed here were over-represented compared to the reference sets. Interestingly, carbohydrate metabolism and cell wall-related processes were much apparent in the female-specific transcripts, while various catabolic processes for biological compounds dominated the male-specific transcripts.

In light of this transcriptome analysis, we cross-referenced the 60 male and 29 female-specific clones from the SSH experiment to the transcript abundance data. Noting that

Table 3 The list of enzymes in starch and sucrose metabolism (KEGG map00500) identified in male and female transcripts from inflorescences of Asian Palmyra palm.

Enzyme	EC number	Number of transcripts
decarboxylase	EC:4.1.1.35	11
phosphodismutase	EC:2.7.1.41	1
endo-1,4-beta-D-glucanase	EC:3.2.1.4	56
saccharogen amylase	EC:3.2.1.2	42
glycogenase	EC:3.2.1.1	20
1,4-alpha-galacturonidase	EC:3.2.1.67	5
isomerase	EC:5.3.1.9	6
alpha-glucosidase	EC:3.2.1.48	42
1,3-beta-glucosidase	EC:3.2.1.58	6
adenylyltransferase	EC:2.7.7.27	15
maltase	EC:3.2.1.20	2
gentiobiase	EC:3.2.1.21	106
synthase	EC:2.4.1.34	39
trehalase	EC:3.2.1.28	5
invertase	EC:3.2.1.26	45
beta-glucuronide glucuronohydrolase glucuronidase	EC:3.2.1.31	6
endo-1,3-beta-D-glucosidase	EC:3.2.1.39	30
4-alpha-galacturonosyltransferase	EC:2.4.1.43	59
1,4-beta-xylosidase	EC:3.2.1.37	29
phosphorylase	EC:2.4.1.1	28
branching enzyme	EC:2.4.1.18	7
1-naphthol glucuronyltransferase	EC:2.4.1.17	33
synthase (UDP-forming)	EC:2.4.1.15	29
synthase	EC:2.4.1.14	18
synthase	EC:2.4.1.13	44
synthase (UDP-forming)	EC:2.4.1.12	98
synthase	EC:2.4.1.11	24
disproportionating enzyme	EC:2.4.1.25	8
synthase (glycosyl-transferring)	EC:2.4.1.21	13
pectin depolymerase	EC:3.2.1.15	75
diphosphatase	EC:3.6.1.9	9
glucokinase (phosphorylating)	EC:2.7.1.2	14
hexokinase type IV glucokinase	EC:2.7.1.1	44
fructokinase (phosphorylating)	EC:2.7.1.4	35
trehalose 6-phosphatase	EC:3.1.3.12	58
4-epimerase	EC:5.1.3.6	4
diphosphatase	EC:3.6.1.21	7
phosphatase	EC:3.1.3.24	6
uridylyltransferase	EC:2.7.7.9	7
pectin demethoxylase	EC:3.1.1.11	69
synthase	EC:2.4.2.24	6
(alpha-D-glucose-1,6-bisphosphate-dependent)	EC:5.4.2.2	9
beta-pgm (gene name)	EC:5.4.2.6	3
6-dehydrogenase	EC:1.1.1.22	14

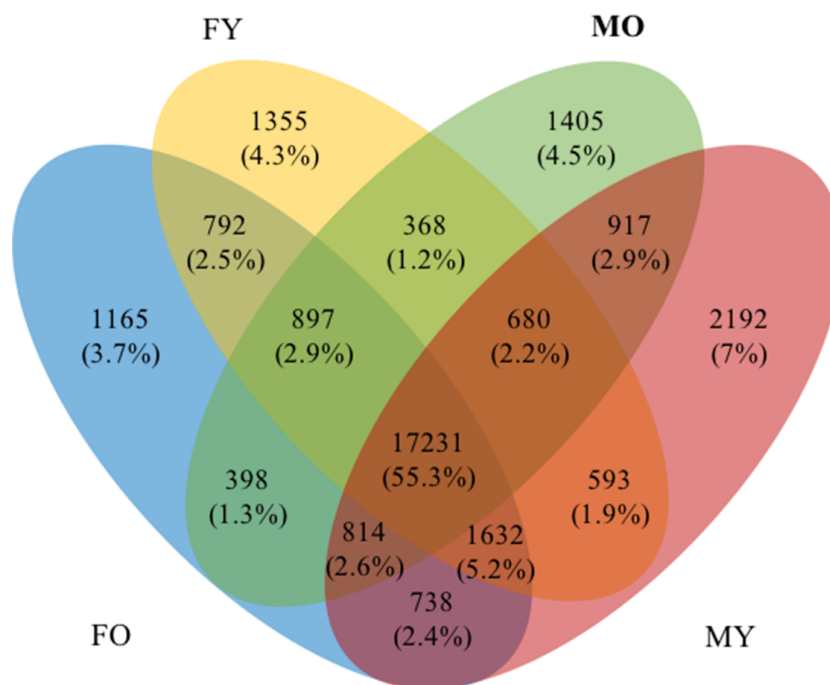


Figure 4 A Venn diagram of four transcript datasets. Numbers and percentages of transcripts are indicated.

Full-size [DOI: 10.7717/peerj.7268/fig-4](https://doi.org/10.7717/peerj.7268/fig-4)

RNA samples used in the SSH experiment were obtained from an older inflorescence stage than those used for transcriptome analysis. The transcript IDs were readily identified with more than 90% identical matches and > 300 transcript length (Table S15). However, we found that, based on FPKM values, the sex-specific clones were indeed uncorrelated to almost all of the transcript data, and only F152 (predicted proline-rich protein 4-like) clone could be identified for their expression towards the female. Though, this clone has no direct relationship to sex development based on the annotation.

To further verify sex-related genes, we analyzed differentially expressed genes among the four datasets by using Cuffdiff with FDR < 0.05 cutoff. Initially, 816 transcripts were identified, but only 43 transcripts displayed differential expression between sexes with transcript length > 300 bp, FPKM > 5, GO terms and significant q value (< 0.05) (Table 5). Among these 43 transcripts, seven transcripts were annotated with genes previously identified to be involved in sex determination and flower development: two and five transcripts for female and male datasets, respectively. Furthermore, we observed six and one transcripts encoding transcription factors that were highly expressed in female and male datasets, respectively. Although being identified for differential expression between the sexes, c1819_g1_i1 and c142400_g1_i1 transcripts encoding ethylene-responsive transcription factors were highly expressed throughout the four datasets, and these genes may be required for the floral development. The list of gene candidates indicated here could be used for a further study on sex determination and floral development in Asian Palmyra palm and other related palm species.

Table 4 Enriched GO terms of male- and female-specific transcripts (FDR <1E−02).

Enriched female			Enriched male		
Go term	Cat	FDR	Go term	Cat	FDR
carbohydrate metabolic process [GO:0005975]	P	4.98E−05	L-allo-threonine aldolase activity [GO:0008732]	F	1.12E−09
phosphoribulokinase activity [GO:0008974]	F	4.98E−05	threonine aldolase activity [GO:0043876]	F	1.12E−09
cell wall organization or biogenesis [GO:0071554]	P	4.98E−05	aspartate family amino acid catabolic process [GO:0009068]	P	1.34E−08
plant-type cell wall organization or biogenesis [GO:0071669]	P	1.09E−03	threonine catabolic process [GO:0006567]	P	1.34E−08
fucose metabolic process [GO:0006004]	P	1.09E−03	glycine biosynthetic process [GO:0006545]	P	3.60E−08
cell wall biogenesis [GO:0042546]	P	1.92E−03	isovaleryl-CoA dehydrogenase activity [GO:0008470]	F	9.54E−06
plant-type secondary cell wall biogenesis [GO:0009834]	P	2.50E−03	threonine metabolic process [GO:0018927]	P	1.20E−05
plant-type cell wall biogenesis [GO:0009832]	P	2.50E−03	simple leaf morphogenesis [GO:0060776]	P	8.86E−05
cellular carbohydrate biosynthetic process [GO:0034637]	P	2.50E−03	branched-chain amino acid catabolic process [GO:0009083]	P	1.65E−04
carbohydrate biosynthetic process [GO:0016051]	P	2.50E−03	leucine catabolic process [GO:0006552]	P	1.34E−03
pyrimidine nucleoside salvage [GO:0043097]	P	3.12E−03	AMP biosynthetic process [GO:0006167]	P	1.41E−03
O-acetyltransferase activity [GO:0004026]	F	3.43E−03	AMP metabolic process [GO:0046033]	P	1.41E−03
Membrane [GO:0016020]	C	4.19E−03	methylated histone binding [GO:0035064]	F	1.41E−03
uridine kinase activity [GO:0004849]	F	4.50E−03	adenylosuccinate synthase activity [GO:0004019]	F	1.41E−03
nucleoside salvage [GO:0043174]	P	4.50E−03	PeBoW complex [GO:0070545]	C	1.41E−03
plant-type cell wall organization [GO:0009664]	P	5.70E−03	cellular amino acid catabolic process [GO:0009063]	P	1.41E−03
single-organism carbohydrate metabolic process [GO:0005975]	P	5.83E−03	phosphoacetylglucosamine mutase activity [GO:0004610]	F	1.74E−03
Endosome [GO:0005768]	C	6.68E−03	de novo' AMP biosynthetic process [GO:0044208]	P	2.33E−03
cell wall organization [GO:0071555]	P	8.97E−03	leucine metabolic process [GO:006551]	P	3.20E−03
			geranylgeranyl-diphosphate geranylgeranyltransferase activity [GO:0016767]	F	3.26E−03
			ELL-EAF complex [GO:0032783]	C	6.81E−03
			regulation of rRNA processing [GO:2000232]	P	9.23E−03

Notes.
P, Process; F, Function; C, Cellular.

DISCUSSION

Previously, there were only two reports that attempted to develop sex identification markers in Asian Palmyra palm. [George et al. \(2007\)](#) presented a male-specific OPA-06₆₀₀ marker after screening 180 RAPD primers using ten male and ten female samples from several populations in India and later used this marker to verify the sexes of more than 100 seedlings ([George & Karun, 2011](#)). [Vinayagam et al. \(2009\)](#) reported another attempt, but could not identify any sex-linked marker after screening up to 130 ISSR markers (with 65 polymorphic bands) based on 20 accessions from another population in India. In this work, we have also tested the OPA-06₆₀₀ marker in Thailand populations, which were collected from the central, northeastern and southern regions, but this marker failed to confirm the sexes in our experiment. The limitation of sex-linked markers across varieties and populations is common ([Heikrujam et al., 2015](#)), and it is most likely due to sequence variation at loci used for developing the markers. Thus, it is required that sex-linked markers for populations in Thailand have to be specifically developed.

The success of identifying molecular markers via DNA fingerprinting may lie upon the genome size and the number of screening PCR primers, and this has been reflected in many reports for developing sex-linked markers (see [Heikrujam et al., 2015](#) for an extensive review on the numbers of primers used in different dioecious plants). Surprisingly, date palm, which is closely related to Asian Palmyra palm ([Barrett et al., 2016](#); [Sakulsathaporn et al., 2017](#)), readily yielded sex-linked markers when tested with only small numbers of screening primers. For example, [Younis, Ismail & Soliman \(2008\)](#) obtained one male-specific and two female-specific markers after screening 30 RAPD primers and five male-specific markers after screening 20 ISSR primers, [Elmeer & Mattat \(2012\)](#) identified 22 microsatellite loci for sex-determination in some date palm cultivars after screening 14 SSR primer pairs, and [Dhawan et al. \(2013\)](#) identified a male-specific marker after screening 100 RAPD primers. Because the genome size of Asian Palmyra palm has not yet been reported, it is difficult to estimate the number of screening primers to cover the genome for effective screening. Considering that our experiment used an extensive set of 1,204 primer pairs among ten techniques and there was only a single sex-linked marker identified in the previous attempts using up to 180 RAPD and 130 SSR primers ([Vinayagam et al., 2009](#); [George & Karun, 2011](#)), this demonstrated a complexity in identifying sex-linked loci in Asian Palmyra palm.

Sex identification of date palm was subjected to a debate as newly developed sex-linked markers via DNA fingerprinting often failed when tested in other varieties or populations ([Rania & Younis, 2008](#); [Al-Mahmoud et al., 2012](#); [Elmeer & Mattat, 2012](#); [Dhawan et al., 2013](#); [Maryam et al., 2016](#)). Indeed, varieties of date palm germplasm and geographical populations were included as much when developing those markers, but the power of sex identification was limited by the variation of DNA sequences at the sex loci and, at that time, an unclear sex determination system in date palm. This problem has recently been solved by the availability of date palm genomes ([Al-Dous et al., 2011](#)), verification of sex determination loci ([Cherif et al., 2013](#)) and construction of genetic map ([Mathew et al., 2014](#)), validating that date palm has a homomorphic XY chromosome system.

Table 5 Transcripts differentially expressed between sexes.

Transcript IDs	Tested samples		FPKM S1	FPKMS2	Fold change (log2)	q value	Blast2GO annotation
	S1	S2					
c129593_g1_i1 ^a	FO	MO	7.700	63.365	3.041	0.0241	auxin-induced 15A-like (<i>Cao et al., 2006</i>)
c42126_g1_i1 ^a	FO	MO	25.601	163.217	2.673	0.0471	glucan endo-1,3-beta-glucosidase 13 (<i>Nagai et al., 1999</i>)
c1819_g1_i1 ^{a,b}	FO	MY	1738.540	99.632	−4.125	0.0383	ethylene-responsive transcription factor 4-like (<i>Liu et al., 2008; Tao et al., 2018</i>)
c142400_g1_i1 ^{a,b}	FO	MY	949.073	68.455	−3.793	0.0087	ethylene-responsive transcription factor ERF017-like (<i>Liu et al., 2008; Tao et al., 2018</i>)
c128418_g1_i1 ^a	FY	MO	6.541	146.348	4.484	0.0151	floricaula leafy homolog (<i>Blázquez et al., 1997</i>)
c119992_g1_i1 ^a	FO	MO	4.834	59.838	3.630	0.0383	galactinol synthase 1-like (<i>Fan et al., 2017</i>)
c138805_g1_i2 ^a	FO	MY	4.267	34.496	3.015	0.0491	argonaute 10 (<i>Nonomura et al., 2007</i>)
c211381_g1_i1	FY	MO	69.478	575.104	3.049	0.0241	anti-muellerian hormone type-2 receptor [<i>Arabidopsis thaliana</i>]
c140985_g1_i1	FY	MY	130.511	20.195	−2.692	0.0409	ASPARTIC PROTEASE IN GUARD CELL 2-like
c195337_g1_i1	FY	MO	44.752	547.920	3.614	0.0151	CASP 2C1
c127228_g1_i1	FY	MO	71.119	6.415	−3.471	0.0328	cellulose synthase A catalytic subunit 4 [UDP-forming]
c131838_g1_i1	FY	MY	20.253	2.104	−3.267	0.0289	cellulose synthase A catalytic subunit 9 [UDP-forming]-like
c145948_g1_i1	FY	MO	26.670	177.451	2.734	0.0448	F-box GID2-like
c124632_g1_i2	FY	MO	3.413	43.070	3.658	0.0383	high mobility group B 7 [<i>Nelumbo nucifera</i>]
c125833_g1_i2	FO	MY	181.228	30.473	−2.572	0.0460	ingression fic1-like
c104934_g1_i1	FY	MY	5.703	70.065	3.619	0.0442	long-chain-alcohol oxidase FAO4A
c138788_g1_i1	FY	MY	483.513	19.217	−4.653	0.0133	NDR1 HIN1-Like 3-like
c106063_g1_i1	FO	MY	458.018	47.014	−3.284	0.0198	nuclease HARBI1
c181538_g1_i1	FO	MY	431.202	34.566	−3.641	0.0087	nuclease HARBI1
c134717_g2_i1	FO	MY	580.064	54.540	−3.411	0.0210	nuclease HARBI1
c126031_g2_i3	FO	MY	142.996	13.803	−3.373	0.0210	nuclease HARBI1
c126031_g2_i2	FO	MY	233.676	27.069	−3.110	0.0380	nuclease HARBI1
c116702_g1_i1	FO	MY	116.085	6.274	−4.210	0.0133	peroxidase 3-like
c129100_g1_i2	FO	MO	2.660	33.248	3.644	0.0375	probable 2-aminoethanethiol dioxygenase
c98445_g1_i1 ^b	FY	MY	172.349	8.685	−4.311	0.0380	probable WRKY transcription factor 25

(continued on next page)

Table 5 (continued)

Transcript IDs	Tested samples		FPKM S1	FPKM S2	Fold change (log2)	q value	Blast2GO annotation
	S1	S2					
c129627_g1_i4 ^b	FY	MO	3.978	64.185	4.012	0.0292	probable WRKY transcription factor 65
c141291_g1_i12	FO	MY	309.255	51.579	−2.584	0.0409	probable xyloglucan endotransglucosylase hydrolase 23
c141291_g1_i9	FO	MY	141.769	24.912	−2.509	0.0500	probable xyloglucan endotransglucosylase hydrolase 23
c128725_g1_i1	FO	MY	149.511	9.942	−3.911	0.0472	probable xyloglucan endotransglucosylase hydrolase 23
c105494_g1_i1	FO	MY	102.967	14.931	−2.786	0.0409	probable xyloglucan endotransglucosylase hydrolase 23
c118421_g1_i2	FY	MY	54.923	2.674	−4.361	0.0241	remorin [<i>Eucalyptus grandis</i>]
c138049_g1_i1	FY	MO	56.348	336.410	2.578	0.0472	signal [<i>Medicago truncatula</i>]
c108722_g1_i1	FY	MO	37.435	654.878	4.129	0.0472	signal [<i>Medicago truncatula</i>]
c135861_g1_i2	FY	MO	38.893	3.975	−3.290	0.0383	trans-resveratrol di-O-methyltransferase-like
c126812_g1_i2 ^b	FO	MY	117.456	12.564	−3.225	0.0133	transcription factor bHLH94-like
c126812_g1_i1 ^b	FO	MY	71.762	5.114	−3.811	0.0241	transcription factor bHLH94-like
c136093_g2_i1 ^b	FO	MY	91.163	8.193	−3.476	0.0133	transcription factor bHLH94-like
c155603_g1_i1	FY	MY	155.137	17.684	−3.133	0.0380	U-box domain-containing 25-like
c82639_g1_i1	FY	MO	87.025	10.838	−3.005	0.0292	U-box domain-containing 26-like
c139774_g4_i1	FY	MY	192.281	28.590	−2.750	0.0472	UPF04964
c135252_g1_i1	FO	MY	40.178	2.638	−3.929	0.0198	VQ motif-containing 4-like
c137964_g3_i1	FY	MO	30.285	441.859	3.867	0.0380	Ycf68 (chloroplast) [<i>Carnegieia gigantea</i>]
c137478_g2_i2	FY	MO	50.030	369.658	2.885	0.0241	ycf68 [<i>Medicago truncatula</i>]

Notes.

Transcript length > 300 bp, FPKM > 5.

^ainvolved in sex determination and flower development.

^bwith high FPKM values.

This sex chromosome is currently being used to develop precise sex-specific markers for specific varieties and populations of date palm (Mohei et al., 2017; Ali et al., 2018; Intha & Chaiprasart, 2018). Thus, it is potential that genome sequencing of Asian Palmyra palm could help us verify the sex determination region and, perhaps, the sex determination system in Asian Palmyra palm.

Potential sex-specific transcripts of Asian Palmyra palm were identified from SSH and *de novo* transcriptome analyses. We have tested many of these sex-specific transcripts for developing sex identification markers by PCR analysis, but none was achieved. Based on a chromosome study by Sharma & Sarkar (1956) proposing the XY chromosomes as the sex determination system of Asian Palmyra palm (XY for male and XX for female), it

can be anticipated that transcripts from X chromosome should be detected in both sexes, while transcripts from the male-specific region of the Y chromosome should only be found in the male. Although most male-specific transcripts identified here are expected to be autosomal genes that support anther development or female sterility, some will be encoded by genes on the male specific region of the Y chromosome (MSY). Those male-specific transcripts can be used as markers for sex and are candidates for sex determination genes. However, PCR analysis of male-specific transcripts so far cannot identify any male-specific marker. Studying Y chromosome of Asian Palmyra palm would be essential for further development on male-specific markers.

Through the exhaustive screening of sex-specific markers using the three approaches performed in this work, this may reveal some aspects of the sex chromosome evolution of this palm species. Firstly, it is possible that the Asian Palmyra palm may have evolutionary young sex chromosomes, which recently diverged, and that the MSY is too small to be identified by the scope of this work. It could be interesting to add Asian Palmyra palm as a specimen for studying the evolution of sex chromosomes ([Charlesworth, 2015](#)). Secondly, although the work by [Sharma & Sarkar \(1956\)](#) had depicted clearly large difference of the sex chromosomes in the shape and size, it is still uncertain whether the XY chromosomes is the sex determination system of this palm as it has not been confirmed at molecular levels. Thirdly, because [George et al. \(2007\)](#) had identified a male-specific marker, which could not be used in the palm population in Thailand, there may be a unique haplotype of the MSY in Thailand. This haplotype may arise from selected individuals during migration to the southeast Asia ([Pipatchartlearnwong et al., 2017a](#)) or the sex chromosomes were recently evolved with small sex determination regions that are difficult to detect.

By comparing the SSH and transcriptome data, to our surprise, most SSH clones were uncorrelated to the sexes, given that the cDNA clones were screened through two rounds of dot blot hybridization. This observation may indicate the limitation of the SSH technique for identifying genes in a complex system. Alternatively, this may be because the RNA samples used for the SSH analysis were obtained from different flower stages from those used for the transcriptome sequencing. A potential weakness of the SSH and transcriptome sequencing was that RNA samples were isolated from developing inflorescence tissues, not from inflorescence primordia that initiate sex organs and that key sex determination transcripts may not be present or difficult to identify in the inflorescence stages ([Harkess et al., 2015](#)). Collecting the primordial tissues for this study is a very challenging task, as the Asian Palmyra palm tip is covered by many layers of thick and hard leaf sheets standing at 20 m height with no indication whether the primordial tissues will develop to be an inflorescence or a leaf, and the plant usually dies after tip removal. Moreover, the variation of sequencing depth among the samples (47–90 million reads per sample) and the lack of replicates could attribute to the complication for extracting robust conclusions from the transcriptome analysis.

Nonetheless, this work provides transcriptome data that would particularly benefit to two research areas: plant sucrose metabolism and sex development in palms and monocots. Sucrose production is one of the important areas in plant biotechnology, and many transcriptome studies have devoted to understand the control of metabolic

flux towards sucrose. Most of the study was performed in major sugar crops including sugarcane ([Cardoso-Silva et al., 2014](#); [Huang et al., 2016](#)), sugar beets ([Mutasa-Göttgens et al., 2012](#)) and sorghum ([Mizuno, Kasuga & Kawahigashi, 2016](#)), and our data add transcript candidates for sucrose metabolism in the inflorescence of palm species. For sex development, male- and female-specific transcripts and differentially expressed transcripts between sexes were listed to provide data for further study in sex determination and male and female inflorescence development. A number of studies with similar objectives in identifying sex determination genes and underlying mechanisms of sex development in dioecious plants via the transcriptome sequencing have been reported in asparagus ([Harkess et al., 2015](#); [Li et al., 2017](#)), *Idesia polycarpa* ([Mei et al., 2017](#)), shrub willows (*Salix suchowensis*) ([Liu et al., 2013](#)), poplar ([Song et al., 2013](#)) and *Coccinia grandis* ([Mohanty et al., 2017](#)). Most of the identified genes belong to floral development, phytohormone biosynthesis, hormone signaling and transduction, transcriptional regulation and DNA methyltransferase activity. However, with many genes playing the role during the complex developmental process, it is difficult to determine the mechanism underlying the sexual development and sex determination in dioecious species. Future progress in functional genomics addressing the identified genes would be an essential tool to solve this long-standing question.

CONCLUSIONS

Although no sex-linked marker was obtained from exhaustive DNA fingerprinting, SSH and *de novo* transcriptome analysis, this work provides transcripts based on male and female inflorescences of the Asian Palmyra palm. Further attempts on developing sex identification markers in Asian Palmyra palm should be directed towards genomic-based approaches, particularly at the MSY. Genome analysis using SNPs have been successful in accessing sex determination loci in date palm ([Al-Mahmoud et al., 2012](#); [Ali et al., 2018](#)) and other dioecious plant species ([Zhou et al., 2018](#); [Jia et al., 2019](#)). Whole genome sequencing, genetic mapping, SNPs and genome-wide association study (GWAS) between male and female populations would be essential tools for further identification of sex-linked loci in Asian Palmyra palm.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was financially supported by Kasetsart University Research and Development Institute (KURDI), Faculty of Science Research Fund (ScRF) and OmiKU. Supachai Vuttipongchaikij was supported by Thailand Research Fund (TRF-RSA6080031). Kwanjai Pipatchartlearnwong was supported by the PhD studentship from Kasetsart University and Faculty of Science. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

Kasetsart University Research and Development Institute (KURDI).
 Faculty of Science Research Fund (ScRF).
 Thailand Research Fund: TRF-RSA6080031.
 Kasetsart University and Faculty of Science.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Kwanjai Pipatchartlearnwong performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Piyada Juntawong and Passorn Wonnapijit performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables, approved the final draft.
- Somsak Apisitwanich conceived and designed the experiments, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.
- Supachai Vuttipongchaikij conceived and designed the experiments, analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.

Field Study Permissions

The following information was supplied relating to field study approvals (i.e., approving body and any reference numbers):

Sample collections were conducted on privately-owned lands.

Data Availability

The following information was supplied regarding data availability:

The sequence of SSH clones are available in GenBank ([JZ977504–JZ977592](#)). The Transcriptome Shotgun Assembly are available at DDBJ/ENA/GenBank (accession [GFYQ00000000](#)). Other raw data are available in [Figs. S1–S4](#) and [Tables S1–S15](#).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.7268#supplemental-information>.

REFERENCES

- Al-Dous EK, George B, Al-Mahmoud ME, Al-Jaber MY, Wang H, Salameh YM, Al-Azwani EK, Chaluvadi S, Pontaroli AC, DeBarry J, Arondel V. 2011. De novo genome sequencing and comparative genomics of date palm (*Phoenix dactylifera*). *Nature Biotechnology* **29**(6):521–527 DOI [10.1038/nbt.1860](#).
- Al-Mahmoud ME, Al-Dous EK, Al-Azwani EK, Malek JA. 2012. DNA-based assays to distinguish date palm (Arecaceae) gender. *American Journal of Botany* **99**(1):e7–e10 DOI [10.3732/ajb.1100425](#).

- Ali HBM, Abubakari A, Wiehle M, Krutovsky KV. 2018. Gene-specific sex-linked genetic markers in date palm (*Phoenix dactylifera* L.). *Genetic Resources and Crop Evolution* 65(1):1–10 DOI 10.1007/s10722-017-0564-7.
- Arabnezhad H, Bahar M, Mohammadi HR, Latifian M. 2012. Development, characterization and use of microsatellite markers for germplasm analysis in date palm (*Phoenix dactylifera* L.). *Scientia Horticulturae* 134:150–156 DOI 10.1016/j.scienta.2011.11.032.
- Barfod AS, Balhara M, Dransfield J, Balslev H. 2015. SE Asian palms for agroforestry and home gardens. *Forests* 6(12):4607–4616 DOI 10.3390/f6124389.
- Barrett CF, Baker WJ, Comer JR, Conran JG, Lahmeyer SC, Leebens-Mack JH, Li J, Lim GS, Mayfield-Jones DR, Perez L, Medina J. 2016. Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. *New Phytologist* 209(2):855–870 DOI 10.1111/nph.13617.
- Beck HT, Balick MJ. 1990. Useful palms of the world. New York: Columbia University Press.
- Billotte N, Marseillac N, Brottier P, Noyer JL, Jacquemoud-Collet JP, Moreau C, Couvreur T, Chevallier MH, Pintaud JC, Risterucci AM. 2004. Nuclear microsatellite markers for the date palm (*Phoenix dactylifera* L.): characterization and utility across the genus *Phoenix* and in other palm genera. *Molecular Ecology Notes* 4(2):256–258 DOI 10.1111/j.1471-8286.2004.00634.x.
- Blázquez MA, Soowal LN, Lee I, Weigel D. 1997. LEAFY expression and flower initiation in *Arabidopsis*. *Development* 124(19):3835–3844.
- Bureau TE, Wessler SR. 1994. Mobile inverted-repeat elements of the Tourist family are associated with the genes of many cereal grasses. *Proceedings of the National Academy of Sciences of the United States of America* 91(4):1411–1415 DOI 10.1073/pnas.91.4.1411.
- Cao D, Cheng H, Wu W, Soo HM, Peng J. 2006. Gibberellin mobilizes distinct DELLA-dependent transcriptomes to regulate seed germination and floral development in *Arabidopsis*. *Plant Physiology* 142(2):509–525 DOI 10.1104/pp.106.082289.
- Cardoso-Silva CB, Costa EA, Mancini MC, Balsalobre TWA, Canesin LEC, Pinto LR, Carneiro MS, Garcia AAF, De Souza AP, Vicentini R. 2014. *De novo* assembly and transcriptome analysis of contrasting sugarcane varieties. *PLOS ONE* 9(2):e88462 DOI 10.1371/journal.pone.0088462.
- Charlesworth D. 2015. Plant contributions to our understanding of sex chromosome evolution. *New Phytologist* 208(1):52–65 DOI 10.1111/nph.13497.
- Cherif E, Zehdi S, Castillo K, Chabrillange N, Abdoukader S, Pintaud JC, Santoni S, Salhi-Hannachi A, Glémin S, Aberlenc-Bertossi F. 2013. Male-specific DNA markers provide genetic evidence of an XY chromosome system, a recombination arrest and allow the tracing of paternal lineages in date palm. *New Phytologist* 197(2):409–415 DOI 10.1111/nph.12069.
- Collard BC, Mackill DJ. 2009. Start codon targeted (SCoT) polymorphism: a simple, novel DNA marker technique for generating gene-targeted markers in plants. *Plant Molecular Biology Reporter* 27:86–93 DOI 10.1007/s11105-008-0060-5.

- Davis TA, Johnson DV. 1987. Current utilization and further development of the palmyra palm (*Borassus flabellifer* L. Arecaceae) in Tamil Nadu State, India. *Economic Botany* 41(2):247–266 DOI 10.1007/BF02858972.
- Dhawan C, Kharb P, Sharma R, Uppal S, Aggarwal RK. 2013. Development of male-specific SCAR marker in date palm (*Phoenix dactylifera* L.). *Tree Genetics & Genomes* 9(5):1143–1150 DOI 10.1007/s11295-013-0617-9.
- Dransfield J, Uhl NW, Asmussen CB, Baker WJ, Harley MM, Lewis CE. 2005. A new phylogenetic classification of the palm family, Arecaceae. *Kew Bulletin* 60(4):559–569.
- Elmeer K, Mattat I. 2012. Marker-assisted sex differentiation in date palm using simple sequence repeats. 3 *Biotech* 2(3):241–247 DOI 10.1007/s13205-012-0052-x.
- Elmeer K, Sarwath H, Malek J, Baum M, Hamwieh A. 2011. New microsatellite markers for assessment of genetic diversity in date palm (*Phoenix Dactylifera*). 3 *Biotech* 1(2):91–97 DOI 10.1007/s13205-011-0010-z.
- Fan Y, Yu M, Liu M, Zhang R, Sun W, Qian M, Duan H, Chang W, Ma J, Qu C, Zhang K. 2017. Genome-wide identification, evolutionary and expression analyses of the GALACTINOL SYNTHASE gene family in rapeseed and tobacco. *International Journal of Molecular Sciences* 18:2768–2788 DOI 10.3390/ijms18122768.
- Ferguson W. 1888. Description of the Palmyra palm of Ceylon. Colombo: Observer Press.
- Fox JJ. 1977. Harvest of the palm. In: *Ecological change in eastern Indonesia*. Cambridge: Harvard Univ. Press.
- George J, Karun A. 2011. Marker assisted detection of seed sex ratio in palmyrah palm (*Borassus flabellifer* L.). *Current Science* 100(6):922–925.
- George J, Karun A, Manimekalai R, Rajesh MK, Remya P. 2007. Identification of RAPD markers linked to sex determination in palmyrah (*Borassus flabellifer* L.). *Current Science* 93(8):1075–1077.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29(7):644–652 DOI 10.1038/nbt.1883.
- Harkess A, Mercati F, Shan HY, Sunseri F, Falavigna A, Leebens-Mack J. 2015. Sex-biased gene expression in dioecious garden asparagus (*Asparagus officinalis*). *New Phytologist* 207(3):883–892 DOI 10.1111/nph.13389.
- Heikrujam M, Sharma K, Prasad M, Agrawal V. 2015. Review on different mechanisms of sex determination and sex-linked molecular markers in dioecious crops: a current update. *Euphytica* 201(2):161–194 DOI 10.1007/s10681-014-1293-z.
- Huang DL, Gao YJ, Gui YY, Chen ZL, Qin CX, Wang M, Liao Q, Yang LT, Li YR. 2016. Transcriptome of High-sucrose sugarcane variety GT35. *Sugar Tech* 18(5):520–528 DOI 10.1007/s12355-015-0420-z.
- Intha N, Chaiprasart P. 2018. Sex determination in date palm (*Phoenix dactylifera* L.) by PCR based marker analysis. *Scientia Horticulturae* 236:251–255 DOI 10.1016/j.scienta.2018.03.039.


- Jia HM, Jia HJ, Cai QL, Wang Y, Zhao HB, Yang WF, Wang GY, Li YH, Zhan DL, Shen YT, Niu QF. 2019. The red bayberry genome and genetic basis of sex determination. *Plant Biotechnology Journal* 17(2):397–409 DOI 10.1111/pbi.12985.
- Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* 28(1):27–30 DOI 10.1093/nar/28.1.27.
- Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology* 14(4):Article R36 DOI 10.1186/gb-2013-14-4-r36.
- Kovoor A. 1983. The palmyrah palm: potential and perspectives (No. 52). In: *FAO plant production and protection paper 52*. Rome: Food and Agriculture Organization.
- Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12(1):323 DOI 10.1186/1471-2105-12-323.
- Li SF, Zhang GJ, Zhang XJ, Yuan JH, Deng CL, Gao WJ. 2017. Comparative transcriptome analysis reveals differentially expressed genes associated with sex expression in garden asparagus (*Asparagus officinalis*). *BMC Plant Biology* 17(1):143 DOI 10.1186/s12870-017-1091-6.
- Lim TK. 2012. *Edible medicinal and non-medicinal plants* (Vol. 1, 656-687). Dordrecht: Springer.
- Liu J, Yin T, Ye N, Chen Y, Yin T, Liu M, Hassani D. 2013. Transcriptome analysis of the differentially expressed genes in the male and female shrub willows (*Salix suchowensis*). *PLOS ONE* 8(4):e60181 DOI 10.1371/journal.pone.0060181.
- Liu S, Xu L, Jia Z, Xu Y, Yang Q, Fei Z, Lu X, Chen H, Huang S. 2008. Genetic association of ETHYLENE-INSENSITIVE3-like sequence with the sex-determining M locus in cucumber (*Cucumis sativus* L.). *Theoretical and Applied Genetics* 117:927–933 DOI 10.1007/s00122-008-0832-1.
- Maryam JM, Awan FS, Ahmad S, Khan IA. 2016. Development of molecular method for sex identification in date palm (*Phoenix dactylifera* L.) plantlets using novel sex-linked microsatellite markers. *3 Biotech* 6(1):22 DOI 10.1007/s13205-015-0321-6.
- Mathew LS, Spannagl M, Al-Malki A, George B, Torres MF, Al-Dous EK, Al-Azwani EK, Hussein E, Mathew S, Mayer KF, Mohamoud YA. 2014. A first genetic map of date palm (*Phoenix dactylifera*) reveals long-range genome structure conservation in the palms. *BMC Genomics* 15(1):285 DOI 10.1186/1471-2164-15-285.
- Mei L, Dong N, Li F, Li N, Yao M, Chen F, Tang L. 2017. Transcriptome analysis of female and male flower buds of *Idesia polycarpa* Maxim. var. *vestita* Diels. *Electronic Journal of Biotechnology* 29:39–46 DOI 10.1016/j.ejbt.2017.07.002.
- Mizuno H, Kasuga S, Kawahigashi H. 2016. The sorghum SWEET gene family: stem sucrose accumulation as revealed through transcriptome profiling. *Biotechnology for Biofuels* 9(1):Article 127 DOI 10.1186/s13068-016-0546-6.
- Mohanty JN, Nayak S, Jha S, Joshi RK. 2017. Transcriptome profiling of the floral buds and discovery of genes related to sex-differentiation in the dioecious cucurbit *Coccinia grandis* (L.) Voigt. *Gene* 626:395–406 DOI 10.1016/j.gene.2017.05.058.

- Mohei EL, Mohasseb HAA, Al-Khateeb AA, Al-Khateeb SA, Chowdhury K, El-Shemy HA, Aldaej MI. 2017. Identification and sequencing of Date-SRY Gene: A novel tool for sex determination of date palm (*Phoenix dactylifera* L.). *Saudi Journal of Biological Sciences* 26(3):514–523 DOI 10.1016/j.sjbs.2017.08.002.
- Morton JF. 1988. Notes on distribution, propagation, and products of *Borassus* Palms (Arecaceae). *Economic Botany* 42(3):420–441 DOI 10.1007/BF02860166.
- Mutasa-Göttgens ES, Joshi A, Holmes HF, Hedden P, Göttgens B. 2012. A new RNASeq-based reference transcriptome for sugar beet and its application in transcriptome-scale analysis of vernalization and gibberellin responses. *BMC Genomics* 13(1):99 DOI 10.1186/1471-2164-13-99.
- Nagai JI, Yamato KT, Sakaida M, Yoda H, Fukuzawa H, Ohyama K. 1999. Expressed sequence tags from immature female sexual organ of a liverwort, Marchantia polymorpha. *DNA Research* 6(1):1–11 DOI 10.1093/dnares/6.1.1.
- Nonomura KI, Morohoshi A, Nakano M, Eiguchi M, Miyao A, Hirochika H, Kurata N. 2007. A germ cell-specific gene of the ARGONAUTE family is essential for the progression of premeiotic mitosis and meiosis during sporogenesis in rice. *The Plant Cell* 19(8):2583–2594 DOI 10.1105/tpc.107.053199.
- Pipatchartlearnwong K, Swatdipong A, Vuttipongchaikij S, Apisitwanich S. 2017a. Genetic evidence of multiple invasions and a small number of founders of Asian Palmyra palm (*Borassus flabellifer*) in Thailand. *BMC Genetics* 18(1):88 DOI 10.1186/s12863-017-0554-y.
- Pipatchartlearnwong K, Swatdipong A, Vuttipongchaikij S, Apisitwanich S. 2017b. Cross-genera transferability of microsatellite loci for Asian Palmyra Palm (*Borassus flabellifer* L.). *HortScience* 52(9):1164–1167 DOI 10.21273/HORTSCI12175-17.
- Rania AA, Younis O. 2008. Identification of Sex Specific DNA Markers for Date Palm (*Phoenix dactylifera* L.) Using RAPD and ISSR. *Research Journal of Agriculture and Biological Sciences* 4(4):278–284.
- Rebrikov DV, Desai SM, Siebert PD, Lukyanov SA. 2004. Suppression subtractive hybridization. In: *Gene expression profiling*. Totowa: Humana press, 107–134.
- Sakulsathaporn A, Wonnapijit P, Vuttipongchaikij S, Apisitwanich S. 2017. The complete chloroplast genome sequence of Asian Palmyra palm (*Borassus flabellifer*). *BMC Research Notes* 10(1):740 DOI 10.1186/s13104-017-3077-8.
- Sharma AK, Sarkar SK. 1956. Veratrine: its use in Cytochemistry: (with Plate XIX). *Caryologia* 8(2):240–249 DOI 10.1080/00087114.1956.10797563.
- Sirajuddin S, Mulyadi M, Dirawan GD, Amir F, Pertiwi N. 2016. Conservation status of lontar palm trees (*Borassus flabellifer* Linn) in Jeneponto district, South Sulawesi, Indonesia. *Journal of Tropical Crop Science* 3(1):28–33.
- Song Y, Ma K, Ci D, Chen Q, Tian J, Zhang D. 2013. Sexual dimorphic floral development in dioecious plants revealed by transcriptome, phytohormone, and DNA methylation analysis in *Populus tomentosa*. *Plant Molecular Biology* 83(6):559–576 DOI 10.1007/s11103-013-0108-2.

- Takata M, Kishima Y, Sano Y. 2005.** DNA methylation polymorphisms in rice and wild rice strains: detection of epigenetic markers. *Breeding Science* **55**(1):57–63 DOI [10.1270/jsbbs.55.57](https://doi.org/10.1270/jsbbs.55.57).
- Tao Q, Niu H, Wang Z, Zhang W, Wang H, Wang S, Zhang X, Li Z. 2018.** Ethylene responsive factor ERF110 mediates ethylene-regulated transcription of a sex determination-related orthologous gene in two Cucumis species. *Journal of Experimental Botany* **69**(12):2953–2965 DOI [10.1093/jxb/ery128](https://doi.org/10.1093/jxb/ery128).
- Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, Van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010.** Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology* **28**(5):511–515 DOI [10.1038/nbt.1621](https://doi.org/10.1038/nbt.1621).
- Ukoskit K, U-thoomporn S. 2016.** Development of ILP markers from genes involving in oil palm (*Elaeis guineensis* Jacq.) inflorescence. *Thai Science and Technology Journal* **24**:299–308 DOI [10.14456/tstj.2016.23](https://doi.org/10.14456/tstj.2016.23).
- Vinayagam P, Dhandapani J, Raman P, Alagarsamy RK, Muthusamy P, Balaraman SS. 2009.** Molecular characterization for genetic diversity of Palmyrah (*Borassus flabellifer*) accessions using inter simple sequence repeat (ISSR) markers. *The Asian and Australasian Journal of Plant Science and Biotechnology* **3**(1):11–5.
- Younis RA, Ismail OM, Soliman SS. 2008.** Identification of sex-specific DNA markers for date palm (*Phoenix dactylifera* L.) using RAPD and ISSR techniques. *Research Journal of Agriculture and Biological Sciences* **4**(4):278–284.
- Zhou R, Macaya-Sanz D, Rodgers-Melnick E, Carlson CH, Gouker FE, Evans LM, Schmutz J, Jenkins JW, Yan J, Tuskan GA, Smart LB. 2018.** Characterization of a large sex determination region in *Salix purpurea* L. (Salicaceae). *Molecular Genetics and Genomics* **293**(6):1437–1452 DOI [10.1007/s00438-018-1473-y](https://doi.org/10.1007/s00438-018-1473-y).



RNA editing in the chloroplast of Asian Palmyra palm (*Borassus flabellifer*)

Arpakorn Sakulsathaporn^{1,2,3}, Passorn Wonnapijit^{4,5,6}, Anongpat Suttangkakul^{4,5}, Somsak Apisitwanich^{1,2,4,5} and Supachai Vuttipongchaikij^{4,5,6} 

¹Center for Agricultural Biotechnology, Kasetsart University, Kamphaeng Saen Campus, Nakhon Pathom 73140, Thailand.

²Center of Excellence on Agricultural Biotechnology: (AG-BIO/PERDO-CHE), Bangkok 10900, Thailand.

³School of Natural Resource and Environmental Management, Faculty of Applied Science and Engineering, Khon Kaen University, Nong Khai Campus, Nong Khai 43000, Thailand.

⁴Department of Genetics, Faculty of Science, Kasetsart University, 50 Ngarm Wong Wan road, Chatuchak, Bangkok 10900, Thailand.

⁵Center of Advanced studies for Tropical Natural Resources, Kasetsart University, Ngam Wong Wan, Chatuchak, Bangkok 10900, Thailand.

⁶Omics Center for Agriculture, Bioresources, Food and Health, Kasetsart University (OmiKU), 50 Ngarm Wong Wan road, Chatuchak, Bangkok 10900, Thailand.

Abstract

We have identified 46 RNA editing sites located in 20 chloroplast (cp) genes of *Borassus flabellifer* (Asian Palmyra palm), family Arecaceae, and tested these genes for supporting phylogenetic study among the commelinids. Among the 46 sites, 43 sites were found to cause amino acid alterations, which were predicted to increase the hydrophobicity and transmembrane regions of the proteins, and one site was to cause a premature stop codon. Analysis of these editing sites with data obtained from seed plants showed that a number of shared-editing sites depend on the evolutionary relationship between plants. We reconstructed a deep phylogenetic relationship among the commelinids using seven RNA edited genes that are orthologous among monocots. This tree could represent the relationship among subfamilies of Arecaceae family, but was insufficient to represent the relationship among the orders of the commelinid. After adding eight gene sequences with high parsimony-informative characters (PICs), the tree topology was improved and could support the topology for the commelinid orders ((Arecales, Dasypogonaceae) (Zingiberales+Commelinales, Poales)). The result provides support for inherent RNA editing along the evolution of seed plants, and we provide an alternative set of loci for the phylogenetic tree reconstruction of Arecaceae's subfamilies.

Keywords: Arecaceae; Arecales; chloroplast genome; Commelinids; Phylogeny.

Received: December 06, 2018; Accepted: October 02, 2019.

Introduction

RNA editing, a crucial post-transcriptional RNA modification process, allows changes in the genetic information on the primary transcripts in the plastids and mitochondria in many living organisms. Base modifications on the primary transcripts often result in amino acid changes and affect the functionality of the protein products (Tillich *et al.*, 2006). This process has been proposed as an evolutionary mean to restore the original sequence of amino acids of genes that possess mutations (Castandet and Araya, 2011). Generally, 26-64 RNA editing sites are observed in the chloroplast of seed plants (Wakasugi *et al.*, 1996;

Huang *et al.*, 2013), and more than a hundred sites have been found in some species such as hornwort (*Anthoceros formosa*), fern (*Adiantum capillus-veneris*) and lycophyte (*Isoetes engelmannii*). So far, only liverwort (*Marchantia polymorpha*) was found to contain no RNA editing in its plastid genome, and this species is currently used as a baseline for comparing RNA editing in plants (Kugita *et al.*, 2003; Wolf *et al.*, 2004). Besides, RNA editing was found in the plastid of dinoflagellates (Dorrell *et al.*, 2016), but it has not yet been observed in any green microalgae (Cha-teigner-Boutin and Small, 2011).

Commelinid clade, a group of monocots is categorized by the presence of ferulate/coumarate in the primary cell wall (Chase *et al.*, 1993; Peña *et al.*, 2016). The clade is composed of monophyletic groups of four orders: Arecales (A), Zingiberales (Z), Commelinales (C), Poales (P), and one family named Dasypogonaceae (D). Molecular classi-

Send correspondence to Supachai Vuttipongchaikij Department of Genetics, Faculty of Science, Kasetsart University, 50 Ngarm Wong Wan road, Chatuchak, Bangkok 10900, Thailand. E-mail: fsciscv@ku.ac.th.

fication of this clade has long been problematic with conflict topologies. The phylogenetic tree built based on plastome data of 83 taxa showed a monophyletic group of ((P,A)(D,ZC)); however, the relationship within sub-clade of Poales and Arecales, as well as a subclade of (Zingiberales + Commelinales) and Dasypogonaceae, were not strongly supported by bootstrap samplings (Givnish *et al.*, 2010). The topology conflict of the commelinid clade is currently being resolved by dense taxon sampling and availability of plastome sequences. The phylogenetic tree reconstructed from plastome sequences presents that the commelinid clade is comprised of two subclades: a monophyletic clade of Poales being sister to Commelinales+Zingiberales and a monophyletic group of Arecales being sister to Dasypogonaceae ((A,D)(ZC, P)). These subclades are strongly supported by bootstrap analysis (Barrett *et al.*, 2013; Barrett *et al.*, 2016).

Borassus flabellifer or Asian Palmyra palm, a member of Arecaceae family, is widespread in Southeast Asia (Pipatchartlearnwong *et al.*, 2017). It is grouped in Coryphoideae subfamily with Phoenix dactylifera (date palm). The economic utilities of *B. flabellifer* are manifold including building woods, ornamentals, fruits and flower sap for palm sugar production. Currently, only few information is available regarding molecular genetics of *B. flabellifer*. In this work, we sequenced some chloroplast (cp) genes that have been reported to carry RNA editing sites, deduced the pattern of the RNA editing and verified these DNA sequences whether they could support the evolutionary relationship among palm species as well as a deep phylogeny of the commelinids. The results of this study provide insight into the evolution of RNA editing and the evolutionary relationship among subfamilies of Arecaceae family.

Material and Methods

RNA isolation, cDNA synthesis and sequencing

Total RNA of *B. flabellifer* was isolated from unexpanded green fan-shaped leaves using Spectrum™ Plant Total RNA kit (Sigma-Aldrich, USA). cDNA was synthesized using Protoscrip M-MuLV First Strand cDNA kit (NEB, UK) with random hexamer-primers. One microgram of RNA was used for each reaction as recommended in the kit.

Thirty-four Arecaceae cp genes, which were previously reported for carrying RNA editing sites in other plant cp genomes, were chosen for this study (Kugita *et al.*, 2003; Junior *et al.*, 2004; Zeng *et al.*, 2007; Chen *et al.*, 2011; Huang *et al.*, 2013; Uthaipaisanwang *et al.*, 2012). These were *accD*, *atpA*, *atpB*, *atpF*, *atpI*, *clpP*, *matK*, *ndhA*, *ndhB*, *ndhD*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhK*, *petA*, *petB*, *psaB*, *psaJ*, *psbC*, *psbH*, *rpl2*, *rpl14*, *rpl20*, *rpl22*, *rpl23*, *rpoA*, *rpoC1*, *rps3*, *rps7*, *rps8*, *rps16*, *ycf2*, *ycf3* and *ycf4*. These 35 transcripts were amplified from the cDNA using PrimeSTAR® GXL DNA Polymerase (Takara-Bio Inc., Japan) with PCR conditions: initial denaturation for 2 min

at 98°C, followed by 35 cycles of 15 s at 98°C, 30 s at 55°C or 60°C depending on melting temperatures of primers, 60 s at 68°C and final extension for 5 min at 68°C (see primer sequences in Table S1). PCR products were resolved by 1% agarose gel electrophoresis and purified using a gel extraction kit (Flavorgen, Taiwan) before Sanger's sequencing (Macrogen, Korea). The transcription sequences were then aligned against predicted RNA editing and cpDNA (GenBank accession number KP_901247; (Sakulsathaporn *et al.*, 2017) using Clustal Omega.

Analysis of RNA editing sites

The editing sites of the 35 cp genes of *B. flabellifer* were predicted using PREP-Cp (<http://prep.unl.edu/cgi-bin/cp-input.pl>) with 0.8 cutoff value and PREPACT 3.0 (<http://www.prepact.de/prepact-main.php>; Lenz *et al.*, 2018). The actual RNA editing sites of these 35 genes were obtained by sequence alignments between the transcripts and the cp DNA sequence (NCBI accession KP_901247) using Clustal Omega.

Protein structure prediction

Non-synonymous mutation sites and consequent amino acid changes were manually analyzed. The protein secondary structures derived from the edited and non-edited coding sequences were predicted by using SOPMA (<http://nhjy.hzau.edu.cn/kech/swxxx/jakj/dianzi/Bioinf7/Expasy/Expasy8.htm>). The proportion of alpha helix, extended strand, beta turn and random coil structures between protein derived from edited and non-edited coding sequences were manually counted. The transmembrane structures and signal peptides were predicted by Phobius program (<http://phobius.sbc.su.se/instructions.html>).

Comparison of RNA editing sites among plant species

RNA editing sites obtained from *B. flabellifer* (this study), *Cocos nucifera* (identified by Next Generations Sequencing or NGS; Huang *et al.*, 2013) and *Elaeis guineensis* (by NGS; Uthaipaisanwang *et al.*, 2012), in total 53 sites, were compared to those of other 15 plant species. These included six monocots: *S. polyrhiza* (21 sites by NGS), *P. aphrodite* (13 sites by direct cDNA sequencing), *Z. mays* (13 sites by direct cDNA sequencing), *O. sativa* (14 sites by direct cDNA sequencing), *S. officinarum* (12 sites by NGS) and *H. vulgare* (8 sites by direct cDNA sequencing), six dicots: *A. thaliana* (10 sites by direct cDNA sequencing), *P. sativum* (7 sites by direct cDNA sequencing), *A. belladonna* (8 sites by direct cDNA sequencing), *H. niger* (8 sites by NGS), *J. curcas* (7 sites by NGS), *N. tabacum* (10 sites by direct cDNA sequencing) and two gymnosperms: *C. taitungensis* (5 sites by using CURE-Chloroplast software and direct sequencing) and *Pinus thunbergii* (0 sites by direct cDNA sequencing) (Wakasugi *et al.*, 1996; Corneille *et al.*, 2000; Sasaki *et al.*, 2003; Inada

et al., 2004; Zeng *et al.*, 2007; Asif *et al.*, 2010; Uthai-paisanwong *et al.*, 2012; Huang *et al.*, 2013; Sanchez-Puerta and Abbona, 2014; Wang *et al.*, 2015; He *et al.*, 2016). The list of genes is presented in Table S2.

Phylogenetic tree reconstruction

Phylogenetic analyses were carried out using two datasets: (1) DNA sequences of the cp genes carrying RNA editing sites that are shared among monocots and orthologs in *Marchantia* (7 genes: *accD*, *atpB*, *atpI*, *ndhB*, *ndhD*, *ndhF* and *rps8*) and (2) DNA sequences of the dataset (1), and additional eight cp genes (*matK*, *rpoA*, *ndhA*, *rpoC1*, *yef1*, *rpoC2*, *ccsA* and *atpA*) with PIC/length ratio above 0.3. All DNA sequences of cp genes were retrieved from GenBank with accession numbers presented in Table S3. Multiple sequence alignment was performed for each gene using MACSE (Ranwez *et al.*, 2011), and aligned sequences were concatenated using SequenceMatrix (Vaidya *et al.*, 2011). The phylogenetic trees were reconstructed based on maximum likelihood (ML) and Bayesian (Bayes) methods with GTR+I+G model using RAxMLGUI (Silvestro and Michalak, 2012) and MrBayes (Ronquist *et al.*, 2012), respectively. This evolutionary model was determined by jModelTest based on AICc value (Darriba *et al.*, 2012). Ten-thousand replicates and 10M generations with stop value at 0.01 were applied to calculate statistical supports for maximum likelihood and Bayesian phylogeny, respectively. *Doryanthes palmeri* (Asparagales) was used as an outgroup.

Results

RNA editing sites in the cp genome of *B. flabellifer*

Based on previous reports on RNA editing in the transcripts of 35 chloroplast genes in monocots, we aimed to analyze the transcripts of these genes in the cp genome of *B. flabellifer*. PREP-Cp program predicted 67 potential editing sites located in 19 out of the 35 target genes, while PREPACT 3.0 predicted 57 editing sites in 23 genes. We sequenced the transcripts of these 35 genes, aligned them against the *B. flabellifer* cp DNA (NCBI accession KP_901247) and found that there were in total 46 RNA editing sites located in 20 genes (Table S4). Although only 32 sites out of 67 predicted sites were correct (47.76%), it is important to note that the PREP-Cp was unable to predict editing sites at the third base of the codon. The numbers of the observed editing sites per gene were varied from one to several, and most frequently edited genes were *ndhB*, *ndhD* and *rpoC1*, with 11, 4 and 4 editing sites, respectively. All of the 46 editing sites were with C-to-U editing type. Examination of the adjacent sequences of (*_C_*) editing sites revealed that the frequency of nucleotides preceding the C editing sites were U (65.21%), C (23.91%), A (6.52%) and G (4.34%), and that of nucleotides following the editing sites were A (73.91%), G (10.67%), U (8.70%) and C

(5.52%). Thus, U_A is the highest context of RNA editing sites (52.17%) (Figure 1a).

Analysis of the editing sites based on base positions within codons showed that seven (15.21%), 37 (80.43%) and two (4.36%) editing sites were located at the first, second and third base, respectively. Follow-up analysis of the edited coding sequences showed that 43 out of 46 editing sites caused amino acid changes, one site generated a premature stop codon and the other two were silent mutations. The amino acid changes were found to be preferably converting neutral (26 sites) and hydrophilic (16 sites) amino acids to hydrophobic amino acids (Figure 1b). The most frequent amino acid alteration was from serine to leucine, the neutral to the hydrophobic amino acid, followed by alteration from proline to leucine, the hydrophilic to hydrophobic amino acid.

Impacts of RNA editing on polypeptides and protein structures

To evaluate the impact of the RNA editing in *B. flabellifer* on the protein structures, we analyzed the modified protein sequences using SOPMA program. Analysis of the predicted secondary structures showed that the amino acid changes could result in the increase of the alpha helix structure, extended strand and beta-turn or random coil (Figure 2a). In particular, the RNA editing of *ndhA*, *ndhB*, *ndhD*, *ndhF*, *ndhG*, *ndhI* and *ndhK* genes showed increases

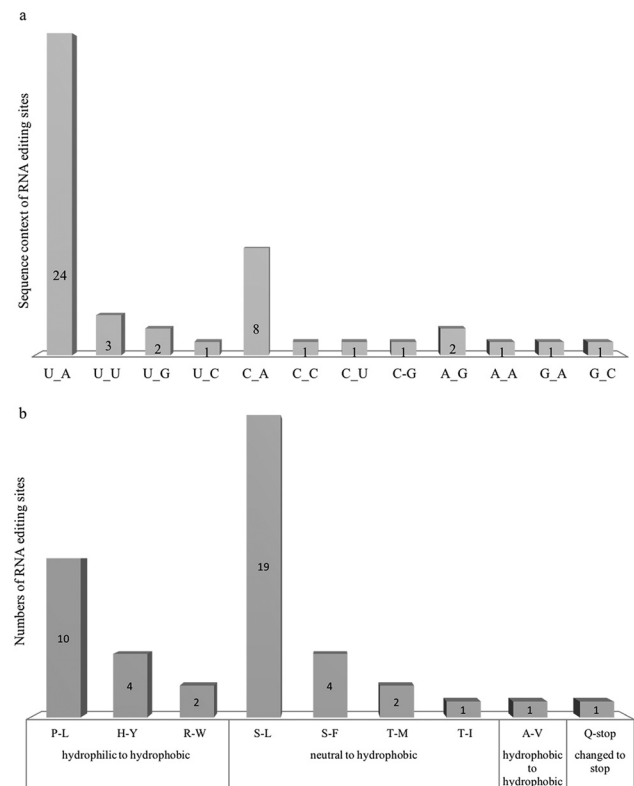


Figure 1 - The frequency of sequence contexts of RNA editing sites (a) and amino acid changes caused by RNA editing (b) in the chloroplast of *B. flabellifer*.

in the alpha helix of their encoded protein: NADH-plastoquinone oxidoreductase. As the observed RNA editing mainly resulted in hydrophobic amino acid and changes in the hydrophobicity have been suggested to affect the transmembrane properties of proteins (Kugita *et al.*, 2003; Wang *et al.*, 2015; He *et al.*, 2016), we compared the predicted transmembrane and signal peptide regions between the edited and non-edited versions of NdhA, NdhB, NdhD, NdhF, NdhG, NdhI and NdhK proteins. While there was no alteration of the signal peptide regions, expansions of transmembrane regions were observed in many regions of these proteins, particularly NdhA, NdhB, and NdhK (Figure 2b). Especially, amino acid changes at ndhA_S159L and ndhA_S189L were found to generate two new transmembrane regions between the amino acid positions 159 and 169 and the positions 182 and 208, respectively. Moreover, NdhB and NdhK proteins resulted from RNA editing including ndhB_50SL, ndhB_156PL, ndhB_181TM, ndhB_277SL and ndhK_44SL gained new transmembrane regions between codon 35 and 50, 155 and 181, 178 and 182, 270 and 283, and 43 and 51, respectively.

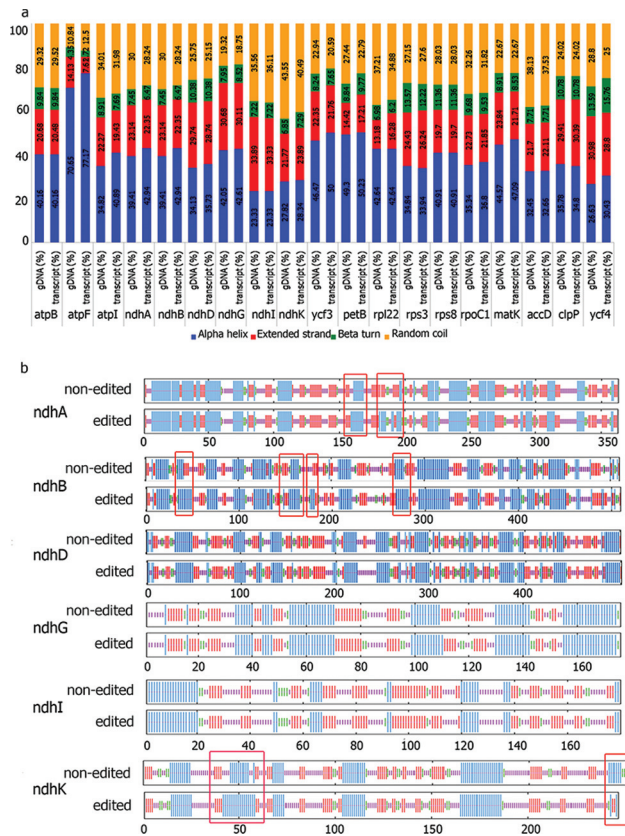


Figure 2 - Comparison of protein structure before and after editing. (a) The proportion of protein secondary structures (in percentages) derived from gDNA sequence and transcript sequences. Blue: alpha-helix, red: extended strand, green: beta-turn, orange: random coil. (b) Patterns of transmembrane regions in NDH proteins compared between non-edited and edited sequences. Squares indicate altered transmembrane regions.

RNA editing sites among *B. flabellifer* and other plant species

The RNA editing sites and resulted amino acid alterations observed in *B. flabellifer* were compared to those previously observed in other 17 plant species that belong to the order Arecales (palms), monocots, dicots, and gymnosperms (Figure 3 and Table S2). Among 46 editing sites observed in *B. flabellifer*, we found that nine sites (ndhD_326SL, ndhI_130SF, ndhK_248QST, petB_129AV, rpl22_83SL, rpoC1_169SL, matK-55SL, matK-63SL and matK426HY) were unique to this species, and 33 and 28 sites were shared among Arecales and monocots, respectively. Furthermore, 12 sites were shared between monocots and dicots, and five sites were shared between monocots and gymnosperms. This result suggested that more common RNA editing sites are likely to be found in evolutionary related species and that these editing sites could be useful for studying the evolutionary relationship of closely related species.

Phylogenetic trees of commelinids based on cp genes carrying RNA editing sites

Following the orthologous genes with common RNA editing sites identified among the commelinid clade, we tested whether these genes are able to support the evolutionary relationship among the orders within this clade. First, we used the DNA sequence of seven RNA edited genes; *accD*, *atpB*, *atpI*, *ndhB*, *ndhD*, *ndhF* and *rps8*, which have orthologous editing sites among the monocots and are orthologues to *M. polymorpha*. The evolutionary relationship among the orders was reconstructed as (D(P(A,ZC))) (Figure 4a). This tree topology did not corre-

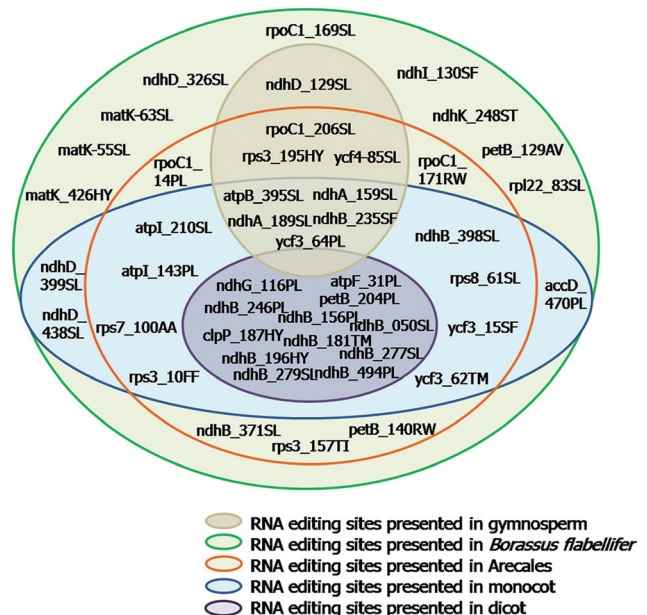


Figure 3 - Comparisons of 46 RNA editing sites observed in *B. flabellifer* to those observed in other 16 plants, which include gymnosperms, dicots, monocots and palms.

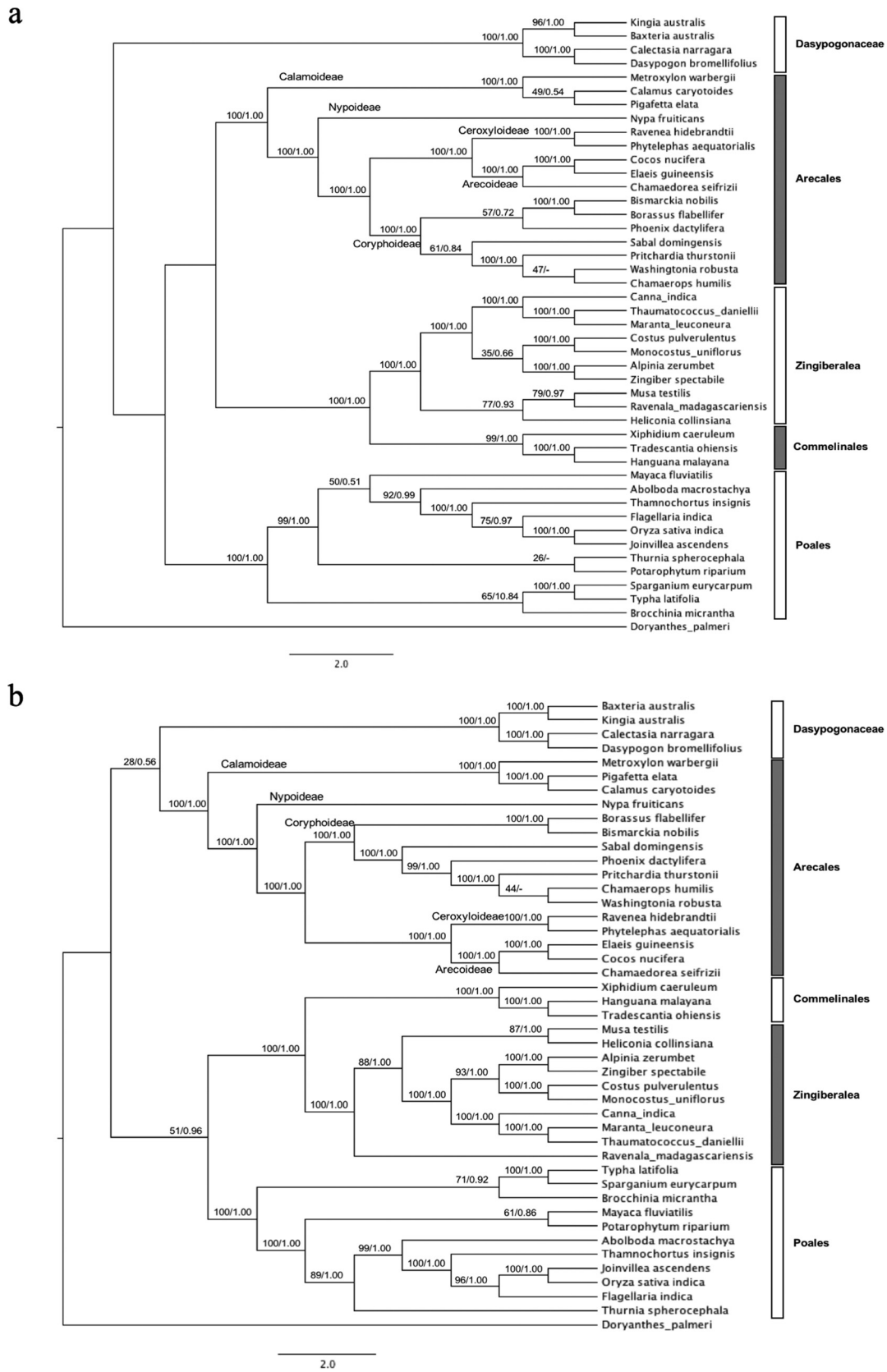


Figure 4 - Cladograms of commelinids reconstructed from seven cp genes (a) and 15 cp genes (b) based on maximum likelihood and Bayesian method with GTR+I+G model. The numbers on each branch present the bootstrap support (BP) and posterior probability value (PP), respectively. 10,000 replicates and 10M generations with stop value at 0.01 were applied to calculate statistical supports for maximum likelihood and Bayesian phylogeny, respectively.

spond well with the report from Barratt *et al.* (2016), which used an extensive dataset including the complete coding sequence of 75 genes of the cp genomes of 132 monocot taxa. This result indicated that a dataset using these seven gene sequences is insufficient for reconstructing a representative phylogeny for the commelinid clade members. We, therefore, built another phylogenetic tree using a new dataset, which included the seven RNA edited cp genes, additional four RNA edited cp genes: *matK*, *rpoA*, *ndhA* and *rpoC1*, and four non-RNA edited cp genes: *ycf1*, *rpoC2*, *ccsA* and *atpA* (Figure 4b). These additional eight genes were selected based on the number of parsimony-informative characters (PICs), which the PIC/length ratios of these genes were above 0.3 (Barratt *et al.*, 2013). Based on these 15 cp genes, the reconstructed phylogeny showed ((D,A)(P,ZC)) clustering, which was similar to that proposed by Barrett *et al.* (2016). Because *ycf1* is often considered variable, we therefore tested the phylogenetic analysis using 14 cp genes (without *ycf1*) and found that the clustering pattern remained the same with less bootstrap support than that of 15 cp genes (Figure S1). Thus, this result showed that the 15 cp sequences were able to provide reliable reconstruction of the topology of the commelinid clade.

Discussion

In this work, we reported for the first time RNA editing in the chloroplast of *B. flabellifer*. Analysis of 35 cp genes revealed authentically 46 RNA editing sites in 20 genes, and the editing was all with C-to-U type. U_A is the highest editing context observed in this species. Most of the editing occurred at the first and second base of codons resulting in amino acid changes in 43 out of 46 codons. The codon changes resulted in increases of the hydrophobicity and extension of the transmembrane regions in a number of proteins, particularly NdhA, NdhB, NdhD, NdhG, NdhI, and NdhK. Comparison of the amino acid changes via RNA editing from 18 plant species showed high numbers of shared editing sites in closely related species suggesting the use of DNA sequences of the genes carrying these editing sites for evaluating the evolutionary relationship. The phylogenetic tree built based on DNA sequences of seven orthologous genes carrying RNA editing sites could not represent the deep phylogeny of commelinids; however, this tree could well represent the evolutionary relationship among subfamilies of Arecaceae, suggesting that these seven loci could potentially be used for solving the evolutionary relationship among the members of the family Arecaceae.

All of the 46 RNA editing sites observed in the chloroplast of *B. flabellifer* demonstrated only the C-to-U type. Previous reports show that the C-to-U type is widespread among seed plants, whereas the reverse editing U-to-C type is present in some lower land plants including a hornwort (*Anthoceros formosae*), a fern (*Adiantum capillis-veneris*) and a lycophyte (*Selaginella uncinata*) (Yoshinaga *et al.*, 1996; Kugita *et al.*, 2003; Wolf *et al.*, 2004; Chateigner-

Boutin and Small, 2010; Grewe *et al.*, 2011), but not in liverworts, mosses and gymnosperm (Rüdinger *et al.*, 2009; Wu *et al.*, 2014; Guo *et al.*, 2015). This observation suggests that the U-to-C type was likely to be originated in the common ancestor of hornwort and tracheophytes and was lost via the separation between mosses and hornwort as well as via the separation between ferns and seed plants (Chateigner-Boutin and Small, 2010). However, in subfamily Arecoideae, Uthapaisanwong *et al.* (2012) showed that both editing types are present in *Elaeis guineensis* (oil palm), while *Cocos nucifera* (coconut) was reported to possess only C-to-U type (Huang *et al.*, 2013). This contradiction suggests that the U-to-C type may not be completely lost during the separation between ferns and seed plants or it may arise during the evolution of seed plants. Since we were unable to analyze all of the cp transcripts, we could not rule out the possibility for the U-to-C type in *B. flabellifer*. Further study to identify the U-to-C editing in *B. flabellifer* is needed for understanding the evolution of RNA editing in the cp genome in the family Arecaceae.

The predominant U_A context of RNA editing sites in *B. flabellifer* is similar to the context bias in other seed plants including *Phalaenopsis aphrodite*, *Nicotiana tabacum*, *Zea mays*, *Pinus thunbergii*, *Atropa belladonna*, and *Arabidopsis thaliana* (Tillich *et al.*, 2006; Zeng *et al.*, 2007). It has been proposed that approximately 30 nucleotide sequences surrounding the editing site are the recognized region of RNA editing factors, and conserved sequences within this region have not yet been identified (Okuda and Shikanai, 2012). Likewise, no conserved sequence could be deduced from alignments of 30 nucleotide-regions surrounding the 46 editing sites observed in the cp genome of *B. flabellifer*.

The increases in the hydrophobicity of the chloroplast proteins through the majority of the RNA editing in *B. flabellifer* are a widespread phenomenon. Indeed, our analysis showed evidence of the impacts of this hydrophobicity on the protein secondary structures. It was proposed that these consequently cause the extensions of transmembrane regions in many chloroplast proteins, particularly Ndh, which increase the stability of the NDH complex and makes it tightly bind to the thylakoid membrane (Jobson and Qiu, 2008).

Generally, RNA editing can generate both new start and stop codons. Many cp genes in angiosperm gain new start codons through RNA editing, which modifies C-to-U of ACG initiator codon to AUG standard start codon such as that in *ndhD*, *psbL* and *rpl2* (Sugiura *et al.*, 1998; Wu *et al.*, 2014). However, the analysis of RNA editing in *B. flabellifer* demonstrated the loss of ACG RNA editing in the initiator codon in *ndhD* and *rpl2*. The premature stop codon observed in the *ndhK* resulting in deleted eight amino acids at the C-terminus of NdhK protein has never been observed in the chloroplast of plants in the family Arecaceae, though this was previously observed in other monocots such as *Wolffia australiana* (NC_015899.1),

Colocasia esculenta (NC_016753.1) and *Zea mays* (NC_001666.2). Although we have analyzed 35 cp genes of *B. flabellifer* for RNA editing, in which 20 genes are positive, the assessment of RNA editing in this species is still far from being complete. There are other 45 out of 79 protein coding genes in the *B. flabellifer* cp genome left to be analyzed (Sakulsathaporn *et al.*, 2017).

The evolutionary relationship among the orders of commelinid was initially reconstructed using several sets of organelle genes including a set of cp genes: *rbcL*, *atpA*, *matK* and *ndhF* and a set of chloroplast and mitochondrial genes: *rbcL*, *matK*, *atp1*, and *nad5*. Several tree topologies were proposed such as ((ZC,A)(D,P)), (A(ZC(D,P))), (A(D(P,ZC))), (D(A(P,ZC))), ((P,ZC)(A,D)) and (A(P(ZC,D))) (Kress *et al.*, 2002; Davis *et al.*, 2004; Chase *et al.*, 2006; Graham *et al.*, 2006; Specht *et al.*, 2006; Saarela *et al.*, 2008; Givnish *et al.*, 2010). Barrett *et al.* (2013) reported that both the single most parsimonious tree and the best-scoring likelihood tree reconstructed based on plastome sequences of 46 taxa supported the ((A,D)(ZC,P)) topology. Nonetheless, Barrett *et al.* (2013) also proposed six alternative topologies: (ZC(A(D,P))), ((ZC,A)(D,P)), (A(ZC(D,P))), (A(D(P,ZC))), (D(A(P,ZC))), ((P,A)(ZC,D))), which were not significantly less likely present the evolutionary relationship among orders belonging to the commelinid clade. From our observation that the closely related plants have more shared editing sites compared to the distantly related plants, the phylogenetic tree reconstructed based on seven cp genes with RNA editing sites presented (D(P(A,ZC))) topology, which was reported as significantly less likely topology compared to the best-scoring likelihood tree (Barrett *et al.*, 2013). These results suggested that only the presence of RNA editing sites would not be a good criterion for selecting orthologous genes to represent deep phylogeny, the evolutionary relationship among plant orders.

Most recently, Barrett *et al.* (2016) reported a maximum likelihood phylogenetic tree with ((A,D)(ZC,P)) topology, which was reconstructed from plastome sequences comprised of 75 protein-coding genes. The (A,D) and (ZC,P) clusters of the trees reconstructed from 75 protein-coding genes were supported by 90% and 96% bootstrap samples, respectively, while those that reconstructed from 46 protein-coding genes previously reported by Barratt *et al.* (2013) were supported by 72% and 85% bootstrap samples, respectively. Compared to our topology as shown in Figure 4b, the tree was reconstructed using only 15 protein-coding genes from the cp genomes of plants belonging to the family Arecaceae, and this can provide a similar topology to those that used large datasets, though the (A,D) and (ZC,P) clusters were supported by only 28% and 51% of bootstrap samples, respectively. Noting that these 15 protein-coding genes comprised of seven edited genes, four edited genes with high PICs and four non-edited genes with high PICs. The similarity of the overall topology suggested that adding DNA sequence data of the genes carrying a high

number of PICs could improve the tree topology presenting evolutionary relationship among orders of the commelinids. This result also supports Barratt *et al.* (2013) that only a particular set of cp genes with high PICs could well represent the deep phylogeny of the commelinids.

It was notable that even though the two phylogenetic trees reconstructed using seven and 15 protein-coding cp genes present different evolutionary relationship among several orders of the commelinid clade, the close relationship between Zingiberales (Z) and Commelinales (C) is maintained as supported by 100% of bootstrap samples. And, each order presents a monophyletic cluster of plant species supported by at least 99% bootstrap samples. By focusing only on the evolutionary relationship among the subfamilies of palms (family Arecaceae, Order Arecales), both phylogenetic trees support that the relationship among these subfamilies as (Calamoideae, (Nypa, (Coryphoideae (Ceroxyloideae, Arecoideae)))) with almost 100% bootstrap samples. This relationship corresponds well with the clusters that were previously reconstructed by using nine plastid markers, four nuclear markers, a morphological dataset and a RFLP dataset (Baker *et al.*, 2011; Faurby *et al.*, 2016) as well as the phylogenetic relationship reconstructed by using 75 protein-coding genes of 132 taxa with 100% of bootstrap (Barratt *et al.*, 2016). Hence, the use of the seven genes, which have RNA editing and share with monocots, is sufficient for verifying the evolutionary relationship among the subfamilies of the Arecaceae family.

Acknowledgments

This work has been financial supported by Kasetsart University Research and Development Institute (KURDI) and OmiKU. Arpakorn Sakulsathaporn is supported by the Center of Excellence on Agricultural Biotechnology, Science and Technology Postgraduate Education and Research Development Office, Office of Higher Education Commission, Ministry of Education. (AG-BIO/PERDO-CHE). Supachai Vuttipongchaikij is supported by Thailand Research Fund (TRF-RSA6080031).

References

- Asif MH, Mantri SS, Sharma A, Srivastava A, Trivedi I, Gupta P, Mohanty CS, Sawant SV and Tuli R (2010) Complete sequence and organisation of the *Jatropha curcas* (Euphorbiaceae) chloroplast genome. *Tree Genet Genomes* 6:941-952.
- Baker WJ, Norup MV, Clarkson JJ, Couvreur TL, Dowe JL, Lewis CE, Pintaud JC, Savolainen V, Wilmot T and Chase MW (2011) Phylogenetic relationships among areoid palms (Arecaceae: Arecoideae). *Annu Bot* 108:1417-1432.
- Barrett CF, Baker WJ, Comer JR, Conran JG, Lahmeyer SC, Leebens-Mack JH, Li J, Lim GS, Mayfield-Jones DR and Perez L (2016) Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. *New Phytol* 209:855-870.

- Barrett CF, Davis JJ, Leebens-Mack J, Conran JG and Stevenson DW (2013) Plastid genomes and deep relationships among the commelinid monocot angiosperms. *Cladistics* 29:65-87.
- Castandet B and Araya A (2011) RNA editing in plant organelles Why make it easy? *Biochem (Mosc)* 76:924-931.
- Chase MW, Fay MF, Devey DS, Maurin O, Rønsted N, Davies TJ, Pillon Y, Petersen G, Seberg O and Tamura MN (2006) Multigene analyses of monocot relationships: a summary. *Aliso* 22:63-75.
- Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH, Mishler BD, Duvall MR, Price RA, Hills HG and Qiu YL (1993) Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. *Ann Mo Bot Gard* 80:528-580.
- Chateigner-Boutin AL and Small I (2011) Organellar RNA editing. *Wiley Interdiscip Rev RNA* 2:493-506.
- Chen H, Deng L, Jiang Y, Lu Pand Yu J (2011) RNA editing sites exist in protein-coding genes in hte chloroplast genome of *Cycas taitungensis*. *J Integr Plant Biol* 53:961-970.
- Corneille S, Lutz K and Maliga P (2000) Conservation of RNA editing between rice and maize plastids: are most editing events dispensable? *Mol Genet Genomics* 264:419-424.
- Darriba D, Taboada GL, Doallo R and Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* 9:772.
- Davis JJ, Stevenson DW, Petersen G, Seberg O, Campbell LM, Freudenstein JV, Goldman DH, Hardy CR, Michelangeli FA, Simmons MP and Specht CD (2004) A phylogeny of the monocots, as inferred from *rbcL* and *atpA* sequence variation, and a comparison of methods for calculating jackknife and bootstrap values. *Syst Bot* 29:467-510.
- Dorrell RG, Hinksman GA and Howe CJ (2016) Diversity of transcripts and transcript processing forms in plastids of the dinoflagellate alga *Karenia mikimotoi*. *Plant Mol Biol* 90:233-247.
- Faurby S, Eiserhardt WL, Baker WJ and Svenning JC (2016) An all-evidence species-level supertree for the palms (Arecaceae). *Mol Phylogenet Evol* 100:57-69.
- Givnish TJ, Ames M, McNeal JR, McKain MR, Steele PR, Graham SW, Pires JC, Stevenson DW, Zomlefer WB and Briggs BG (2010) Assembling the the tree of the monocotyledons: plastome sequence phylogeny and evolution of Poales.. *Ann Mo Bot Gard* 97:584-616.
- Graham SW, Zgurski JM, McPherson MA, Cherniawsky DM, Saarela JM, Horne EF, Smith SY, Wong WA, O'Brien HE, Biron VL and Pires JC (2006) Robust inference of monocot deep phylogeny using an expanded multigene plastid data set. *Aliso* 22:3-20.
- Grewe F, Herres S, Viehöver P, Polsakiewicz M, Weisshaar B and Knoop V (2010) A unique transcriptome: 1782 positions of RNA editing alter 1406 codon identities in mitochondrial mRNAs of the lycophyte *Isoetes engelmannii*. *Nucleic Acids Res* 39:2890-2902.
- Guo W, Grewe F and Mower JP (2015) Variable frequency of plastid RNA editing among ferns and repeated loss of uridine-to-cytidine editing from vascular plants. *PLoS One* 10:e0117075.
- He P, Huang S, Xiao G, Zhang Y and Yu J (2016) Abundant RNA editing sites of chloroplast protein-coding genes in Ginkgo biloba and an evolutionary pattern analysis. *BMC Plant Biol* 16:257.
- Huang YY, Matzke AJ and Matzke M (2013) Complete sequence and comparative analysis of the chloroplast genome of coconut palm (*Cocos nucifera*). *PLoS One* 8:e74736.
- Inada M, Sasaki T, Yukawa M, Tsudzuki T and Sugiura M (2004) A systematic search for RNA editing sites in pea chloroplasts: an editing event causes diversification from the evolutionarily conserved amino acid sequence. *Plant Cell Physiol* 45:1615-1622.
- Jobson RW and Qiu YL (2008) Did RNA editing in plant organellar genomes originate under natural selection or through genetic drift? *Biol Direct* 3:43.
- Junior TC, Carraro DM, Benatti MR, Barbosa AC, Kitajima JP and Carrer H (2004) Structural features and transcript-editing analysis of sugarcane (*Saccharum officinarum* L) chloroplast genome. *Curr Genet* 46:366-373.
- Kress WJ, Prince LM and Williams KJ (2002) The phylogeny and a new classification of the ginger (Zingiberaceae): evidence from molecular data. *Am J Bot* 89:1682-1696.
- Kugita M, Kaneko A, Yamamoto Y, Takeya Y, Matsumoto T and Yoshinaga K (2003) The complete nucleotide sequence of the hornwort (*Anthoceros formosae*) chloroplast genome: insight into the earliest land plants. *Nucleic Acids Res* 31:716-721.
- Lenz H, Hein A and Knoop V (2018) Plant organelle RNA editing and its specificity factors: enhancements of analyses and new database features in PREPACT 3.0. *BMC Bioinformatics* 19:255.
- Okuda K and Shikanai T (2012) A pentatricopeptide repeat protein acts as a site-specificity factor at multiple RNA editing sites with unrelated cis-acting elements in plastids. *Nucleic Acids Res* 40:5052-5064.
- Peña MJ, Kulkarni AR, Backe J, Boyd M, O'Neill MA and York WS (2016) Structural diversity of xylans in the cell walls of monocots. *Planta* 244:589-606.
- Pipatchartlearnwong K, Swatdipong A, Vuttipongchaikij S and Apisitwanich S (2017) Genetic evidence of multiple invasions and a small number of founders of Asian Palmyra palm (*Borassus flabellifer*) in Thailand. *BMC Genet* 18:88.
- Ranwez V, Harispe S, Delsuc F and Douzery EJ (2011) MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons. *PLoS One*, 6:e22594.
- Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA and Huelsenbeck JP (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 61:539-542.
- Rüdinger M, Funk HT, Rensing SA, Maier UG and Knoop V (2009) RNA editing: only eleven sites are present in the *Physcomitrella patens* mitochondrial transcriptome and a universal nomenclature proposal. *Mol Genet Genomics* 281:473-481.
- Sakulsathaporn A, Wonnapijit P, Vuttipongchaikij S and Apisitwanich S (2017) The complete chloroplast genome sequence of Asian Palmyra palm (*Borassus flabellifer*). *BMC Res Notes* 10:740.
- Sanchez-Puerta MV and Abbona CC (2014) The chloroplast genome of *Hyoscyamus niger* and a phylogenetic study of the tribe Hyoscyameae (Solanaceae). *PLoS One* 9:e98353.
- Sasaki T, Yukawa Y, Miyamoto T, Obokata J and Sugiura M (2003) Identification of RNA editing sites in chloroplast transcripts from the maternal and paternal progenitors of to-

- bacco (*Nicotiana tabacum*): comparative analysis shows the involvement of distinct trans-factors for *ndhB* editing. *Mol Biol Evol* 20:1028-1035.
- Saarela JM, Prentis PJ, Rai HS and Graham SW (2008) Phylogenetic relationships in the monocot order Commelinales, with a focus on Philydraceae. *Botany* 86:719-731.
- Silvestro D and Michalak I (2012) raxmlGUI: a graphical front-end for RAxML. *Org Divers Evol* 12:335-337.
- Specht CD and Stevenson DW (2006) A new phylogeny-based generic classification of Costaceae (Zingiberales). *Taxon* 55:153-163.
- Sugiura M, Hirose T and Sugita M (1998) Evolution and mechanism of translation in chloroplasts. *Annu Rev Genet* 32:437-459.
- Tillich M, Lehwick P, Morton BR and Maier UG (2006) The evolution of chloroplast RNA editing. *Mol Biol Evol* 23:1912-1921.
- Uthapaisanwong P, Chanprasert J, Shearman J, Sangsarakru D, Yoocha T, Jomchai N, Jantassuriyarat C, Tragoonrung S and Tangphatsornruang S (2012) Characterization of the chloroplast genome sequence of oil palm (*Elaeis guineensis* Jacq). *Gene* 500:172-180.
- Vaidya G, Lohman DJ and Meier R (2011) SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics* 27:171-180.
- Wakasugi T, Hirose T, Horihata M, Tsudzuki T, Kössel H and Sugiura M (1996) Creation of a novel protein-coding region at the RNA level in black pine chloroplasts: the pattern of RNA editing in the gymnosperm chloroplast is different from that in angiosperms. *Proc Natl Acad Sci U S A* 93:8766-8770.
- Wang W, Zhang W, Wu Y, Maliga P, Messing J (2015) RNA editing in chloroplasts of *Spirodela polyrrhiza*, an aquatic monocotyledonous species. *PLoS One* 10:e0140285.
- Wolf PG, Rowe CA and Hasebe M (2004) High levels of RNA editing in a vascular plant chloroplast genome: analysis of transcripts from the fern *Adiantum capillus-veneris*. *Gene* 339:89-97.
- Wu CS and Chaw SM (2014) Highly rearranged and size-variable chloroplast genomes in conifers II clade (cupressophytes): evolution towards shorter intergenic spacers. *Plant Biotechnol J* 12:344-353.
- Yoshinaga K, Inuma H, Masuzawa T and Uedal K (1996) Extensive RNA editing of U to C in addition to C to U substitution in the *rbcL* transcripts of hornwort chloroplasts and the origin of RNA editing in green plants. *Nucleic Acids Res* 24:1008-1014.
- Zeng WH, Liao SC and Chang CC (2007) Identification of RNA editing sites in chloroplast transcripts of *Phalaenopsis aphrodite* and comparative analysis with those of other seed plants. *Plant Cell Physiol* 48:362-368.

Supplementary material

The following online material is available for this article:
 Figure S1 - Cladograms of commelinids reconstructed from 14 cp genes (without *ycf1*) based on maximum likelihood and Bayesian method with GTR+I+G model.
 Table S1 - Oligonucleotide primers for amplification of 35 cp genes.
 Table S2 - Comparison of RNA editing sites in chloroplast transcripts from 18 plant species.
 Table S3 - Accessions numbers of cp genomes from 59 plant species.
 Table S4 - Predicted and experimentally identified RNA editing sites in 35 chloroplast genes of *B. flabellifer*. + = editing site, - = no editing site and # = not predicted by the PREP-Cp.

Associate Editor: Rogério Margis

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License (type CC-BY), which permits unrestricted use, distribution and reproduction in any medium, provided the original article is properly cited.

Agriculture and Natural Resources. year. Vol(No): xx–xx.

Agr. Nat. Resour. year. Vol(No): xx–xx.

Straw digestibility of Thai rice accessions

**Yuppayao Kophimai^a, Saran Cheenacharoen^b, Rachael Simister^c, Leonardo D. Gomez^c,
Simon J. McQueen-Mason^c, Supachai Vuttipongchaikij^{d,e,*}**

^a Program in Genetics, Faculty of Science, Maejo University, Chiang Mai 50290, Thailand.

^b Department of Biology and Centre of Excellence for Biodiversity Research and
Implementation for Community, Faculty of Science and Technology, Chiang Mai Rajabhat
University, Chiang Mai 50300, Thailand.

^c CNAP, Department of Biology, University of York, Heslington, York YO10 5YW, United
Kingdom.

^d Department of Genetics, Faculty of Science, Kasetsart University, 50 Ngarm Wong Wan
road, Chattuchak, Bangkok, 10900, Thailand.

^e Center of Advanced studies for Tropical Natural Resources, Kasetsart University, Ngam
Wong Wan, Chattuchak, Bangkok 10900, Thailand.

Article history:

Received

Accepted

Available online

Keywords: Biomass, Plant cell wall, Polysaccharides, Rice, Straw

* Corresponding author.

E-mail address: fsciscv@ku.ac.th (S. Vuttipongchaikij)

Abstract

Rice straw is major agricultural biomass in Southeast Asia. However, it has not been utilized to full potential due to its recalcitrant to breakdown by an enzymatic process. Recent work demonstrated a distribution of straw digestibility among rice varieties, but this trait has not been shown for the rice population in Thailand. Identifying rice varieties with highly digestible straw would be useful for biofuel, biorefinery and animal feed. Here, we assessed the digestibility of straw from 49 Thai rice accessions using the enzymatic-based high-throughput platform. Straw saccharification released sugars in the range of 50.53-114.52 nmol mg⁻¹ hr⁻¹. The highest and lowest saccharification potential were from a glutinous variety “Muey Nawng 62 M” and an upland variety “Dawk Pa-Yawm”, respectively. Rice accessions with high digestibility for potential use are presented. These data provide a preliminary assessment of rice straw from the Thai population and would be useful for studying genes responsible for the digestibility trait and breeding programs for both rice grain production and straw utilization.

1 Introduction

2 Rice straw is a major agricultural biomass waste produced in Southeast Asia and
3 particularly in Thailand, and yet most of this biomass are left unutilized and often burned in
4 the field causing severe environment problems at both local and global levels (Oanh, et al.,
5 2011; Junpen et al., 2018). It is estimated that over 20 million tonnes of rice straw are
6 produced in Thailand every year (Gadde et al., 2009) and that, during the peak seasons of rice
7 harvesting, straw is burned on 90% of rice paddies to get rid of it (Tipayarom and Oanh,
8 2007). Open rice straw burning results in incomplete combustion giving rise to high levels of
9 particulate black carbon, carbon monoxide and ozone, which cause respiratory problems and
10 premature deaths among the human population, crop yield loss and atmospheric warming
11 (UNEP/WMO, 2011). With the burden as waste, rice straw is posing both a critical threat
12 from our current ways of elimination and, at the same time, a great opportunity to benefit
13 humankind by converting it into biomass-based products (Sattlewal et al., 2018).

14 Rice straw is rich in polysaccharides (~70% dry weight), and the large volume of
15 straw produced annually could, therefore, provide a valuable lignocellulose feedstock for
16 producing biofuels, chemicals and animal feed (Chen and Dixon, 2007; Harris and DeBolt,
17 2010). Unfortunately, compositional characteristics of rice straw currently hinder its
18 applications as it is inherently difficult to digest by enzymes, making it costly for biofuel and
19 chemical production through fermentation. Furthermore, rice straw has notably high levels of
20 silica (up to 15% dry weight), which makes it particularly unpalatable to animals and,
21 thereby, poor quality as animal feed (Virk et al., 2019). Nonetheless, recent work has shown
22 that there is large natural variation in all of these parameters among commercial rice
23 varieties, suggesting that there is substantial scope for improving the characteristics of rice
24 straw without having negative impacts on plant performance and grain yield (Marriott et al.,
25 2016; Garrido et al., 2018). Moreover, recent work using Genome-Wide Association Studies
26 (GWAS) have identified some regions in genomes that are responsible for variation in rice
27 straw digestibility (Liu et al., 2016).

28 In this work, we report rice straw saccharification analysis of 49 Thai rice accessions
29 (taken from both commercial and landrace varieties) by enzymatic digestion using a platform
30 developed by Gomez et al., (2010). The central aim of this work is to lay the groundwork to
31 drive widespread use of rice straw in animal feed, biofuel and bioenergy applications in
32 Thailand. Selection of specific varieties with high straw saccharification potential will pave
33 ways for developing rice breeds with high straw and hush values for further exploitation.

Materials and methods

Plant materials

Seeds of 49 Thai rice accessions were obtained from Pathum-Thani Rice Research Centre (Thailand) (Table 1). For straw sample collection, rice was grown in a pot (30 cm diameter and 22.5 cm depth) using one germinating seed per pot and five replicating pots per variety. Rice pots were arrayed in a completely randomized design on a plain field in Chiang Mai Province, Thailand (18°52'47.5"N, 99°02'43.2"E) during January-April 2016. Rice cultivation was performed using a submerged irrigation method with regular feeding for four months or until flowered and then left to dry up to two weeks. Stems from the main tillers were collected, and straw samples were taken from the third and the fourth node. Straw samples were further dried in an incubator at 45°C for two days before cut into ~4 mm length and pulverized to a fine powder using a robotic grinding platform described in Gomez et al. (2010).

Enzymatic saccharification analysis

Rice straw saccharification was performed in 96-well-plate formats using Celluclast (cellulose from *Trichoderma reesei*) and Novozyme 188 (Novozymes A/S, Bagsvaerd, Denmark) at a ratio of 4:1 at an enzyme loading of 22.5 Filter Paper Unit (FPU) per gram of material and a robotic platform (Tecan Evo200; Tecan Group Ltd.), as described by Gomez et al. (2010). Briefly, with equipped robotic tools, 4 mg of straw powder were loaded into 96-deep-well plates, pre-treated with 350 µL of 0.5 M NaOH solution at 90°C for 30 min, washed for five times with 500 µL of 25 mM sodium acetate buffer pH 4.5 and then incubated with the enzyme cocktail at 50°C for 8 hrs. Noting that this 8 hrs digestion was programmed for verifying the saccharification potential or digestibility of straw among rice accessions, not for the total saccharification values of straw biomass. Biomass hydrolysates were aliquoted by the robotic platform into 96-well-plates and analyzed for released reducing sugar using a modified MBTH method (Gomez et al., 2010; 2011). Our pre-treatment using 0.5 M NaOH is required to remove the majority of hemicellulose from straw materials before the enzymatic hydrolysis, and saccharified products consist mainly of glucose from cellulose (Moradi et al., 2013; Wood et al., 2016). The experiment was performed using three biological replicates, each with four technical replicates. OD reads were converted to amounts of reducing sugars released in nanomoles using glucose standards, as previously described.

Statistical analysis

Data analysis was performed in R environment with Agricola for statistical analysis. The dataset was accessed for significant differences among rice accessions using analysis of variance (ANOVA), and samples were grouped using the TUKEY method ($p < 0.05$).

Results and Discussion

To explore the saccharification potential, we analyzed straw collected from 49 Thai rice varieties using an automated enzymatic hydrolysis platform developed by Gomez et al., (2010). The means of sugar released from straw biomass from these 49 varieties distributed between 50.53-114.52 nmol mg⁻¹ hr⁻¹ (Fig. 1). The highest and lowest saccharification potential were from R30, a glutinous variety “Muey Nawng 62 M”, and R13, an upland variety “Dawk Pa-Yawm”, respectively (Table 1). Normal distribution was observed for the sugar released values of the 49 accessions (Fig. 2). Because of data continuity, statistical analysis was unable to clearly distinguish the saccharification potential from one another. Nonetheless, some accessions could be grouped based on TUKEY analysis for those with high saccharification potential, for example, R30, R47, R31, R02, R43, R34, R37, R44, R40 and R03. No correlation was observed between the saccharification values and rice groups such as grain types and high or low land accessions.

Previously, rice straw saccharifications by enzymatic hydrolysis were reported with various ranges of sugar releases using different rice populations, including those from China (59-116 nmol mg⁻¹ hr⁻¹; Liu et al., 2016) and Vietnam (20-134 nmol mg⁻¹ hr⁻¹; Nguyen, 2016). Here, our analysis among selected Thai rice accessions showed a similar distribution and range of saccharification potential, demonstrating that the trait diversity is presented within the Thai population. Noting that these analyses were performed using the same platform and experimental conditions, allowing the comparison of the saccharification potential. Furthermore, our highest saccharification value up to 114 nmol mg⁻¹ hr⁻¹ suggests that the Thai accessions could hold competency compared to the populations from China and Vietnam. However, most of our top saccharification values were from accessions with restricted propagation, local uses or low production, thereby limiting a direct utilization as a large-scale biomass resource. Thus, these accessions are more likely to be useful for introgression breeding for high digestibility traits. A recent analysis of biomass digestibility and productivity traits showed that both traits are non-correlated and indicated that we could breed these crops for both high productivity and high digestibility traits (Garrido et al., 2018).

Various studies of biomass saccharification have revealed some insight into the plant cell wall recalcitrant, which has a multitude of pathways contributing to this end, particularly the cell wall polysaccharide components particularly cellulose and hemicelluloses (Demartini et al., 2013; Marriott et al., 2014). Followed up studies of biomass from grasses and rice demonstrated that lignins also play key roles in the recalcitrant (Fu et al., 2011; Bartley et al., 2013; Bouvier d'Yvoire et al., 2013). Interestingly, silica appeared to be one of the key important parts for the recalcitrant of rice straw and husk (Zhang et al., 2015), acting as a rice defense to pathogen and even forage feeding animals (Agbagla-Dohnani et al., 2003). Changes in the complexity of biomass in many aspects including composition, structure, molecular arrangement could affect the saccharification potential (Gomez et al., 2010; Marriott et al., 2014; Whitehead et al., 2018). Though non-cell wall related, starch remained in rice straw biomass also provides an add-on to final saccharification products (Teramura et al., 2013). Other aspects of plant cell walls responsible for biomass recalcitrant are now being discovered, and much recent progress is owed to the work of genetic variation and genome analysis, including GWAS and QTL analysis. Newly identified genes will be beneficial both for marker-assisted selection for selecting breeding lines with low recalcitrant and providing further insight into the genetics underlying straw digestibility. When the genomic data of the rice accessions used here become available, our saccharification data would be useful for identifying loci responsible for the straw recalcitrant.

Quantitative traits are largely affected by the environment, and, indeed, saccharification is also prone to this effect. Given the result from Nguyen (2016), straw saccharification results were highly different between two consecutive years: 20-134 and 23-72 nmol mg⁻¹ hr⁻¹ from 2013 and 2014, respectively, under well-planned paddy field experiments in Vietnam. Thus, it is difficult to directly compared biomass saccharification results among different experiments. Although our analysis was based on the straw materials collected from rice grown in pots, which may pose a different situation from those grown in the paddy fields, this result is useful for uniformly comparing the saccharification potential among the tested accessions. Pot-cultivation with a controlled environment including well-mixed compost, watering and fertilizer feeding would minimize the environmental effects usually encountered in samples collected from the fields. However, it is crucial in the future to perform the analysis using straw samples from rice cultivated in paddy fields for accessing biomass yields on the fields and their saccharified products.

In conclusion, enzymatic straw saccharification of Thai rice accessions presented here serves as preliminary data to support the utilization of rice straw biomass in Thailand. Rice

accessions with high saccharification potential identified here could be beneficial both for direct biomass production and for breeding programs. Besides, these data also prompt us to perform saccharification screening on a more extensive collection of Thai rice accessions. A more detailed study using plot experimental designs, DNA sequence data and genetic analysis tools such as GWAS would allow us to use this straw waste more effectively. Diverting rice straw from burning will not only reduce the harmful consequences but also provide a new source of animal feed, biofuels and bioenergy with additional environmental benefits from reducing the land required to produce forage and the use of fossil fuels.

Ethics Statements

Human or animal ethic or biosafety approval is not required for this work.

Conflict of interest

The authors declare that there are no conflicts of interest.

Acknowledgements

We thank Pathum-Thani Rice Research Centre, Ministry of Agriculture (Thailand) for rice germplasm. This work was supported by Newton Researcher links through the British council and Thailand Research Fund (TRF), the National Research Council of Thailand (MJU-2016 and CMRU-2017) and Thailand Research Fund (TRF-RSA6080031).

References

- Agbagla-Dohnani, A., Noziere, P., Gaillard-Martinie, B., Puard, M. and Doreau, M., 2003. Effect of silica content on rice straw ruminal degradation. *The Journal of Agricultural Science*, 140(2), pp.183–192.
- Bartley, L.E., Peck, M.L., Kim, S.R., Ebert, B., Manisseri, C., Chiniquy, D.M., Sykes, R., Gao, L., Rautengarten, C., Vega-Sánchez, M.E. and Benke, P.I., 2013. Overexpression of a BAHD acyltransferase, OsAt10, alters rice cell wall hydroxycinnamic acid content and saccharification. *Plant Physiology*, 161(4), pp.1615–1633.
- Bouvier d'Yvoire, M., Bouchabke-Coussa, O., Voorend, W., Antelme, S., Cézard, L., Legée, F., Lebris, P., Legay, S., Whitehead, C., McQueen-Mason, S.J. and Gomez, L.D., 2013. Disrupting the cinnamyl alcohol dehydrogenase 1 gene (Bd CAD 1) leads to altered

- lignification and improved saccharification in *Brachypodium distachyon*. *The Plant Journal*, 73(3), pp.496–508.
- Chen, F. and Dixon, R.A., 2007. Lignin modification improves fermentable sugar yields for biofuel production. *Nature biotechnology*, 25(7), p.759.
- DeMartini, J.D., Pattathil, S., Miller, J.S., Li, H., Hahn, M.G. and Wyman, C.E., 2013. Investigating plant cell wall components that affect biomass recalcitrance in poplar and switchgrass. *Energy & Environmental Science*, 6(3), pp.898–909.
- Fu, C., Mielenz, J.R., Xiao, X., Ge, Y., Hamilton, C.Y., Rodriguez, M., Chen, F., Foston, M., Ragauskas, A., Bouton, J. and Dixon, R.A., 2011. Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass. *Proceedings of the National Academy of Sciences*, 108(9), pp.3803–3808.
- Gadde, B., Menke, C. and Wassmann, R., 2009. Rice straw as a renewable energy source in India, Thailand, and the Philippines: Overall potential and limitations for energy contribution and greenhouse gas mitigation. *Biomass and bioenergy*, 33(11), pp.1532–1546.
- Garrido, F.J.O., Pistón, F., Gómez, L.D. and McQueen-Mason, S.J., 2018. Biomass recalcitrance in barley, wheat and triticale straw: Correlation of biomass quality with classic agronomical traits. *PloS one*, 13(11), p.e0205880.
- Gomez, L.D., Whitehead, C., Barakate, A., Halpin, C. and McQueen-Mason, S.J., 2010. Automated saccharification assay for determination of digestibility in plant materials. *Biotechnology for biofuels*, 3(1), p.23.
- Gomez, L.D., Whitehead, C., Roberts, P. and McQueen-Mason, S.J., 2011. High-throughput Saccharification assay for lignocellulosic materials. *JoVE (Journal of Visualized Experiments)*, (53), p.e3240.
- Harris, D. and DeBolt, S., 2010. Synthesis, regulation and utilization of lignocellulosic biomass. *Plant biotechnology journal*, 8(3), pp.244–262.
- Junpen, A., Pansuk, J., Kamnoet, O., Cheewaphongphan, P. and Garivait, S., 2018. Emission of Air Pollutants from Rice Residue Open Burning in Thailand, 2018. *Atmosphere*, 9(11), p.449.
- Liu, B., Gómez, L.D., Hua, C., Sun, L., Ali, I., Huang, L., Yu, C., Simister, R., Steele-King, C., Gan, Y. and McQueen-Mason, S.J., 2016. Linkage mapping of stem saccharification digestibility in rice. *PloS one*, 11(7), p.e0159117.

- Marriott, P.E., Gómez, L.D. and McQueen-Mason, S.J., 2016. Unlocking the potential of lignocellulosic biomass through plant science. *New Phytologist*, 209(4), pp.1366–1381.
- Marriott, P.E., Sibout, R., Lapierre, C., Fangel, J.U., Willats, W.G., Hofte, H., Gómez, L.D. and McQueen-Mason, S.J., 2014. Range of cell-wall alterations enhance saccharification in *Brachypodium distachyon* mutants. *Proceedings of the National Academy of Sciences*, 111(40), pp.14601–14606.
- Moradi, F., Amiri, H., Soleimanian-Zad, S., Ehsani, M.R. and Karimi, K., 2013. Improvement of acetone, butanol and ethanol production from rice straw by acid and alkaline pretreatments. *Fuel*, 112, pp.8-13.
- Nguyen, D., 2016. Using association mapping to identify Quantitative Trait Loci (QTL) for straw digestibility in rice straw. Master dissertation, University of York. York, United Kingdom.
- Oanh, N.T.K., Ly, B.T., Tipayarom, D., Manandhar, B.R., Prapat, P., Simpson, C.D. and Liu, L.J.S., 2011. Characterization of particulate matter emission from open burning of rice straw. *Atmospheric Environment*, 45(2), pp.493–502.
- Sattlewal, A., Agrawal, R., Bhagia, S., Das, P. and Ragauskas, A.J., 2018. Rice straw as a feedstock for biofuels: Availability, recalcitrance, and chemical properties. *Biofuels, Bioproducts and Biorefining*, 12(1), pp.83–107.
- Teramura, H., Oshima, T., Matsuda, F., Sasaki, K., Ogino, C., Yamasaki, M. and Kondo, A., 2013. Glucose content in the liquid hydrolysate after dilute acid pretreatment is affected by the starch content in rice straw. *Bioresource technology*, 149, pp.520–524.
- Tipayarom, D. and Oanh, N.K., 2007. Effects from open rice straw burning emission on air quality in the Bangkok Metropolitan Region. *Science Asia*, 33(3), pp.339–345.
- UNEP, W., 2011. Integrated assessment of black carbon and tropospheric ozone: summary for decision makers. Nairobi: United Nations Environment Programme.
- Virk, P., Xianglin, L. and Blümmel, M., 2019. A note on variation in grain and straw fodder quality traits in 437 cultivars of rice from the varietal groups of aromatic, hybrids, Indica, new planting types and released varieties in the Philippines. *Field crops research*, 233, pp.96–100.
- Whitehead, C., Ostos Garrido, F.J., Reymond, M., Simister, R., Distelfeld, A., Atienza, S.G., Piston, F., Gomez, L.D. and McQueen-Mason, S.J., 2018. A glycosyl transferase family 43 protein involved in xylan biosynthesis is associated with straw digestibility in *Brachypodium distachyon*. *New Phytologist*, 218(3), pp.974-985.

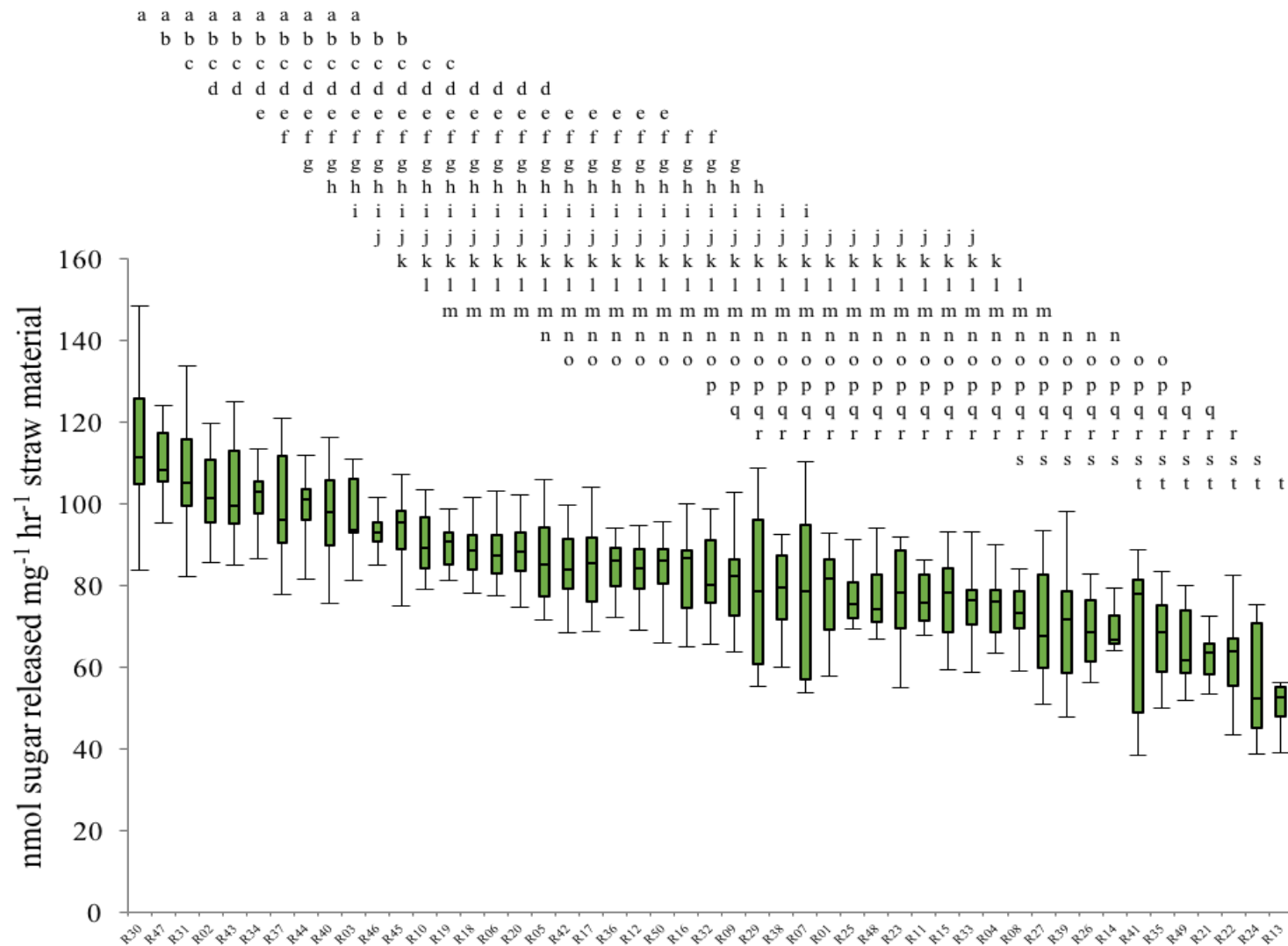
- 1 Wood, I.P., Cao, H.G., Tran, L., Cook, N., Ryden, P., Wilson, D.R., Moates, G.K., Collins,
2 S.R., Elliston, A. and Waldron, K.W., 2016. Comparison of saccharification and
3 fermentation of steam exploded rice straw and rice husk. *Biotechnology for*
4 *biofuels*, 9(1), p.193.
- 5 Zhang, J., Zou, W., Li, Y., Feng, Y., Zhang, H., Wu, Z., Tu, Y., Wang, Y., Cai, X. and Peng,
6 L., 2015. Silica distinctively affects cell wall features and lignocellulosic
7 saccharification with large enhancement on biomass production in rice. *Plant*
8 *Science*, 239, pp.84–91.

1 Table 1 Forty-nine Thai rice accessions used in this experiment. Accessions are listed from
 2 the highest to lowest mean values of saccharification. NG = non-glutinous variety; and G =
 3 glutinous variety.

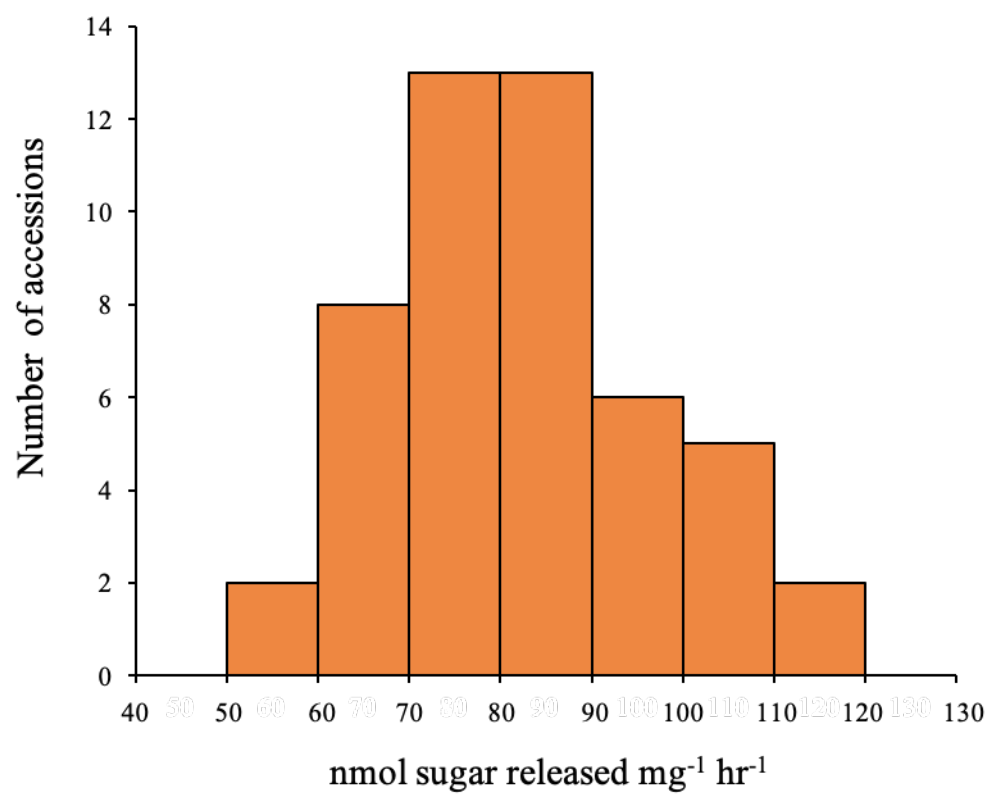
	ID	Accession	Grain type
1	R30	Muey Nawng 62M	G
2	R47	Riceberry	NG
3	R31	RD16	G
4	R02	RD15	NG
5	R43	RD33	NG
6	R34	Phitsanulok 2	NG
7	R37	RD39	NG
8	R44	RD41	NG
9	R40	RD21	NG
10	R03	Leb Meu Nahng 111	NG
11	R46	Homnil	NG
12	R45	RD47	NG
13	R10	Pin Gaew 56	NG
14	R19	Hantra 60	NG
15	R18	Khem Tawng Phatthalung	NG
16	R06	Khao Tah Haeng 17	NG
17	R20	Ta-pow Gaew 161	NG
18	R05	Goo Meuang Luang	NG
19	R42	Hawm Suphan	NG
20	R17	Gaen Jan	NG
21	R36	Chai Nat 80 (RD29)	NG
22	R12	Sao Hai	NG
23	R50	RD14	G
24	R16	Hommali Daeng	NG
25	R32	Pathum Thani 1	NG
26	R09	Prachin Buri 2	NG
27	R29	Sew Mae Jan	G
28	R38	RD43	NG
29	R07	Nahng Mon S-4	NG
30	R01	KDML 105	NG
31	R25	Hahng Yi 71	G
32	R48	San-pah-tawng 1	G
33	R23	RD6	G
34	R11	Plai Ngahm Prachin Buri	NG
35	R15	Sang Yod Phattalung	NG
36	R33	Suphan Buri 1	NG
37	R04	RD51	NG
38	R08	Leuang Pratew 123	NG
39	R27	Nahng Cha-lawng	G
40	R39	RD49	NG
41	R26	Niaw Kiaw Ngoo	G
42	R14	Sinlek	NG
43	R41	Pathum Thani 80 (RD31)	NG
44	R35	Chai Nat 1	NG

45	R49	RD10	G
46	R21	Leuang no.1	NG
47	R22	RD19	NG
48	R24	Leum Pua	G
49	R13	Dawk Pa-Yawm	NG

- 1 Fig. 1. Rice straw saccharification of 49 Thai rice accessions. Box plots represent distribution of reducing sugars released by enzymatic
 2 hydrolysis of pre-treated rice straw (three biological replicates and each with four technical replicates). Turkey's statistical test is indicated,
 3 where samples without a common letter differ ($p < 0.05$).



1 Fig. 2. Distribution of straw saccharification potential among 49 accessions.



2