Development of DNA marker for selection of intergeneric hybrids (B3-1c) (JIRCAS)

การพัฒนาและคัดเลือกเครื่องหมายโมเลกุล SSR บนลำดับนิวโคลีโอไทด์ของอ้อยป่า *Erianthus arundinaceus*

Development and characterization of SSR marker based on transcriptome sequencing in *Erianthus arundinaceus*

วีรกรณ์ แสงไสย์ เบญจวรรณ รัตวัตร ศุภรัตน์ ศรีทะวงษ์ Masumi Ebina Makoto Kobayashi Terashima Yoshifumi ศุจิรัตน์ สงวนรังศิริกุล อัมราวรรณ ทิพยวัฒน์ Shinichi Tsuruta

W. Saengsai B. Ruttawat S. Srithawong M. Ebina M. Kobayashi Y. Terajima A. Tippayawat S. Sakuanrungsirikul S. Tsuruta

ABSTRACT

We developed simple sequence repeat (SSR) markers from RNA-seq data for *E. arundinaceus.* Using an Illumina HiSeq2000 system, we sequenced RNA from two *E. arundinaceus* accessions from japan and one from Indonesia. All raw sequencing reads were assembled into contigs, which were used as reference sequence for transcriptome analysis in each genotype. SSR-containing regions were used to design about 2,500 primer pairs, of which a total of 672 annotated regions were selected for primer synthesis. Among these primer paires, 344 primer pair were selected through evaluation for reproducibility of amplication and polymorphism of detected fragments with wild *E. arundinaceus* accession in Thailand. Some of the SSR primer the SSR primer pairs from *E. arundinaceus* selected in the screening were applied to *Saccharum* and *Miscanthus* species

Key words: *Erianthus arundinaceus*, SSR, RNA

บทคัดย่อ

การพัฒนาเครื่องหมายโมเลกุล SSR จากข้อมูลลำดับนิวคลีโอไทด์ของอ้อยป่า *E. arundinaceus* ด้วยเทคโนโลยีการวิเคราะห์ลำดับนิวคลีโอไทด์ยุคใหม่ next generation sequencing (NGS) ได้นำลำดับนิวคลีโอไทด์ของอ้อยป่าจากญี่ปุ่น 2 หมายเลข และอินโดนีเซีย 1 หมายเลข พบจำนวน read (เป็นลำดับ

นิวคลีโอไทด์สายสั้นๆ) และตำแหน่งนิวคลีโอไทด์ทับซ้อนกัน (contig) ใช้เป็นลำดับนิวคลีโอไทด์อ้างอิง เพื่อวิเคราะห์ลำดับนิวคลีโอไทด์ในแต่ละจีโนไทป์ บริเวณที่เป็นเครื่องหมายโมเลกุล SSR โดยใช้ออกแบบไพร์เมอร์จำนวนท้งหมด 2,500 คู่ ข้อมูลทั้งหมด 672 ตำแหน่ง เพื่อใช้ในการคัดเลือกสำหรับสังเคราะห์ไพร์เมอร์ จากนั้นเลือกไพรเมอร์ 344 คู่ เพื่อใช้ประเมินสำหรับการเพิ่มปริมาณและตรวจสอบความแตกต่างของขนาดดีเอ็นเอของอ้อยป่า *E. arundinaceus* ในประเทศไทย โดยมีบางคู่ไพร์เมอร์ของยีน SSR จาก *E. arundinaceus* ถูกใช้ในการคัดเลือกอ้อยในกลุ่ม *Saccharum* และ *Miscanthus*

คำหลัก : อ้อยป่า, เครื่องหมายโมเลกุล, นิวคลีโอไทด์

คำนำ

อ้อย *Erianthus arundinaceus* จัดอยู่ในกลุ่มอ้อย *Saccharum* ในประเทศไทยพืชสกุลนี้ เรียกว่า เลา พบทั่วไปทั้งที่ราบ เชิงเขา บนภูเขาริมฝั่งน้ำและที่ลุ่มน้ำท่วม แยกตามแหล่งที่พบออกเป็น 2 ชนิดใหญ่ๆ คือ พวกที่อยู่ในที่ดอนและพวกที่อยู่ในที่ลุ่มหรือริมฝั่งแม่น้ำ พวกที่อยู่ในที่ดอนจะมีจำนวนต้นต่อกอมาก กอแน่น กาบใบมีขนมาก ไส้กลางคล้ายฟองน้ำมีขนาดใหญ่ เนื้อไม้บางของเหลวในลำต้นมีน้อย ส่วนพวกที่อยู่ริมฝั่งแม่น้ำหรือในที่ลุ่ม พบมากในภาคใต้ ลำต้นมีขนาดใหญ่ กอไม่แน่น และกาบใบไม่มีขนหรือมีเล็กน้อย ไส้กลางเล็กกว่าชนิดที่พบในที่ดอน เนื้อไม้หนากว่าและของเหลวในลำต้นมีมากกว่า โดยอ้อยชนิดนี้มีความน่าสนใจและเป็นชนิดที่มีศักยภาพในการปรับปรุงลักษณะพันธุกรรมและให้เป็นพืชในการผลิตพลังงานจากธรรมชาติ การวิเคราะห์ลักษณะพันธุกรรมจากแหล่งเชื้อพันธุกรรมของอ้อย *E. arundinaceus* จึงมีวัตถุประสงค์เพื่อการพัฒนาเครื่องหมายโมเลกุล SSR สำหรับอ้อย *E. arundinaceus*

อุปกรณ์และวิธีการ

ออกแบบไพร์เมอร์บริเวณ SSR ข้อมูลจาก NGS โดยใช้ Illumina HiSeq2000 system วิเคราะห์ลำดับนิวคลีโอไทด์อ้อยป่าจากญี่ปุ่น 2 accession และอินโดนีเซีย 1 หมายเลข ตำแหน่งนิวคลีโอไทด์ทับซ้อนกัน (contig) ใช้เป็นลำดับเบสอ้างอิง วิเคราะห์ลำดับนิวคลีโอไทด์ในแต่ละจีโนไทป์ โดยบริเวณที่เป็นเครื่องหมายโมเลกุล SSR ออกแบบไพร์เมอร์จำนวน 2,500 คู่ ใช้คัดเลือกสำหรับสังเคราะห์ไพร์เมอร์ทั้งหมด 672 ตำแหน่ง

คัดเลือกไพร์เมอร์จาก 672 ตำแหน่ง ใช้ในการคัดเลือกกับอ้อยป่า *E. arundinaceus* (4 หมายเลข) จากประเทศไทย โดยเพิ่มปริมาณสารพันธุกรรมด้วยวิธีพีซีอาร์และคัดเลือกด้วยเจลอิเล็คโตฟอริซิส หลังจากการคัดเลือกประเมินการเพิ่มปริมาณและตรวจสอบความแตกต่างของชิ้นส่วนดีเอ็นเอกับอ้อยป่า *E. arundinaceus* 8 หมายเลข โดยใช้เครื่อง ABI3500 Genetic Analyzer

การคัดเลือก SSR ไพร์เมอร์เพื่อออกแบบในการศึกษานี้ ได้ทดสอบกับอ้อยกลุ่มอื่นๆ 16 จีนัส ที่มีความสัมพันธ์ใกล้ชิดกัน ได้แก่ *Saccharin* (9 หมายเลข), *Miscanthus* (9 หมายเลข), และพันธุ์ลูกผสม *Saccharum* และ *Erianthus* (3 หมายเลข)

ผลการทดลองและวิจารณ์

การวิเคราะห์ลำดับนิวคลีโอไทด์และระบุเครื่องหมายโมเลกุล SSR จากการรวม raw reads ให้เป็นชิ้นดีเอ็นเอที่ยาวขึ้น (contig) ได้ 217,002 contig ผ่านการคัดกรองจีโนมด้วยวิธีการโดยการเปรียบเทียบกับฐานข้อมูลลำดับเบสของ NCBI ด้วยโปรแกรม BLAST อ้างอิงลำดับเบสของยีน *E. arundinaceus* และ *Sorghum bicolor* ทั้งหมด 5,910 ตำแหน่งของ SSR บ่งชี้จากตำแหน่งนิวคลีโอไทด์ทับซ้อนกัน (contig) ด้วยซอร์ฟแวร์ MSATCOMMANDER v.1.03 (Faircloth 2008) จากการคัดกรอง 5,910 ตำแหน่ง ใช้หลักการจากการอ่านช่วงความยาวของลำดับเบสมากกว่า 150 bp และตำแหน่ง SSR ภายในข้อมูลชิ้นลำดับเบส ออกแบบคู่เบสสำหรับ 2,549 loci ข้อมูลคู่ไพรเมอร์ทั้งหมด 672 ตำแหน่ง มีข้อมูลใช้ประกอบการคัดเลือกสำหรับใช้ในการสังเคราะห์ไพรเมอร์ โดย regions คัดเลือกจากการแสดงของข้อมูลนิวคลีโอไทด์แบบ di-hexanucleotide

การออกแบบไพรเมอร์ และการประเมินสำหรับอ้อย *Erianthus* species ในประเทศไทย จากการออกแบบไพรเมอร์จำนวน 627 คู่ ใช้คัดเลือกในอ้อยป่า *Erianthus* จากเชื้อพันธุกรรมในประเทศไทย มีการคัดด้วยเจลอิเล็คโตฟอริซิสได้ไพรเมอร์ 540 คู่ จากการเพิ่มปริมาณส่วนของ fragment ด้วยวิธีพีซีอาร์ จากอ้อยป่า *E. arundinaceus* 4 accessions การทำงานของคู่ไพรเมอร์ตรวจสอบใน 8 accessions ของอ้อย *Erianthus* ประกอบด้วย 4 hexaploid *E. arundinaceus* (2*n*=60), 2 tetraploid *E. arundinaceus* (2*n*=40), และ 2 *E. proceus* (2*n*=40) โดยการวิเคราะห์ชิ้นส่วนของดีเอ็นด้วยเครื่อง ABI3500 Genetic Analyzer ไพรเมอร์มีการเพิ่มปริมาณ 344 ชิ้นดีเอ็นเอในทุกหมายเลข ที่ใช้ในการทดสอบ ซึ่งตรวจสอบได้ 318 polymorphic loci ไพรเมอร์ส่วนใหญ่เพิ่มปริมาณได้ 1 ถึง 3 ของชิ้นดีเอ็นเอในอ้อยป่า *Erianthus* การกระจายตัวของจำนวนของการเพิ่มปริมาณชิ้นส่วนดีเอ็นเอคล้ายกับการทดสอบหมายเลขในการศึกษานี้ ข้อแนะนำนั้นจำนวนของอัลลีลที่ SSR loci มีบริเวณที่มีความอนุรักษ์สูงใน Hexaploidy และ tetraploidy ของยีนอ้อยพันธุ์ *Erianthus*

จากการทดสอบการใช้ไพรเมอร์ 12 คู่ คัดเลือกโดยการสุ่มข้ามพบว่า 16 Saccharum complex ได้แก่ กลุ่ม *Saccharum* และ *Miscanthus* พบว่ามี 6 คู่ไพรเมอร์ เพิ่มปริมาณส่วนชิ้นส่วนดีเอ็นเอของอ้อย *Miscanthus* ได้ 5 คู่ไพรเมอร์ โดยตรวจสอบได้ 1-4 อัลลีลต่อโลคัส ในอ้อย *Saccharum* และ 1-2 อัลลีลต่อโลคัสในอ้อย *Miscanthus*

สรุปผลการทดลอง

ไพรเมอร์ทั้งหมด 344 คู่ เพิ่มปริมาณ locus ประกอบด้วยตำแหน่ง repeat ที่เป็น di-hexanucleotide พัฒนาจากข้อมูลการหาลำดับนิวคลีโอไทด์ในอ้อย *E. arundinaceus* ชุดของ polymorphic loci จำนวน 318 loci ในหมวดของแหล่งพันธุกรรมสามารถใช้ประโยชน์เพื่อประเมินความสัมพันธ์ทางพันธุกรรมของเชื้อพันธุกรรมอ้อย *Erianthus* สำหรับวิเคราะห์ linkage ส่วนของเครื่องหมายโมเลกุลที่มีการ cross-transportability สามารถใช้ประโยชน์เพื่อวัตถุประสงค์ในการประเมินลูกผสมแบบ interspecific หรือ intergenric ได้

เอกสารอ้างอิง

Faircloth B. 2008. MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. Mol Ecol Resour 8:92–94.

Table 1 Characteristics 672 primer pair designed in this study.

Repeat type	No. of SSR	No. of repeat	Average no. of repeat	Ratio (%)
Dinucleotide	49	8-11	8.7	7.3
Trinucleotide	521	5-7	5.3	77.5
Tetranucleotide	76	4-11	4.3	11.3
Pentranucleotide	18	4-5	4.1	2.7
Hexanucleotide	8	4-8	4.5	1.2

Table 2 Result of screening for 672 primer pairs.

Screening	Method*	Accession tested		
		No.	Origin	Selected primers
1st	Annotation	2	Japan	672
2nd	AGE	4	Thailand	540
3rd	FA	10	Thailand	344

AGE: Agarose gel electrophoresis, FA: Fragment analysis

การวิเคราะห์เปรียบเทียบลำดับนิวคลีโอไทด์จีโนมคลอโรพลาสของอ้อย *Erianthus*

Comparative analysis of complete chloroplast genomes in *Erianthus* species

Shinichi Tsuruta ศุภรัตน์ ศรีทวงษ์ เบญจวรรณ รัตวัตร วีรกรณ์ แสงไสย์ Masumi Ebina
Makoto Kobayashi Terashima Yoshifumi อัมราวรรณ ทิพยวัฒน์ ศุจิรัตน์ สงวนรังศิริกุล
S. Tsuruta S. Srithawong B. Ruttawat W. Saengsai M. Ebina
M. Kobayashi Y. Terajima A. Tippayawat S. Sakuanrungsirikul

ABSTRACT

We determined the complete chloroplast genome sequence of *E. procerus*. Alignment of this sequence with that previously reported for *E. arundinaceus* demonstrated a high degree of conservation in gene content and order. We found indel and SNP variation between these species} which could be useful for species identification and analysis of their evolution. We use these data to characterize reproducibility of inter-specific varieties and phylogenetic relationships in the Thai *Erianthus*. This is the first report of phylogenetic and evolutionary relationships inferred from maternally inherited variation comprehensively detected within *Erianthus* species. Our study provides an important framework for understanding the phylogenetic and evolutionary aspects of economically important genera *Erianthus*.

Key words: *Erianthus*, Chloroplast, genome

บทคัดย่อ

การศึกษาลำดับนิวคลีโอไทด์จีโนมคลอโรพลาสของอ้อย *E. procerus* โดยเปรียบเทียบลำดับนิวคลีโอไทด์กับอ้อยพันธุ์ *E. arundinaceus* ซึ่งมีตำแหน่งอนุรักษ์ของยีนสูง จากการศึกษาพบว่า indel และ SNP มีความแปรปรวนระหว่างชนิด จึงนำมาใช้สำหรับการระบุชนิดและการวิเคราะห์วิวัฒนาการโดยนำข้อมูลลักษณะการถ่ายทอดแบบจำเพาะภายในสายพันธุ์และความสัมพันธ์และความสัมพันธุ์ของสายวิวัฒนาการในอ้อย *Erianthus* ของไทย การรายงานนี้ เป็นรายงานครั้งแรกที่ชี้ให้เห็นถึงความสัมพันธ์ทางพันธุกรรมและสายวิวัฒนาการความสัมพันธ์ของลักษณะที่ได้รับการถ่ายทอดจากต้นแม่ในอ้อยพันธุ์ *Erianthus* การศึกษาครั้งนี้ทำให้ทราบถึงความสัมพันธ์ทางพันธุกรรมและสายวิวัฒนาการของอ้อยพันธุ์ *Erianthus*

คำหลัก : อ้อยป่า, คลอโรพลาส, จีโนม

คำนำ

อ้อย *Erianthus* จัดอยู่ใน Groups of Subtribe: Saccharinae ภายใน Tribe Andropogoneae ประกอบด้วย 7 สปีชีส์ที่มีความสัมพันธ์ใกล้ชิดกัน ในประเทศไทยพืชสกุลนี้เรียกว่า เลา พบทั่วไปทั้งที่ราบ เชิงเขา บนภูเขาริมฝั่งน้ำและที่ลุ่มน้ำท่วม แยกตามแหล่งที่พบออกเป็น 2 ชนิดใหญ่ๆ คือ พวกที่อยู่ในที่ดอนและพวกที่อยู่ในที่ลุ่มหรือริมฝั่งแม่น้ำ พวกที่อยู่ในที่ดอนจะมีจำนวนต้นต่อกอมาก กอแน่น กาบใบมีขนมาก ไส้กลางคล้ายฟองน้ำมีขนาดใหญ่ เนื้อไม้บางของเหลวในลำต้นมีน้อย ส่วนพวกที่อยู่ริมฝั่งแม่น้ำหรือในที่ลุ่ม พบมากในภาคใต้ ลำต้นมีขนาดใหญ่ กอไม่แน่นและกาบใบไม่มีขนหรือมีเล็กน้อย ไส้กลางเล็กกว่าชนิดที่พบในที่ดอน เนื้อไม้หนากว่าและของเหลวในลำต้นมีมากกว่าอ้อย *Erianthus* ได้ถูกใช้เพื่อศึกษาความสัมพันธ์ทางพันธุกรรมของอ้อยในกลุ่ม Saccharum เพื่อพัฒนาและปรับปรุงพันธุ์ให้เป็นพืชพลังงานชีวภาพ ในประเทศไทยมีอ้อยป่าอยู่ 2 ชนิด ได้ถูกจัดจำแนกเป็นอ้อย *E. arundinaceus* ในความแตกต่างของ ploidy level (2*n*=40,60) และอ้อย *E. procerus* (2*n*=40) ได้รวบรวมทรัพยากรพันธุกรรมเหล่านี้ทั่วทั้งประเทศไทย จากปี 2533 จำนวนของอ้อยที่ใช้ศึกษาบนลักษณะทางพันธุกรรมและการปรับปรุงพันธุ์อ้อย *E. arundinaceus* ที่ได้มีการรายงาน ส่วนอ้อย *E. procerus* มีการศึกษาน้อยมาก ในการศึกษาครั้งนี้ เพื่อความเข้าใจยิ่งขึ้นของความสัมพันธ์ทางพันธุกรรมและสายวิวัฒนาการของอ้อย *Erianthus* จากการหาลำดับนิวคลีโอไทด์ complete chloroplast genome ของอ้อย *E. procerus* และเปรียบเทียบลำดับนิวคลีโอไทด์กับอ้อย *E. arundinaceus*

อุปกรณ์และวิธีการ

การวิเคราะห์ลำดับนิวคลีโอไทด์จีโนมคลอโรพลาสของอ้อย *E. procerus* ด้วยวิธี Sanger sequencing จากปฏิกิริยาพีซีอาร์ ทำการออกแบบไพรเมอร์จำนวน 16 คู่ จากลำดับนิวคลีโอไทด์จีโนมคลอโรพลาสของอ้อย *E. arundinaceus* (Tsuruta et al. 2017) ใช้สำหรับเพิ่มปริมาณนิวคลีโอไทด์จีโนมคลอโรพลาสขของอ้อย *E. procerus* หาลำดับนิวคลีโอไทด์จากการเพิ่มปริมาณด้วยปฏิกิริยาพีซีอาร์ โดยใช้ไพรเมอร์ทั้งหมด 258 คู่ ใช้สำหรับหาลำดับนิวคลีโอไทด์จีโนมคลอโรพลาสทั้งหมดของอ้อย *Miscanthus* (Tsuruta et al. 2017)

เปรียบเทียบ whole-genome ในอ้อย *Erianthus* โดยนำลำดับนิวคลีโอไทด์ทั้งหมดของจีโนมคลอโรพลาสของอ้อย *E. procerus* วิเคราะห์ด้วยซอร์ฟแวร์ Dual Oraganellar GenoMe Annotator (DOGMA) เปรียบเทียบกับลำดับนิวคลีโอไทด์จากอ้อย *E. arundinaceus* และวาดแผนที่นิวคลีโอไทด์จีโนมคลอโรพลาส ด้วยซอร์ฟแวร์ GenomeVx

ผลการทดลองและวิจารณ์

ลำดับนิวคลีโอไทด์จีโนมคลอโรพลาสของอ้อย *E. procerus* (141, 215 bp) มีลักษณะเป็นโครงสร้างเป็นวงกลม โดยนิวคลีโอไทด์จีโนมคลอโรพลาส ประกอบด้วย large single-copy (LSC) (83,174 bp) และบริเวณ small single-copy (SSC) (12,515 bp) ซึ่งถูกคั่นด้วยคู่ของ IRs (IRa และ IRb แต่ละส่วนขนาด 22,763 bp) มีปริมาณค่า GC 34 เปอร์เซ็นต์ ในอ้อย *E. procerus* ซึ่งคล้ายกับค่าใน Panicoideae อื่นๆ ซึ่งในอ้อย *E. arundinaceus* มีปริมาณค่า GC 39 เปอร์เซ็นต์ จำนวนยีนของ *E. procerus* มีทั้งหมด 136 ยีน ประกอบด้วย protein-coding 87 ยีน ในจีโนมมี ribosomal RNA (rRNA) 8 ยีน และ transfer RNA (tRNA) 41 ยีน โดยยีน coding 58 เปอร์เซ็นต์ (**Table 1**) และยังพบว่า start codons ของยีน *rpf2* และ *rps19* มีแนวโน้มที่จะเป็น ACG และ GTG โดยมีการแก้ไขยีน RNA ในระหว่างการ translation ในจีโนมของ *E. procerus* เป็นการรายงานในอ้อย *E. arundinaceus*

ความผันแปรของลำดับนิวคีลโอไทด์ของจีโนมอ้อย *E. arundinaceus* จากการเปรียบเทียบลำดับนิวคลีโอไทด์ จากการการศึกษานี้เปรียบเทียบกับฐานข้อมูลของ *E. arundinaceus* ใน GenBank โดยพบความแปรปรวนของลำดับเบสโดยมีการบ่งชี้ 31 ที่ ซึ่งมี 11 ที่ ตรวจพบที่บริเวณ coding ในขณะที่เปรียบเทียบของอ้อย *Erianthus* ตรวจพบ indel และ SNP มีความแปรปรวนระหว่างชนิดในยีน *atp*I และ *rpo*A ในการเปลี่ยนของลำดับนิวคลีโอไทด์ของกรดอะมิโน การศึกษาครั้งนี้ทำให้เข้าใจถึงความสัมพันธ์ทางพันธุกรรมและสายวิวัฒนาการของอ้อยสกุล *Erianthus* ที่สำคัญทางเศรษฐกิจ

สรุปผลการทดลอง

การศึกษาลำดับนิวคลีโอไทด์จีโนมคลอโรพลาสของอ้อย *E. procerus* โดยเปรียบเทียบลำดับนิวคลีโอไทด์กับอ้อยพันธุ์ *E. arundinaceus* ซึ่งมีตำแหน่งอนุรักษ์ของยีนสูงระหว่างชนิดอ้อย เป็นการรายงานครั้งแรกที่ชี้ให้เห็นถึงความสัมพันธ์ทางพันธุกรรมและสายวิวัฒนาการความสัมพันธ์ของลักษณะที่ได้รับการถ่ายทอดจากต้นแม่ในอ้อยพันธุ์ *Erianthus* การศึกษาครั้งนี้ทำให้ทราบถึงความสัมพันธ์ทางพันธุกรรมและสายวิวัฒนาการของอ้อยพันธุ์ *Erianthus*

เอกสารอ้างอิง

Tsuruta S-i, Ebina M, Kobayashi M, Takahashi W (2017) Complete Chloroplast Genomes of *Erianthus arundinaceus* and *Miscanthus sinensis*: Comparative Genomics and Evolution of the *Saccharum* Complex. PLoS ONE 12(1): e0169992.

Table 1 Characteristics of the chloroplast genome in the *E. procerus*

Species	Genome organization[1]					Number of genes[2]				Reference
	Length of total genome	Length of LSC	Length of SSC	Length of IR	GC/AC contents (%)	Total	CDS	rRNA	tRNA	
E. procerus	141,215	83,174	12,515	22,763	38.4/61.6	136	87 (8)	8 (4)	41 (8)	This study
E. arundinaceus	141,210	83,170	12,516	22,762	38.5/61.5	136	87 (8)	8 (4)	41 (8)	Tsuruta et al. 2017
Saccharum spp.	121,182	83,048	12,544	22,795	38.4/61.6	136	87 (8)	8 (4)	41 (8)	Asano et al. 2004

[1]Length is included in base pairs.

[2]The numbers of duplicated genes are shown in parentheses