

อิทธิพลของแบบจำลองทางสถิติและจำนวนเครื่องหมาย  
ที่ส่งผลต่อความแม่นยำของการทำนายค่าจีโนม  
สำหรับลักษณะผลผลิตน้ำยางในต้นยางพารา  
Effect of Statistical Models and Number of Markers on  
the Accuracy of the Genomic Prediction for Latex Yield  
in Rubber Trees (*Hevea brasiliensis* Muell. Arg.)

ภัทรวี นิลพลับ และกิตติพัฒน์ อุโฆษกิจ\*

สาขาวิชาเทคโนโลยีชีวภาพ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์

ศูนย์วิจัย ตำบลคลองหนึ่ง อำเภอคลองหลวง จังหวัดปทุมธานี 12120

Patrawee Nilplub and Kittipat Ukoskit\*

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

Rangsit Centre, Khlong Nueng, Khlong Luang, Pathum Thani 12120

Received: April 16, 2020; Accepted: April 28, 2020

## บทคัดย่อ

การทำนายค่าจีโนมเป็นวิธีการทำนายค่าการผสมพันธุ์จีโนมของลักษณะปริมาณในพืชและสัตว์ โดยใช้ความผันแปรทางพันธุกรรมในระดับนิวคลีโอไทด์ที่มีการกระจายอยู่ทั่วทั้งจีโนม ซึ่งสามารถลดระยะเวลาวงจรการปรับปรุงพันธุ์พืชและสัตว์ ความแม่นยำของการทำนายค่าจีโนมขึ้นอยู่กับหลายปัจจัย การศึกษาที่จึงศึกษาผลกระทบของจำนวนเครื่องหมายและแบบจำลองทางสถิติที่มีผลต่อความแม่นยำการทำนายค่าจีโนมลักษณะผลผลิตน้ำยางในฤดูฝนและฤดูแล้งของยางพารา (*Hevea brasiliensis*) 170 ต้น เครื่องหมาย SNP จำนวน 14,155 เครื่องหมาย แบบจำลองทางสถิติที่ใช้ในการศึกษานี้ ได้แก่ ridge regression-best linear unbiased prediction (RR-BLUP) และ Bayesian LASSO (least absolute shrinkage and selection operator, BL) ซึ่งต่างกันที่ข้อกำหนดสมมติฐานเกี่ยวกับการกระจายตัวของอิทธิพลเครื่องหมาย ผลการศึกษาพบว่า RR-BLUP มีความแม่นยำในการทำนายที่สูงกว่า BL ทั้งลักษณะผลผลิตน้ำยางในฤดูฝนและฤดูแล้ง แสดงให้เห็นว่าลักษณะผลผลิตน้ำยางถูกควบคุมด้วยยีนหลายตำแหน่งไม่มียีนหลักที่ส่งผลต่อความแปรปรวนฟีโนไทป์ นอกจากนี้พบว่าความแม่นยำในการทำนายค่าจีโนมเพิ่มขึ้นเมื่อเพิ่มจำนวนเครื่องหมายขึ้นจนถึงจำนวนที่เหมาะสม ผลการศึกษานี้สามารถนำไปใช้เป็นแนวทางนำไปสู่การปรับปรุงพันธุ์ด้วยวิธีการคัดเลือกจีโนมในยางพารา

คำสำคัญ : ยางพารา; การทำนายค่าจีโนม; RR-BLUP; Bayesian LASSO; SNP

## Abstract

Genomic prediction is a method for predicting genomic breeding values (GEBVs) of quantitative traits for plants and animals, using nucleotide variation throughout the genome. This approach can increase genetic gains by accelerating the breeding cycle. Several factors affecting prediction accuracy should be well evaluated if breeders exploit genomic selection to its full potential. In this study, a panel of 170 natural rubber trees genotyped with 14,155 single nucleotide polymorphism (SNP) markers to investigate the effect of marker density and statistical models was examined for genomic prediction accuracy of latex yield in the dry (YD) and wet (YW) seasons. The performance of two different genomic prediction methods that differ with respect to assumptions regarding distribution of marker effects, including ridge regression-best linear unbiased prediction (RR-BLUP), and Bayesian LASSO (least absolute shrinkage and selection operator, BL) was evaluated. The predictive ability of the methods was evaluated using a cross-validation approach. RR-BLUP had higher predictive ability than BL for both YD and YW. This suggests that latex yield in the dry and wet seasons is controlled by many genes of equal contribution of all markers to the observed variation. Accuracy can be improved by increasing the optimal marker diversity. These findings represent a resource for plant breeders and contribute to the collective knowledge for genomic selection in rubber tree.

**Keywords:** rubber tree; genomic prediction; RR-BLUP, Bayesian LASSO; SNP

## 1. คำนำ

ยางพารา (*Hevea brasiliensis* Muell. Arg.) เป็นแหล่งผลิตยางธรรมชาติ (1,4 cis-polyisoprene) เพียงแหล่งเดียว 90 เปอร์เซ็นต์ ของผลผลิตยางธรรมชาติทั่วโลกที่ผลิตจากประเทศไทยและอินโดนีเซีย (Cros *et al.*, 2019) ยางพาราจึงเป็นพืชเศรษฐกิจที่สำคัญของประเทศไทย ซึ่งมีบทบาทสำคัญต่อชีวิตและความเป็นอยู่ของเกษตรกรไทย ปัจจุบันมีการขยายพื้นที่ปลูกยางเพิ่มมากขึ้น ความต้องการยางพันธุ์ใหม่ที่เหมาะสมจึงมีมากขึ้น ดังนั้นการคัดเลือกพันธุ์ยางที่สามารถให้ผลผลิตอย่างมีประสิทธิภาพจึงมีความจำเป็นและเป็นประโยชน์อย่างยิ่งต่อเกษตรกรในอนาคต

ต้นยางพารามีจีโนมเป็นดิพลอยด์ ( $2n = 36$ ) โดยจีโนมแฮพลอยด์มีขนาดประมาณ 2.15 Gb มีถิ่นกำเนิดบริเวณลุ่มน้ำแอมะซอน ประเทศบราซิล

และเปรู ทวีปอเมริกาใต้ โดยทั่วไปต้นยางเปิดกรีดได้เมื่ออายุประมาณ 7 ปีครึ่ง ขนาดเส้นรอบต้นไม่ต่ำกว่า 50 เซนติเมตร ความสูง 150 เซนติเมตร จากพื้นดิน (พิชิต และคณะ, 2550) ดังนั้นจึงต้องใช้ระยะเวลายาวนานในการทดสอบและคัดเลือกพันธุ์เพื่อให้แน่ใจว่าลูกผสมที่ได้มีผลผลิตตามที่ต้องการ ซึ่งหากผลลัพธ์ที่ได้ไม่เป็นไปตามความคาดหวังจะทำให้เสียเวลาในการพัฒนาเพื่อหาพันธุ์ใหม่ที่มีประสิทธิภาพตามต้องการ ความไม่แน่นอนของการคัดเลือกพันธุ์แบบมาตรฐานเกิดจากสมบัติที่แสดงออกจากการกำหนดโดยพันธุกรรมและสภาพแวดล้อม ดังนั้นการนำความรู้ทางพันธุศาสตร์ด้านการวิเคราะห์จีโนมมาใช้ควบคู่กับการคัดเลือกพันธุ์ด้วยวิธีแบบมาตรฐาน จะช่วยให้เกิดความมั่นใจในผลการคัดเลือกพันธุ์ว่าจะได้พันธุ์ที่มีสมบัติตามความคาดหวังในระยะเวลาที่สั้นลง

การปรับปรุงพันธุ์ลักษณะปริมาณโดยใช้เครื่องหมายดีเอ็นเอช่วยในการคัดเลือก (marker-assisted selection, MAS) ปัจจุบันส่วนใหญ่ใช้วิธีการค้นหาเครื่องหมายที่มีอิทธิพลสูง และใช้เครื่องหมายที่ค้นหาได้ช่วยในการคัดเลือก วิธีนี้อาจมีประสิทธิภาพน้อยเมื่อใช้กับการคัดเลือกลักษณะปริมาณที่มีจำนวนยีนเกี่ยวข้องกับจำนวนมากและไม่มียีนที่มีอิทธิพลหลัก เช่น ลักษณะผลผลิตพืช การทำนายค่าจีโนมหรือการคัดเลือกจีโนมเป็นวิธีหนึ่งในการพัฒนาล่าสุดที่ใช้เครื่องหมายทางพันธุกรรมในการคัดเลือก โดยใช้ข้อมูลเครื่องหมายโมเลกุลที่มีการกระจายทั่วทั้งจีโนม (ไม่ใช่เฉพาะเครื่องหมายโมเลกุลบริเวณของจีโนมที่สนใจที่สัมพันธ์กับลักษณะที่ต้องการคัดเลือก) ติดตามและใช้ทำนายค่าการผสมพันธุ์และลักษณะปรากฏหรือฟีโนไทป์ (Meuwissen *et al.*, 2001) การทำนายค่าจีโนมสามารถทำให้ข้อมูลพันธุกรรมมีประโยชน์ในโปรแกรมการปรับปรุงพันธุ์พืชได้โดยตรง เพราะสามารถวิเคราะห์จีโนมได้ตั้งแต่พืชยังเป็นต้นอ่อนและทำนายฟีโนไทป์ได้จากการทำนายค่าจีโนม (Grattapaglia, 2017) โดยการสร้างแบบจำลองทางสถิติเพื่อใช้ในการทำนายจะใช้ข้อมูลจีโนมและฟีโนไทป์จากประชากรฝึกหรือประชากรตัวอย่าง (training population) ในการสร้างแบบจำลองทางสถิติ และสามารถนำแบบจำลองทางสถิติที่สร้างได้มาใช้ทำนายค่าการผสมพันธุ์จีโนมและฟีโนไทป์ของประชากรปรับปรุงพันธุ์ (breeding population) อื่นต่อไป

วิธีการทางสถิติหรือแบบจำลองทางสถิติที่นิยมใช้สำหรับการทำนายค่าจีโนมทั่วไป เช่น regression best linear unbiased predictor (RR-BLUP), Bayesian least absolute shrinkage and selection operator (LASSO) (BL) (Park and Casella, 2008), reproducing kernel Hilbert space (RKHS) (Gianola and van Kaam, 2008) RR-

BLUP และ BL เป็นสถิติพาราเมตริกที่มีการกำหนดสมมติฐานของความสัมพันธ์ระหว่างจีโนมไทป์และฟีโนไทป์เช่นเดียวกัน แต่ต่างกันที่การกระจายตัวของค่าประมาณอิทธิพลเครื่องหมาย (marker effect) RR-BLUP กำหนดสมมติฐานให้ค่าประมาณอิทธิพลเครื่องหมายมีการกระจายตัวแบบปกติ (normally distribution) หรือทุกเครื่องหมายส่งผลต่อความแปรปรวนฟีโนไทป์เล็กน้อยและเท่ากัน (Meuwissen *et al.*, 2001) ส่วน BL กำหนดสมมติฐานให้ค่าประมาณอิทธิพลมีการแจกแจงอิสระที่มีความแปรปรวนไม่เท่ากัน หรือเครื่องหมายมีอิทธิพลต่อความแปรปรวนฟีโนไทป์ต่างกัน และมีการกระจายตัวอิทธิพลของเครื่องหมายแบบ double exponential (Park and Casella, 2008) และ RKHS เป็นตัวแบบกึ่งพาราเมตริก (semiparametric) และเป็นการถดถอยแบบไม่เป็นเส้นตรง (nonlinear regression) ตัวแบบกึ่งพาราเมตริกวิเคราะห์ได้ถึงอิทธิพลเนื่องจากพันธุกรรมแบบบวกสะสม (additive effect) และอิทธิพลเนื่องจากพันธุกรรมแบบไม่บวกสะสม (non-additive effect) (Howard *et al.*, 2014) โดย RKHS สามารถวิเคราะห์ถึงปฏิริยาการข่มระหว่างยีนต่างตำแหน่ง (epistatic interaction) ได้อีกด้วย (Gianola and van Kaam, 2008)

ความแม่นยำของการทำนายค่าจีโนมขึ้นอยู่กับหลายปัจจัย ได้แก่ ระดับของ linkage disequilibrium (LD) ระหว่างเครื่องหมายพันธุกรรมกับยีนลักษณะปริมาณ (quantitative trait loci, QTL) จำนวนตัวอย่างของพืชที่จะใช้เป็นประชากรตัวอย่างในการสร้างแบบจำลองทางสถิติ (training set, TS) ค่าอัตราพันธุกรรมของลักษณะที่ศึกษา (heritability) การกระจายตัวของ QTL (Hayes *et al.*, 2009) และความถี่และอิทธิพลของเครื่องหมายพันธุกรรม (Solberg *et al.*, 2008) การศึกษานี้มีวัตถุประสงค์เพื่อศึกษาปัจจัยที่มีผลต่อความแม่นยำในการ

ทำนายค่าจีโนม โดยศึกษาแบบจำลองทางสถิติและอิทธิพลของจำนวนเครื่องหมายพันธุกรรม ที่มีผลต่อความแม่นยำในการทำนายค่าจีโนมลักษณะผลผลิตน้ำยางของยางพารา

## 2. วิธีการวิจัย

### 2.1 ข้อมูลที่ใช้ในงานวิจัย

ข้อมูลจีโนไทป์และฟีโนไทป์ผลผลิตน้ำยางที่ใช้ในงานวิจัยนี้เป็นข้อมูลจากงานวิจัยก่อนหน้า (Chanroj *et al.*, 2017) พันธุ์ยางที่ใช้เป็นยางปามาจากประเทศมาเลเซียและประเทศไอวอรีโคสต์ ซึ่งมีต้นกำเนิดในเขตป่าแอมะซอนจาก 3 รัฐ ในประเทศบราซิล รัฐ Acre 14 สายพันธุ์ รัฐ Rondonia 91 สายพันธุ์ และรัฐ Mato Grosso 64 สายพันธุ์ รวม 170 สายพันธุ์ ต้นยางทั้งหมดนำมาขยายพันธุ์ที่ศูนย์วิจัยยางหนองคาย จังหวัดหนองคาย ซึ่งอยู่ในภาคตะวันออกเฉียงเหนือของประเทศไทย

#### 2.1.1 ข้อมูลฟีโนไทป์และการวิเคราะห์

ประเมินผลฟีโนไทป์จากลักษณะผลผลิตน้ำยาง โดยการกรีดยางที่เหนือระดับพื้นดิน 1.5 เมตร โดยกรีดครั้งต้นวันเว้นวัน ( $1/2 S d / 2$ ) ชั่งน้ำหนักน้ำยางที่ได้ในหน่วยกรัมต่อการกรีดเปรียบเทียบในช่วงฤดูฝนและฤดูแล้ง ทดสอบการแจกแจงแบบปกติของข้อมูลตามข้อตกลงเบื้องต้นในการวิเคราะห์ทางสถิติโดยใช้การทดสอบของ Kolmogorov-Smirnov ที่ระดับนัยสำคัญ 0.01 วิเคราะห์ความสัมพันธ์ลักษณะระหว่างฤดูฝนและฤดูแล้งด้วยโปรแกรม SPSS

#### 2.1.2 ข้อมูลจีโนไทป์

เครื่องหมายทางพันธุกรรมที่ใช้ในงานวิจัย คือ เครื่องหมาย SNP (single nucleotide polymorphism) จำนวน 14,155 เครื่องหมาย ซึ่งเป็นเครื่องหมายที่พัฒนาจากการใช้เทคโนโลยีการวิเคราะห์ลำดับเบสยุคใหม่ วิเคราะห์ลำดับเบสทรานสคริปโทรม (transcriptome sequencing) และผ่าน

การคัดกรองคุณภาพเครื่องหมายจากงานวิจัยก่อนหน้า (Chanroj *et al.*, 2017)

### 2.2 ปัจจัยที่มีผลต่อความแม่นยำการทำนายค่าจีโนมแบบจำลองทางสถิติ

งานวิจัยนี้ใช้แบบจำลองทางสถิติ 2 แบบ คือ RR-BLUP และ BL แบบจำลองทางสถิติทั้งสองต่างกันที่สมมติฐานการแจกแจงก่อน (prior distribution) ของอิทธิพลของเครื่องหมาย SNP โดยมีรูปแบบแบบจำลองพื้นฐานที่เหมือนกัน คือ

$$y = \mu + Xg + e$$

โดย  $y$  คือ เวกเตอร์ของค่าสังเกต  $\mu$  คือ เวกเตอร์ค่าเฉลี่ยประชากร (ค่าเฉลี่ยค่าสังเกตโดยรวมของ training set แต่ละเครื่องหมาย SNP)  $X$  คือ เมทริกซ์ขนาด  $n \times m$  แสดงอิทธิพลเครื่องหมาย SNP จำนวน  $m$  เครื่องหมาย โดยที่  $X_{ij} = -1, 0, 1$  เมื่อจีโนไทป์ของตัวอย่างที่  $i$  เครื่องหมาย SNP ตำแหน่ง  $j$  เป็น AA, AB, BB ตามลำดับ  $n$  คือ จำนวนตัวอย่าง  $m$  คือ จำนวนเครื่องหมาย  $g$  คือ เวกเตอร์ขนาด  $m \times 1$  อิทธิพลของเครื่องหมาย SNP แต่ละตำแหน่ง  $e$  คือ เวกเตอร์ค่าความคลาดเคลื่อน

วิธี RR-BLUP กำหนดสมมติฐานให้  $g$  มีการกระจายตัวแบบปกติ โดย  $g \sim N(0, \sigma_g^2)$  ส่วนวิธี BL กำหนดสมมติฐานให้  $g$  มีการแจกแจงอิสระที่มีความแปรปรวนไม่เท่ากันและมีการกระจายตัวอิทธิพลของเครื่องหมายแบบ double exponential (DE)

$$DE(g_j | \lambda) = \prod_{j=1}^p (\lambda/2) \exp(-\lambda|g_j|)$$

การกระจายตัวของ DE ประกอบด้วย การประมาณค่าอิทธิพลเครื่องหมาย SNP ค่า (ค่าอิทธิพลเครื่องหมายเข้าใกล้ศูนย์) จำนวนมาก และประมาณค่าอิทธิพลเครื่องหมาย SNP ระดับปานกลางถึงสูงจำนวนน้อย (Campos *et al.*, 2009)

#### 2.2.1 ชุดเครื่องหมาย SNP

คัดเลือกเครื่องหมายแบบสุ่มตั้งแต่ 50-14,150 เครื่องหมาย เพื่อศึกษาอิทธิพลของ

จำนวนเครื่องหมายพันธุกรรมต่อความแม่นยำการทำนายค่าจีโนม โดยเพิ่มจำนวนเครื่องหมายขึ้นครั้งละ 50 เครื่องหมาย เปรียบเทียบกับการใช้เครื่องหมายทั้งจีโนมจำนวน 14,155 เครื่องหมาย

### 2.2.2 การวิเคราะห์ความแม่นยำการทำนาย

ความแม่นยำการทำนายรายงานผลเป็นค่าเฉลี่ยและค่าคลาดเคลื่อนมาตรฐานจากการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ (correlation coefficient) ระหว่างค่าที่ทำนาย (predict) กับค่าที่สังเกตได้ (observ) ของการวิเคราะห์ 500 รอบ โดยแต่ละรอบจะแบ่งประชากรตัวอย่าง (170 ตัวอย่าง) เป็น 2 กลุ่ม คือ กลุ่มประชากรฝึกหัด (training set) 102 ตัวอย่าง (60 เปอร์เซนต์) และกลุ่มประชากรทดสอบ (validate set) 68 ตัวอย่าง (40 เปอร์เซนต์)

## 3. ผลการวิจัยและวิจารณ์

### 3.1 ข้อมูลฟีโนไทป์

การตรวจสอบรูปแบบการกระจายของข้อมูลฟีโนไทป์ พบว่าลักษณะผลผลิตน้ำยางไม่มีการกระจายตัวแบบปกติจึงต้องแปลงข้อมูลเพื่อแก้ไขให้ข้อมูลฟีโนไทป์เป็นไปตามข้อกำหนดเบื้องต้น โดยตัดแปลงข้อมูลเดิมโดยใช้ลอการิทึม (logarithmic transformation) ในฤดูฝนมีผลผลิตน้ำยาง 0.30-18.35 กรัม เฉลี่ย 4.52 กรัม ในฤดูแล้งมีผลผลิตน้ำยาง 0.64-31.34 กรัม เฉลี่ย 6.88 กรัม งานวิจัยก่อนหน้าของ Chanroj และคณะ (2017) ที่ใช้ข้อมูลในการศึกษาชุดเดียวกัน ผลการประเมินค่าอัตราพันธุกรรมแบบกว้าง ( $h^2$ ) ของทุกลักษณะพบว่า มีค่าอยู่ในระดับสูงปานกลาง โดยผลผลิตน้ำยางฤดูฝนและฤดูแล้งมีอัตราพันธุกรรมแบบกว้าง 52.16 และ 57.42 เปอร์เซนต์ ตามลำดับ แสดงให้เห็นว่าความแปรปรวนของทุกลักษณะเกิดขึ้นเนื่องจากได้รับอิทธิพลจากยีนมากกว่าสภาพแวดล้อม การคัดเลือกจีโนไทป์ที่มีประสิทธิภาพจะสามารถช่วย

เพิ่มความก้าวหน้าทางพันธุกรรม (genetic gain) ลักษณะผลผลิตน้ำยาง

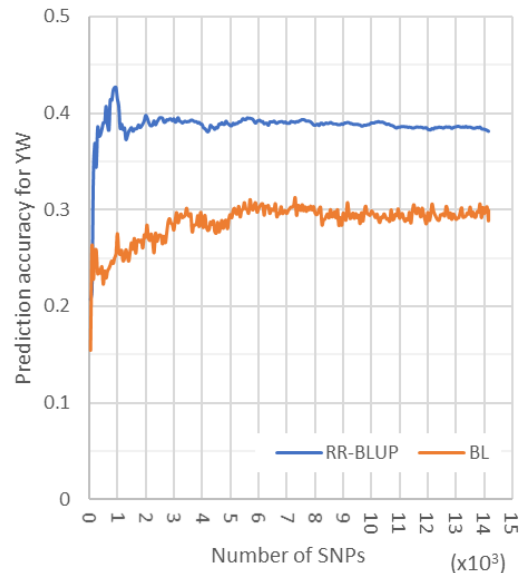
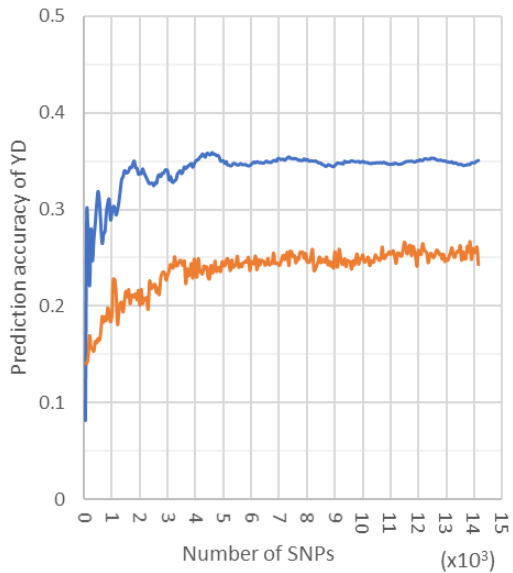
### 3.2 ปัจจัยที่มีผลต่อความแม่นยำการทำนายค่าจีโนม

#### 3.2.1 อัตราพันธุกรรม

การวิเคราะห์ความแม่นยำการทำนายค่าจีโนมโดยใช้ข้อมูลฟีโนไทป์ 60 เปอร์เซนต์ ของประชากรทั้งหมด (102 ตัวอย่าง) เครื่องหมาย SNP 14,155 เครื่องหมาย ในการสร้างแบบจำลองทางสถิติ (ตารางที่ 1) พบว่าความแม่นยำการทำนายค่าจีโนมสอดคล้องกับค่าอัตราพันธุกรรม โดยลักษณะผลผลิตน้ำยางในฤดูแล้งที่มีค่าอัตราพันธุกรรมที่สูงกว่าผลผลิตน้ำยางในฤดูฝนมีค่าความแม่นยำการทำนายค่าจีโนมที่สูงกว่าเช่นเดียวกันไม่ว่าจะใช้แบบจำลองทางสถิติแบบใด ความสัมพันธ์นี้สอดคล้องกับการศึกษาหลายงานก่อนหน้า (Hayes *et al.*, 2009; Duangjit *et al.*, 2016; Kwong *et al.*, 2017) ซึ่งสามารถอธิบายว่าลักษณะที่มีอัตราพันธุกรรมสูงมีความสัมพันธ์เชิงบวกกับความแม่นยำการทำนายค่าจีโนม จึงนำมาใช้คาดการณ์ความแม่นยำการทำนายค่าจีโนมเบื้องต้นได้ (Resende *et al.*, 2012)

ตารางที่ 1 ค่าเฉลี่ยความแม่นยำการทำนาย (500 ซ้ำ) ของลักษณะผลผลิตน้ำยางในฤดูฝน (YD) และฤดูแล้ง (YW) จากการใช้แบบจำลองสถิติที่ต่างกัน โดยใช้ SNP จำนวน 14,155 เครื่องหมาย

ลักษณะ	$h^2$ (%)	แบบจำลองทางสถิติ			
		RR-BLUP		BL	
		mean	SE	mean	SE
YD	52.16	0.35	0.001	0.25	0.087
YW	57.42	0.38	0.001	0.30	0.083



รูปที่ 1 ค่าเฉลี่ยความแม่นยำการทำนายค่าจีโนมลักษณะผลผลิตน้ำยางในฤดูฝน (YD) และฤดูแล้ง (YW) โดยใช้ชุดจำนวนเครื่องหมายที่ต่างกัน

### 3.2.2 แบบจำลองทางสถิติ

การเปรียบเทียบแบบจำลองทางสถิติ RR-BLUP และ BL ดังตารางที่ 1 และรูปที่ 1 พบว่า RR-BLUP ให้ความแม่นยำการทำนายค่าจีโนมสูงกว่า BL ความแม่นยำการทำนายลักษณะผลผลิตน้ำยางในฤดูฝนและฤดูแล้งจาก RR-BLUP มีค่า 0.35 และ 0.38 ตามลำดับ ซึ่งมากกว่า BL ที่ให้ความแม่นยำการทำนายลักษณะผลผลิตน้ำยางในฤดูฝนและฤดูแล้ง 0.25 และ 0.30 ตามลำดับ

มีหลายการศึกษาก่อนหน้านี้แสดงให้เห็นว่าวิธี Bayesian เช่น BL มักจะมีความแม่นยำมากกว่า RR-BLUP (Meuwissen *et al.*, 2001; Clark *et al.*, 2011; Honarvar *et al.*, 2013) แต่มีเหตุผลหนึ่งที่ผลการวิเคราะห์อาจไม่เป็นเช่นนั้น คือ ลักษณะนั้น ๆ อาจถูกควบคุมด้วย QTL ที่มีอิทธิพลต่อลักษณะต่ำและมีการกระจายอยู่ทั่วจีโนม (Daetwyler *et al.*, 2010) วิธี Bayesian มีสมมติฐานให้ QTL มีอิทธิพลต่อฟีโนไทป์ต่างกันขึ้นอยู่กับลักษณะที่ใช้ในการทำนาย (Park and Casella,

2008) จะมีประสิทธิภาพดีกว่า RR-BLUP เมื่อใช้ทำนายลักษณะฟีโนไทป์ที่เครื่องหมายมีอิทธิพลต่อ QTL มาก และจำนวนน้อย (Clark *et al.*, 2011) ส่วน RR-BLUP จะทำนายได้แม่นยำกว่าเมื่อใช้ทำนายลักษณะที่ถูกควบคุมด้วย QTL ที่มีอิทธิพลไม่สูงมากและถูกควบคุมด้วยเครื่องหมายจำนวนมาก โดย RR-BLUP มีสมมติฐานว่า QTL ทั้งหมดจะอธิบายถึงความแปรปรวนฟีโนไทป์เท่ากัน (Meuwissen *et al.*, 2001) อาจสรุปได้ว่าลักษณะผลผลิตน้ำยางถูกควบคุมด้วยยีนหลายตำแหน่ง ไม่มียีนหลักที่ส่งผลต่อความแปรปรวนของฟีโนไทป์และยีนควบคุมแต่ละตำแหน่งไม่ได้มีอิทธิพลมากและน้อยต่างกัน ทำให้ได้ค่าความแม่นยำการทำนายจาก RR-BLUP สูงกว่า เนื่องจาก RR-BLUP มีสมมติฐานว่าทุกเครื่องหมายส่งผลต่อความแปรปรวนของฟีโนไทป์เล็กน้อยและเท่ากัน (Meuwissen *et al.*, 2001)

### 3.2.3 จำนวนเครื่องหมาย

จำนวนเครื่องหมาย มีความสำคัญต่อ

ความแม่นยำการทำนายค่าจีโนม จากรูปที่ 1 แสดงความสัมพันธ์ระหว่างจำนวนเครื่องหมาย SNP ที่ใช้ในการสร้างแบบจำลองทางสถิติและความแม่นยำการทำนายค่าจีโนม พบว่าความแม่นยำการทำนายลักษณะผลผลิตน้ำยางทั้งในฤดูฝนและฤดูแล้งเพิ่มขึ้นเมื่อเพิ่มจำนวนเครื่องหมาย SNP จากการเพิ่มจำนวนเครื่องหมาย SNP ในช่วงแรกมีความแปรปรวนของความแม่นยำการทำนายค่าจีโนม อาจเกิดจากการคัดเลือกเครื่องหมายแบบสุ่มที่มีจำนวนไม่มากพอที่จะทำให้การทำนายมีความแม่นยำสูงแบบคงที่ และความแม่นยำการทำนายเริ่มคงที่เมื่อเพิ่มจำนวนเครื่องหมาย SNP จนถึงจำนวนหนึ่ง เช่นเดียวกันทั้งในการใช้แบบจำลองทางสถิติ RR-BLUP และ BL โดย RR-BLUP ให้ค่าความแม่นยำสูงที่สุดและเริ่มคงที่เมื่อใช้เครื่องหมาย SNP ตั้งแต่ 3,350 เครื่องหมาย สำหรับลักษณะผลผลิตน้ำยางในฤดูฝน และตั้งแต่ 1,000 เครื่องหมาย สำหรับลักษณะผลผลิตน้ำยางในฤดูแล้ง ซึ่งมีค่าความแม่นยำการทำนายค่าจีโนม 0.34 และ 0.41 ตามลำดับ ส่วน BL ให้ค่าความแม่นยำสูงที่สุดและเริ่มคงที่เมื่อใช้เครื่องหมาย SNP ตั้งแต่ 2,750 เครื่องหมาย สำหรับลักษณะผลผลิตน้ำยางในฤดูฝน และตั้งแต่ 4,250 เครื่องหมาย สำหรับลักษณะผลผลิตน้ำยางในฤดูแล้ง ซึ่งมีค่าความแม่นยำการทำนายค่าจีโนม 0.23 และ 0.28 ตามลำดับ

การคัดเลือกเครื่องหมายในการทดลองนี้เป็นการคัดเลือกแบบสุ่ม ดังนั้นจำนวนเครื่องหมายที่ถูกคัดเลือก ซึ่งมี LD กับยีนลักษณะปริมาณจึงมีลักษณะสุ่มเช่นกัน จึงอาจเป็นเหตุผลว่าค่าความแม่นยำสูงสุดยังไม่สูงมากนักในการทดลองนี้ เมื่อเปรียบเทียบกับงานวิจัยอื่นที่มีการคัดเลือกเครื่องหมายโดยวิธีแบบไม่สุ่ม เช่น การวิเคราะห์ค่าสัมประสิทธิ์ถดถอยเชิงเดี่ยว (simple linear regression) ของเครื่องหมายแต่ละเครื่องหมาย

(Macciotta *et al.*, 2009) การคัดเลือกโดยวิเคราะห์ค่าอิทธิพลของเครื่องหมาย (marker effect) (Kwong *et al.*, 2017) การคัดเลือกเครื่องหมายโดยเว้นระยะห่างระหว่างเครื่องหมายให้เท่ากันและกระจายทั่วจีโนม (Zhang *et al.*, 2011) นอกจากนี้การเพิ่มจำนวนเครื่องหมายมากขึ้นมีค่าความแม่นยำสูงที่สุดคงที่อาจเกิดจากจำนวนตัวแปรอิสระที่ใส่ในสมการถดถอยเป็นตัวแปรไม่มีความสัมพันธ์กับตัวแปรตามจึงทำให้ประสิทธิภาพของการทำนายไม่เพิ่มขึ้น

เป้าหมายของการศึกษาอิทธิพลจำนวนเครื่องหมาย SNP ในพืชที่ไม่ใช่พืชต้นแบบการศึกษาพันธุศาสตร์ ส่วนใหญ่แนวทางการศึกษาจะศึกษาเพื่อทดสอบความเป็นไปได้ของการลดจำนวนเครื่องหมายที่ใช้ในการทำนายค่าจีโนม การลดจำนวนเครื่องหมาย SNP ได้โดยไม่ลดความแม่นยำการทำนาย จะสามารถลดต้นทุนการวิเคราะห์จีโนมไทป์ในโครงการปรับปรุงพันธุ์ และมีประโยชน์อย่างมากสำหรับพืชที่ยังไม่มีระบบการวิเคราะห์จีโนมไทป์แบบประสิทธิภาพสูง (high throughput genotyping) เช่น SNP อาร์เรย์ (high density SNPs panel array) (Ballesta *et al.*, 2020) นอกจากนี้การที่สามารถลดจำนวนเครื่องหมายโดยที่ความแม่นยำไม่ลดลง ทำให้สามารถนำต้นทุนที่ลดไปใช้ในการเพิ่มจำนวนประชากรตัวอย่างที่ใช้ในการทำนายค่าจีโนม การมีประชากรฝึกหัดในการสร้างแบบจำลองทางสถิติมากจะทำให้การทำนายค่าจีโนมมีประสิทธิภาพในการคัดเลือกมากขึ้น (Cericola *et al.*, 2017)

#### 4.สรุป

อัตราพันธุกรรมแบบกว้างของลักษณะผลผลิตน้ำยางทั้งในฤดูฝนและฤดูแล้งที่มีค่าอยู่ในระดับสูงปานกลาง แสดงให้เห็นว่าความแปรปรวนของทุกลักษณะเกิดขึ้นเนื่องจากได้รับอิทธิพลจาก

ยืนมากกว่าสภาพแวดล้อม สัมพันธ์กับความแม่นยำ การทำนายค่าจีโนม การทำนายค่าจีโนมจะมีความแม่นยำเมื่อใช้ปัจจัยต่าง ๆ ที่มีอิทธิพลต่อความแม่นยำการสร้างแบบจำลองทางสถิติที่เหมาะสม การวิจัยนี้แสดงให้เห็นว่าการเปรียบเทียบแบบจำลองทางสถิติสำหรับการสร้างแบบจำลองสถิติการทำนายค่าจีโนม มีผลต่อความแม่นยำของการทำนาย โดยพบว่า RR-BLUP มีความแม่นยำการทำนายที่สูงกว่า BL ทำให้สรุปได้ว่าลักษณะผลผลิตนี้ น่าจะถูกควบคุมด้วยยีนหลายตำแหน่ง ไม่มียีนหลักที่ส่งผลต่อความแปรปรวนของฟีโนไทป์ ดังนั้นกล่าวได้ว่าการปรับปรุงพันธุ์ลักษณะผลผลิตนี้ของยางพาราโดยใช้เทคโนโลยีการคัดเลือกจีโนม (genomic selection) เป็นอีกแนวทางหนึ่งที่เหมาะสมมากกว่าการใช้เครื่องหมายช่วยคัดเลือก (MAS) แต่การใช้เครื่องหมายจำนวนมากที่ครอบคลุมทั้งจีโนมในการคัดเลือกจีโนมจะทำให้ต้นทุนการคัดเลือกสูง จึงไม่เหมาะสมกับการใช้งานจริงในโครงการปรับปรุงพันธุ์ยางพารา การศึกษาที่พบว่าการใช้จำนวนเครื่องหมายจำนวนมากที่ครอบคลุมทั้งจีโนมไม่ได้ช่วยเพิ่มความแม่นยำของการทำนายค่าจีโนมลักษณะผลผลิตนี้ แต่การคัดเลือกแบบสุ่มจำนวนหนึ่งก็ให้ความแม่นยำการทำนายค่าจีโนมได้ไม่ต่างจากการใช้เครื่องหมายจำนวนมากทั้งจีโนมสำหรับประชากรยางพาราชุดนี้ ดังนั้นแนวทางการศึกษาการคัดเลือกเครื่องหมายเพื่อใช้ในการสร้างแบบจำลองทางสถิติสำหรับการทำนายค่าจีโนมที่ไม่ใช่การคัดเลือกแบบสุ่มและให้ค่าความแม่นยำสูงขึ้น จึงเป็นแนวทางที่ควรมีการศึกษาต่อไป

## 5. รายการอ้างอิง

พิชิต สพโชค, พิศมัย จันทูมา และพนัส แพชนะ, 2550, การกรีดยางและการใช้สารเคมีเร่งน้ำยาง, พิมพ์ครั้งที่ 2, โรงพิมพ์ชุมนุมสหกรณ์

การเกษตรแห่งประเทศไทย จำกัด, กรุงเทพฯ.

- Ballesta, P., Bush, D., Silva, F.F. and Mora, F., 2020, Genomic predictions using low-density SNP markers, pedigree and GWAS information: A case study with the non-model species *Eucalyptus cladocalyx*, *Plants* 9: 99.
- Campos, G.D.L., Naya, H., Gianola, D., Crossa, J., Legarra, A., Manfredi, Weigel, K. and Cotes, J.M., 2009, Predicting quantitative traits with regression models for dense molecular markers and pedigree, *Genetics* 182: 375-385.
- Cericola, F., Jahoor, A., Orabi, J., Andersen, J.R., Janss, L.L. and Jensen, J., 2017, Optimizing training population size and genotyping strategy for genomic prediction using association study results and pedigree information: A case of study in advanced wheat breeding lines, *PLoS One* 12(1): e0169606.
- Chanroj, V., Rattanawong, R., Phunichai, T., Tangphatsornrung, S. and Ukoskit, K., 2017, Genome-wide association mapping of latex yield and girth in amazonian accessions of *Hevea brasiliensis* grown in a suboptimal climate zone, *Genomics* 109: 475-484.
- Clark, S.A., Hickey, J.M. and van der Werf, J.H., 2011, Different models of genetic variation and their effect on genomic evaluation, *Genet. Sel. Evol.* 43: 18.
- Cros, D., Mbo-Nkoulou, L., Bell, J.M., Oum, J., Masson, A., Soumahoro, M., Tran, D.M., Achour, A., Guen, V.L. and Demange, A.C.,



- 2019, Within-family genomic selection in rubber tree (*Hevea Brasiliensis*) increases genetic gain for rubber production, *Ind. Crops Prod.* 138: 111464.
- Daetwyler, H.D., Pong-Wong, R., Villanueva, B. and Woolliams, J.A., 2010, The impact of genetic architecture on genome-wide evaluation methods, *Genetics* 185: 1021-1031.
- Duangjit, J., Causse, M. and Sauvage, C., 2016, Efficiency of genomic selection for tomato fruit quality, *Mol Breed.* 36(3).
- Gianola, D. and van Kaam, J.B.C.H.M., 2008, Reproducing kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits, *Genetics* 178: 2298-2303.
- Grattapaglia, D., 2017, Status and Perspectives of Genomic Selection in Forest Tree Breeding, pp. 199-249, In Varshney, R.K., Roorkiwal, M. and Sorrells, M.E. (Eds.), *Genomic Selection for Crop Improvement*, Springer International Publishing, New York.
- Hayes, B.J., Bowman, P.J., Chamberlain, A.J. and Goddard, M.E., 2009, Invited review – Genomic selection in dairy cattle: Progress and Challenges, *J. Dairy Sci.* 92: 433-443.
- Honarvar, M. and Rostami, M., 2013, Accuracy of genomic prediction using RR-BLUP and Bayesian LASSO, *Eur. J. Exp. Biol.* 3: 42-47.
- Howard, R., Carriquiry, A.L. and Beavis, W.D., 2014, Parametric and nonparametric statistical methods for genomic selection of traits with additive and epistatic genetic architectures, *G3-Genes Genomes Genet.* 4: 1027-1046.
- Kwong, Q.B., Ong, A.L., Teh, C.K., Chew, F.T., Tammi, M., Mayes, S., Kulaveerasingam, H., Yeoh, S.H., Harikrishna, J.A. and Appleton, D.R., 2017, Genomic selection in commercial perennial crops: Applicability and improvement in oil palm (*Elaeis guineensis* Jacq.), *Sci. Rep.* 7: 2872/ 1-2872/9.
- Macciotta, N.P., Gaspa, G., Carnier, P. and Dimauro, C., 2009, Pre-selection of most significant SNPS for the estimation of genomic breeding values, *BMC Proceed.* 3(Suppl. 1): S14.
- Meuwissen, T.H.E., Hayes, B.J. and Goddard, M.E., 2001, Prediction of total genetic value using genome-wide dense marker maps, *Genetics* 157: 1819-1829.
- Park, T. and Casella, G., 2008, The Bayesian Lasso, *ASA* 103: 681-686.
- Resende, J.M., Mun, O.P., Resende, M., Garrick, D., Fernando, R., Dav' s, J., Jokela, E., Martin, T., Peter, G. and Kirst, M., 2012, Accuracy of genomic selection methods in a standard data set of Loblolly Pine (*Pinus taeda* L.), *Genetics* 190: 1503-1510.
- Solberg, T.R., Sonesson, A.K., Woolliams, J.A. and Meuwissen T.H.E., 2008, Genomic selection using different marker types and densities, *J. Animal Sci.* 86: 2447-2454.
- Zhang, Z., Ding, X., Liu, J. and Zhang, Q., 2011, Accuracy of genomic prediction using low-density marker panels, *J. Dairy Sci.* 94: 3642-3650.