國立臺灣大學生物資源暨農學院農藝學系

碩士論文

Department of Agronomy
College of Bioresources and Agriculture
National Taiwan University

Master Thesis

模擬探討基因體選種在水稻育種計畫中之運用

Genomic selection in rice breeding programs — a simulation study

潘芃諭

Peng-Yu Pan

指導教授：黃永芬 博士

Advisor: Yung-Fen Huang, Ph. D.

中華民國 106 年 7 月

July 2017

doi:10.6342/NTU201701806

國立臺灣大學碩士學位論文
口試委員會審定書

模擬探討基因體選種在水稻育種計畫中之運用

Genomic selection in rice breeding programs — a simulation study

本論文係　潘芃諭　君（R04621109）在國立臺灣大學農藝學系完成之碩士學位論文，於民國 106 年 07 月 13 日承下列考試委員審查通過及口試及格，特此證明

口試委員：

國立臺灣大學農藝學系　副教授
胡凱康　博士（召集委員）　胡凱康

行政院農委會農業試驗所　研究員
賴明信　博士　賴明信

國立臺灣大學農藝學系　教授
廖振鐸　博士　廖振鐸

國立臺灣大學農藝學系　副教授
林順福　博士　林順福

國立臺灣大學農藝學系　助理教授
黃永芬　博士（指導教授）　黃永芬

doi:10.6342/NTU201701806

致謝

記得碩一剛進來時，黃永芬老師讓我們閱讀《牧羊少年奇幻之旅》這本書，現在回想這兩年的碩士生活可謂應了書中名言佳句：「當你真心渴望某樣東西時，整個宇宙都會聯合起來幫助你完成。」，一路上感謝許多貴人的幫助，我才能順利完成碩士學業。首先最感謝的當然是我的指導教授黃永芬老師，謝謝老師在我研究過程中有疑問時給予許多建議導正我的方向，整份論文也經過老師用心的修改。另外老師常與我分享人生的體悟，令我思考並開拓視野和心胸。老師總是以溫和與樂觀的態度和我們對話，且時常鼓勵我，真的很謝謝老師不管是研究還是生活方面都十分用心幫助我，讓這兩年學習生活充實又順利。

另外要感謝論文口試委員們對論文的建議，林順福老師對育種方法的設計提出許多精闢的見解，並詳閱論文指出細節不足之處；賴明信博士讓我知道實際育種家們的訴求；廖振鐸老師幫助我解決線性混合效應模型及變方分析的問題；胡凱康老師指出模擬外表型值與 PCA 之流程誤區，以及老師碩一必修課的教導影響我研究的思考邏輯。除口試委員們外，還得謝謝李長沛博士與我分享育種家經驗，讓我在研究初始有思考的方向；謝謝蔡政安老師和蔡欣甫老師同樣幫助我解決許多統計上的難題，包括 R 程式碼錯誤訊息解決辦法；謝謝李欣叡學姐的碩士論文使我認識基因體選種和三個統計方法，以及十折交叉驗證 R 程式碼的思路。

一路上還得感謝許多人的陪伴：同實驗室的翁子涵學長和林譽嵐時常協助我完成一些事，還有仲汶經常聽我訴說生活中的瑣事並給予客觀的意見甚至幫我校稿；昇峰和昱心兩位大學起的僑生朋友，一起吃飯聊天分享彼此的想法；以及碩士認識成為朋友的亭妤學姐和瀞予，幸好有你們一同聊天抒發壓力並彼此鼓勵、共同以畢業為目標努力向前邁進。

特別感謝我的中醫生鄭醫師，拯救我身體的健康，使我能夠撐過這段時間。最後感謝家人對我無條件的支持，媽媽每天為我帶便當、爸爸為我生活上的問題出謀劃策、哥哥在我累時幫我按摩肩頸，使我只需專注在學業上不用煩惱其他事。畢業後將進入職場，期許自己在人生路上能不斷成長，並成為他人的貴人。

doi:10.6342/NTU201701806

摘要

基因體選種 (genomic selection, GS) 是一種新興的分子標誌輔助選種方法 (marker-assisted selection, MAS)，利用覆蓋整個基因體的大量分子標誌計算個體育種價估計值 (genomic estimated breeding value, GEBV)，並依此選拔出優良個體。由於高通量基因型鑑定技術的演進及成本下滑，研究者期盼藉由高通量基因型資料取代部份的外表型調查工作，使得單位時間能產生更佳的選拔效率。本研究欲藉由模擬探討 GS 在水稻育種計畫中之遺傳增進效果，分別探討雙親本雜交 (bi-parental cross) 及輪迴選種 (recurrent selection) 兩種育種流程。在此利用 327 個秈稻 (*Oryza sativa* L. ssp. *indica*) 優良品系的產量、抽穗期及株高之性狀資料，及 5,264 個單核苷酸多型性分子標誌 (single nucleotide polymorphism, SNP) 的基因型資料，以十折交叉驗證 (10-fold cross-validation) 比較 8 個統計方法的預測準確度，選出各性狀最適合之統計方法。比較結果顯示產量和株高最適用之 GS 模型為 RR-BLUP，抽穗期則以 BayesB 最佳，後續以這兩種方法預測個體之 GEBV。從 327 品系中選出 5 個產量表現優異之品系作為雙親本雜交親本，在傳統外表型選種 (phenotypic selection, PS) 架構下，於 F_2 及 F_6 世代另外以 GS 選拔產量性狀，比較 10 個雜交組合之 GS 與 PS 結果。結果顯示九個雜交組合之 GS 平均產量優於 PS，六個雜交組合中 GS 選拔之品系較 PS 早抽穗，八個雜交組合顯示 GS 選拔之品系株高較 PS 矮。輪迴選種以 327 品系中產量與抽穗期各前四名共八個品系作為親本，結果顯示，隨著循環數 (cycle) 增加，族群的產量與抽穗期皆有顯著改進，且對偶基因頻度於 Cycle 2 即大量固定，GS 能有效改良族群並快速累積有利對偶基因頻度。本研究以模擬方式探討 GS 在水稻育種計畫中之應用，未來或可先以輪迴選種改良族群，再從中選出優良親本雜交，於 F_2 早世代及 F_6 以 GS 方式節省大量人力且較精確選拔出優良品系。

關鍵字：基因體選種、十折交叉驗證、雙親本雜交、輪迴選種、水稻 (*Oryza sativa*)

doi:10.6342/NTU201701806

Abstract

Genomic selection (GS) is a new marker-assisted selection (MAS) method. GS uses a large number of molecular markers covering the entire genome to estimate the genomic estimated breeding value (GEBV) based on which the selection is made. Thanks to the evolution of high-throughput genotyping techniques and the consequent decline in costs, researchers expect to replace part of the phenotype survey by high-throughput genotypic data, resulting in a better selection efficiency per unit time. In this study, we wanted to explore the genetic improvement of GS in rice breeding program, and to explore the breeding methods of bi-parental cross and recurrent selection. The data set comprised of yield, flowering time (FL) and plant height (PH) from 327 *indica* rice (*Oryza sativa* L. ssp. *indica*) and 5,264 single nucleotide polymorphisms (SNP). Ten-fold cross-validation was used to assay the prediction accuracy of eight statistical models. The results showed that RR-BLUP was the most suitable model for yield and PH while BayesB for FL. In the traditional phenotypic selection (PS) framework, five lines with excellent yield were selected as parents from the initial 327 varieties. GS was applied in F_2 and F_6 in order to compare GS and PS in 10 bi-parental crosses. In recurrent selection scheme, the top four lines in yield and the earliest four liens in FL from the initial 327 varieties, were used as a parent. The results showed that the yield and FL of the population were significantly improved with the increase of cycle number, and the allele frequency is almost fixed in Cycle 2. GS can effectively improve the population and quickly accumulate favorable alleles. Based on the present study, a possible breeding scheme using GBS could be population improvement by recurrent selection then purify potential lines or select parents for bi-parental cross. Using GS in F_2 early generation and F_6 can be a way to save much manpower and precisely select candidate lines.

Keywords: genomic selection, 10-fold cross-validation, bi-parental cross, recurrent selection, rice (*Oryza sativa*)

doi:10.6342/NTU201701806

目錄

doi:10.6342/NTU201701806

doi:10.6342/NTU201701806

表目錄

doi:10.6342/NTU201701806

圖目錄

doi:10.6342/NTU201701806

中英對照表

中文	英文	縮寫
十折交叉驗證	10-fold cross-validation	
貝氏 A	BayesA	
貝氏 B	BayesB	
貝氏 C	BayesC	
貝氏最小絕對壓縮挑選機制法	Bayesian LASSO	BL
貝氏脊迴歸	Bayesian ridge regression	BRR
雙親本雜交	bi-parental cross	
育種族群	breeding population	
全基因體關聯分析	genome-wide association analysis	GWAS
個體育種價估計值	genomic estimated breeding value	GEBV
基因體選種	genomic selection	GS
連鎖失衡	linkage disequilibrium	LD
主效基因座	major quantitative trait loci	major QTL
分子標誌輔助輪迴選種	marker-assisted recurrent selection	MARS
分子標誌輔助選種	marker-assisted selection	MAS
微效基因座	minor quantitative trait loci	minor QTL
外表型選種	phenotypic selection	PS
預測準確度	prediction accuracy	
主成分分析	principal components analysis	PCA
隨機森林法	random forests	RF
輪迴選種	recurrent selection	
再生核希氏函數空間法	reproducing kernel Hilbert spaces regression	RKHS
脊迴歸最佳線性無偏預測法	ridge regression best linear unbiased prediction	RR-BLUP
選拔強度	selection intensity	

doi:10.6342/NTU201701806

中文	英文	縮寫
單核苷酸多型性分子標幟	single nucleotide polymorphism	SNP
訓練族群	training population	
驗證族群	validation population	

doi:10.6342/NTU201701806

一、前言

1.1 基因體選種簡介

基因體選種 (genomic selection, GS) 是一種新興的分子標誌輔助選種方法 (marker-assisted selection, MAS)，利用全基因體的基因型資料預測任何性狀表現以進行選拔 (Meuwissen et al. 2001)。GS 策略中具有兩種族群：訓練族群 (training population) 和育種族群 (breeding population)，前者用以調查外表型和基因型資料，利用適合之統計方法估計各分子標誌效應，以建立預測模型；後者通常由訓練族群後代或相關之新品種組成 (Desta and Ortiz 2014)，只調查基因型資料並將其代入預測模型，預測個體的育種價估計值 (genomic estimated breeding value, GEBV)，依照 GEBV 選拔個體 (Lorenz et al. 2011)。通常會先使用交叉驗證法 (cross-validation) 將已知外表型之族群作為驗證族群 (validation population) 代入已訓練出參數之模型預測 GEBV，再以 GEBV 與外表型間的相關係數作為預測準確度 (prediction accuracy)，以此決定適用之模型。Taylor et al. (2016) 提到 GS 概念最初由 Meuwissen et al. (2001) 提出，作者以模擬方式評估最小平方法 (least squares)、最佳線性無偏預測模型 (best linear unbiased prediction, BLUP) 和貝氏方法模型估計分子標誌效應及預測 GEBV 的表現；但當時基因型鑑定技術成本高，要獲得全基因體的基因型資料費用較昂貴。隨著近年來次世代定序 (next-generation sequencing, NGS) 演進，高通量分子標誌成本下降，GS 逐漸受到重視，可能徹底改變作物及林木之育種計畫實施方式 (Desta and Ortiz 2014)。

GS 於動物育種發展較作物育種早，因傳統動物育種耗時長、成本高，如乳牛之種牛育種，經後裔檢定選拔優良公牛，總耗時長達 64 個月，藉由 GS 則可於公牛出生時以預測之 GEBV 選拔省下後裔檢定時間與繁殖成本 (Schaeffer 2006)。Schaeffer (2006) 提出以 GS 執行乳牛育種之計畫流程，假設 GS 準確度為 0.75，可提高 2 倍遺傳增進率並省下 92%後裔檢定成本 (Heffner et al. 2009)。

doi:10.6342/NTU201701806

2008 年一月第一個農業物種高通量基因晶片 Illumina BovineSNP50 公開上市後，加速 GS 應用於乳牛育種計畫，Meuwissen et al. (2001) 的研究更受到重視，引用次數大幅增加，美國 Holstein 乳牛基因型鑑定個數與牛奶產量遺傳增進率明顯提高 (Taylor et al. 2016)。

GS 於作物育種之研究發展較晚，Bernardo and Yu 於 2007 年所發表之模擬研究屬於早期研究，結果顯示在不同性狀遺傳結構情境和分子標誌數量下，GS 於玉米雙親本雜交族群輪迴選種之選拔效率優於傳統 MAS。GS 同傳統 MAS 一樣能加速世代推進並提高選拔強度 (selection intensity)，其相較外表型選種 (phenotypic selection, PS) 之優勢為單位時間遺傳增進率較高 (Bernardo and Yu 2007)。因作物無需田間評估，但透過基因型表現即可進行選拔，使得育種族群可大幅增加不同種原、家系的試驗數量 (Lorenz et al. 2011)，並將溫帶玉米輪迴選種之每循環所需之 2 年縮短至 4 個月 (Massman et al. 2013)。

傳統 MAS 只使用與外表型顯著相關的分子標誌，因此適合選拔主效基因座 (major quantitative trait loci, major QTL) 控制之性狀，但不適用於受多個微效基因座 (minor quantitative trait loci, minor QTL) 控制且易受環境影響之數量性狀 (quantitative trait)(Bernardo 2008; Bernardo and Yu 2007; Heffner et al. 2011; Lorenz et al. 2011; Massman et al. 2013)。Moreau et al. (2004) 研究顯示 MAS 快速固定與玉米產量 QTL 連鎖之分子標誌，但其選拔增進不如外表型選拔，推測與使用之分子標誌數不足以含括所有調控產量之 QTL 有關。GS 使用覆蓋全基因體的高通量分子標誌，利用新的統計方法分析數量性狀，同時估計所有分子標誌效應，可能保留效應較小於傳統 MAS 中不易顯著之微效基因座效應，因此較傳統 MAS 更適合選拔複雜性狀 (complex trait)(Lorenz et al. 2011; Massman et al. 2013; Meuwissen et al. 2001)。

doi:10.6342/NTU201701806

1.2 統計方法

GS 使用覆蓋整個基因體的大量分子標誌，因此會有估計的預測變數效應 p (即分子標誌效應) 大於外表型觀測值 n 的問題 ($p>>n$)。就像在解多元一次方程式時，若未知數多於聯立方程式，就無法得到唯一解，因此過去 MAS 估計分子標誌效應所使用的最小平方法在 GS 下會因自由度不足而無法估計全部的預測變數效應。即使自由度足夠，也不一定能得到唯一解，若方程式間線性不獨立還是可能有無限多組解甚至無解，且分子標誌間可能存在高度的多元共線性 (multicollinearity)，產生過度配適 (overfitting) 的模型，高估資料中微小的變動導致預測能力差 (Lorenz et al. 2011)。研究學者們提出各種 GS 統計模型方法來解決這些問題，這些模型可大致分為壓縮模型 (shrinkage model)、變數選擇模型 (variable selection model)、核方法 (kernel method) 和降維方法 (dimension reduction method)，並且將分子標誌效應視為隨機型效應而非固定型效應 (Lorenz et al. 2011)。本研究中共使用八種統計方法，分別為脊迴歸最佳線性無偏預測法 (ridge regression best linear unbiased prediction, RR-BLUP)、貝氏脊迴歸 (Bayesian ridge regression, BRR)、貝氏最小絕對壓縮挑選機制法 (Bayesian LASSO, BL)、BayesA、BayesB、BayesC、再生核希氏函數空間法 (reproducing kernel Hilbert spaces regression, RKHS) 和隨機森林法 (random forests, RF)。

RR-BLUP 是最早被提出的基因體選種方法之一，屬於壓縮模型。假設全部分子標誌效應值小但非 0，且具有相同的變方 V_G/N_M (V_G 為遺傳變方；N_M 為分子標誌數量)，所以除殘差變方外只需要估計一個遺傳變方即解決自由度不足的問題，並一起以脊參數 (ridge parameter) λ 同等地向 0 壓縮，不過效應值大小不一定相同 (Bernardo and Yu 2007; Desta and Ortiz 2014; Lorenz et al. 2011; Meuwissen et al. 2001)。此方法沒有變數選擇而包含全部分子標誌在模型中，故適用於多個微效基因座控制的性狀 (Lorenz et al. 2011)，且因運算快及在不同族群大小和分

doi:10.6342/NTU201701806

子標誌數量下的預測準確度穩定，是 GS 最常使用的統計方法 (Lorenzana and Bernardo 2009)。

BRR 也為壓縮模型且為貝氏方法，將全部分子標誌效應一起向 0 壓縮並假設分子標誌效應的事前分布 (prior distribution) 為高斯分布 (Gaussian distribution)(Perez and de los Campos 2014)，即常態分布，因此與 RR-BLUP 一樣每個效應的壓縮程度相同 (Desta and Ortiz 2014; Perez and de los Campos 2014)。

BL 為 Park and Casella (2008) 提出的貝氏版 LASSO，因此同時是壓縮模型及變數選擇模型。其假設分子標誌效應變方的事前分布為指數分布，使得效應的邊際分布為雙指數 (double exponential, DE) 分布，相較於高斯分布在 0 的位置有更高的質量密度和較厚的兩尾 (Perez and de los Campos 2014)，因此壓縮能力比 RR-BLUP 強，大幅度壓縮小的效應估計值，使某些係數為 0 簡化模型；而較大的效應估計值之壓縮程度小，因此適合由少數主效基因座 (major QTL) 控制的性狀 (Desta and Ortiz 2014; Lorenz et al. 2011; Perez and de los Campos 2014; Perez et al. 2010)。BL 相對非貝氏 LASSO，會基於迴歸係數設定 λ 的事前分布，因此非零迴歸係數解能多於 n-1 個 (n 為訓練族群大小)，或許能增加模型的準確度 (Lorenz et al. 2011)。

BayesA、BayesB、BayesC 皆為貝氏方法的壓縮模型，壓縮較小的分子標誌效應而保留較大的效應，且 BayesB 和 BayesC 也為變數選擇模型。BayesA 令分子標誌變方的事前分布為反卡方分布 (inverse chi-square, χ^{-2}) 使得效應為尺度化 t 分布 (scaled t-distribution)，相較 BL 的 DE 分布在 0 的位置有更高的質量密度 (Desta and Ortiz 2014; Perez and de los Campos 2014)。BayesB 修改自 BayesA，分子標誌效應事前分布為尺度化 t 分布外，另外設定分子標誌完全沒有效應的機率 π 之混合分布，達到變數選擇效果，適合由少數主效基因座控制的性狀；$\pi = 0$ 即為 BayesA (Desta and Ortiz 2014; Lorenz et al. 2011; Perez and de los Campos 2014)。BayesC 的分子標誌效應事前分布與 BayesB 一樣為混合分布，並假設非 0 分子標誌效應抽樣自相同變方的分布，因此由尺度化 t 分布轉變為高斯分布，且為變數

doi:10.6342/NTU201701806

選擇模型 (Perez and de los Campos 2014)，能彈性地應用於由少數主效基因座或多數微效基因座控制的性狀 (Lorenz et al. 2011)。

RKHS 為半母數 (semi-parametric) 的非線性迴歸模型，適合預測由大量非累加性效應組成的性狀 (Lorenz et al. 2011)。RKHS 屬於核方法，利用核函數將分子標誌資料從多維的輸入空間轉換成一維實數特徵空間之觀測項間距離，產生類似譜系關係的個體遺傳距離矩陣，其為半正定矩陣，可代入線性模型中，估計特徵空間中的個體遺傳效應 (Lorenz et al. 2011; Perez and de los Campos 2014)。

RF 是一種整體學習 (ensemble learning) 方法，為無母數迴歸 (non-parametric regression)，由 Breiman (2001) 提出，建造大量相互間獨立的決策樹來做迴歸或類別分類。利用自助重抽法 (bootstrap) 重新取樣 (重複抽樣) 建立新的的訓練資料，再於建造決策樹時加入隨機性 (randomness)，每個節點 (node) 從隨機選擇的預測變數子集合中經過演算法得到決定分支的規則，以此降低決策樹間的變異性 (variance)、偏差 (bias) 與過度配適 的問題，提高穩健性 (robustness)(Breiman 2001; Liaw and Wiener 2002; Onogi et al. 2015)。預測新資料則是總合 (aggregate) 全部決策樹的預測值，迴歸分析取平均值；類別分析多數決 (majority vote) 得到最後的預測結果 (Liaw and Wiener 2002)。RF 因在同個決策樹下連續以多變量分類樣本，故被認為可以捕捉到分子標誌間的交感效應 (Desta and Ortiz 2014; Onogi et al. 2015)。

1.3 目前 GS 研究趨勢

GS 預測能力除受性狀遺傳結構與其適合之統計模型影響外，還受到性狀遺傳率、訓練族群大小、訓練族群與驗證族群相關性、分子標誌數量等因子影響 (Desta and Ortiz 2014; Lorenz et al. 2011; Zhao et al. 2015)。一般訓練族群越大、分子標誌數量越多、性狀遺傳率越高則預測準確度越高 (Bernardo and Yu 2007; Lorenzana and Bernardo 2009)；訓練族群與驗證族群相關性越高則預測準確度越高且比訓練族群大小重要 (Lorenz and Smith 2015)。大部分 GS 實證研究僅以交叉

doi:10.6342/NTU201701806

驗證 (cross-validation) 評估不同目標性狀在各種訓練族群大小、驗證族群、分子標誌數量、統計模型等變因下之預測準確度，涵蓋玉米、小麥、大麥、燕麥、黑小麥、馬鈴薯、大豆、番茄、水稻等作物 (Asoro et al. 2011; Crossa et al. 2014; Duangjit et al. 2016; Habyarimana et al. 2017; Heffner et al. 2011; Lorenz et al. 2012; Ma et al. 2016; Spindel et al. 2015; Wurschum et al. 2017)。少數研究包含世代推進之遺傳重組，早期僅有模擬研究模擬世代推進比較 GS 選拔效率，涵蓋玉米、小麥、油棕、大麥等作物 (Bernardo and Yu 2007; Heffner et al. 2010; Wong and Bernardo 2008; Zhong et al. 2009)，結果顯示 GS 選拔效率優於 MAS；另一篇模擬比較 GS 與 PS 在多年生黑麥草 (ryegrass) 育種計畫中持久性與產量性狀選拔效率之研究，顯示 GS 遺傳增進分別為 PS 之 2 或 3 倍，但近親交配率較高 (Lin et al. 2016)。

Massman et al. (2013) 為第一篇真正於作物育種流程中實施 GS 之實驗，顯示雙親本溫帶玉米 B73 × Mo17 之 intermated recombinant inbred lines (iRILs) 經 GS 輪迴選種後的遺傳增進。研究以 223 個 RILs 經試交試驗調查外表型值和 287 個 SNP 資料建立預測模型，改良目標為莖稈指數 (Stover Index) 及產量和莖稈綜合指數 (Yield + Stover Index)，比較 GS 和分子標誌輔助輪迴選種 (marker-assisted recurrent selection, MARS) 之遺傳增進，經多地點試驗調查外表型證實 GS 之兩種指數性狀遺傳增進比 MARS 多 15-50%。Combs and Bernardo (2013) 實施 4 次 GS 回交將半矮性 (semi dwarf) 玉米種原導入美國玉米帶自交系中，同時也以外表型選拔至 BC_4。結果顯示 GS 從 Cycle 1 至 Cycle 5 之遺傳增進維持 PS 從 Cycle 0 至 Cycle 1 之遺傳增進或有改善，GS 遺傳增進觀測值大致與預測值一致，但最終平均表現不一定最好。Beyene et al. (2015) 以 8 個熱帶玉米雙親本雜交為材料，比較 GS 與傳統基於譜系法之 PS 在乾旱逆境下之穀粒產量選拔效率，結果顯示 GS 於每次選拔循環平均增加 0.086 Mg/ha，Cycle 3 之雜交系產量顯著高於 Cycle 0 之雜交系且比傳統譜系法多 7.3%，顯然 GS 產量選拔效率優於 PS。Rutkoski et al. (2015) 比較單位時間內 GS 與 PS 於小麥稈銹病 (stem rust) 抗性之選拔效率，結

doi:10.6342/NTU201701806

果顯示 GS 遺傳增進等同於 PS，且遺傳變異下降較快。Yamamoto et al. (2017) 以 96 個 F_1 番茄品系和 337 個 SNP 建立可溶性固體含量 (soluble solids content) 及總果重之預測模型，欲預測產生優良後代之親本組合及後代之性狀表現。作者成功以 GS 模型預測優良親本，但預測效率 (GEBV 與外表型值相關係數) 依據雜交組合和選拔性狀而不同。Sallam and Smith (2016) 以實際大麥歷史資料比較產量和抗病性單一世代 GS 相對 PS 之選拔效率，結論是 GS 遺傳增進與 PS 相同且有利對偶基因頻度增加，但 GS 育種循環時間更短及成本更低。

1.4 研究目的

GS 研究作物以玉米和小麥為主，水稻相關研究較少，且皆為交叉驗證比較各種變因之預測準確度，缺乏世代推進之研究，因此本研究欲藉由實際秈稻資料模擬世代推進探討 GS 在水稻育種計畫中之遺傳增進效果。實驗目的有三：(1) 以十折交叉驗證 (10-fold cross-validation) 比較 8 個統計方法的預測準確度，分別選出產量、抽穗期及株高最適合之統計方法。(2) 模擬雙親本雜交 (bi-parental cross) 流程，比較 GS 與 PS 於產量性狀之選拔效率。(3) 模擬輪迴選種 (recurrent selection) 育種流程，目標性狀為產量和抽穗期，探討 GS 之各循環遺傳增進及對偶基因頻度變化。

doi:10.6342/NTU201701806

二、材料與方法

2.1 植物材料之外表型資料及基因型資料

本研究使用 Spindel et al. (2015) 所發表的秈稻 (*Oryza sativa* L. ssp. *indica*) 資料，資料簡述如下：由 IRRI 灌溉水稻育種計畫選出的 369 個優良育種品系，在 2011 和 2012 年的乾濕兩季以隨機完全區集設計 (randomized complete block design, RCBD) 進行區域產量試驗 (regional yield trial, RYT)，共調查七個性狀；本研究使用其中的抽穗期 (flowering time, FL; days)、株高 (plant height, PH; cm) 和穀粒產量 (yield; kg/ha) 共三個性狀資料。

基因型資料以 Genotyping-by-Sequencing (GBS) 方法製備產生，原始資料包含 369 品系和 108,024 個單核苷酸多型性分子標誌 (SNP)，其中每個 SNP 的 call rate≧0.75，個體缺值數＜0.6。Spindel et al. (2015) 補值 (imputation) 後將缺值數＞10%的 SNP 刪除餘下 73,147 個 SNP，並移除個體缺值數≧0.6 的品系至 363 個，再以主成分分析 (principal components analysis, PCA) 排除與 *japonica* 有親緣關係的品系至 332 個品系；此份資料與原始資料以 csv 檔公開於 Rice Diversity 網址 https://ricediversity.org/data/index.cfm (圖一虛線上方)。本研究於檢查資料時發現篩選後 73,147 個 SNP 資料中存在的 A1303 品系在原始 108,024 個 SNP 資料中不存在而刪除，再將作者以原始資料計算的 minor allele frequency (MAF)＜0.05 之分子標誌移除，餘下 50,127 個 SNP。

外表型資料部份，根據基因型資料篩選出之 331 個品系其中 4 個 (M1472、M1473、M1474、M1475) 在 2011 和 2012 年之兩季皆為缺值，刪除後餘下 327 個品系。另，2011 年濕季的 M1396 品系因名稱重複但數值不同所以予以移除；2012 年濕季的 B1203 品系資料重複且數值相同，所以只移除重複部份。本研究最終所使用之 327 個品系在各年份、季節的品系數、區集數和各性狀缺值數如表一所示。

doi:10.6342/NTU201701806

重新計算 327 個品系 50,127 個 SNP 的 call rate 和 MAF，經過篩選後得到 38,639 個 SNP (圖一虛線下方)。

2.2 變方分析與外表型資料校正

本研究皆利用 R 統計軟體 version 3.3.2 (R Core Team, 2016) 整理資料、建立預測模型、模擬水稻外表型與世代推進，以及進行統計分析。因資料有缺值且 2012 年乾季區集數為 2，與其他三季之區集數不同，屬非平衡資料 (unbalance data)，故利用 R 套件 *car* version 2.1-3 (Fox and Weisberg 2011) 函數 *Anova* 計算 Type III Sum of Square。其中 singular.ok 參數在此資料結構下需設定為 TRUE 否則無法運算。將年份和季節組合成四種環境效應 (E_j)，變方分析模型為

$$y_{ijkl} = \mu + \tau_i + E_j + B_{k(j)} + \tau_i \times E_j + \varepsilon_{ijkl} \quad (1)$$

y_{ijkl}為外表型資料，τ_i為品系效應又即基因型效應，E_j為環境效應，$B_{k(j)}$為區集效應巢式於環境效應下，$\tau_i \times E_j$為基因型效應與環境效應之交感效應，皆為固定型效應。

為結合四種環境外表型資料來建立預測模型，以線性混合效應模型 (linear mixed-effects model) 如式(2) 校正外表型資料

$$y_{ijk} = \mu + \tau_i + E_j + B_{k(j)} + \varepsilon_{ijk} \quad (2)$$

並將環境效應與區集效應視為隨機型效應，且本研究只關注單一環境下各品系間相對表現故不考慮交感效應。估算第 i 個品系的基因型效應 (τ_i)，為固定型效應，作為後續全基因體關聯分析 (genome-wide association study, GWAS) 和 GS 模型的應變數 (y)。μ為外表型族群平均值；E_j為E_1至E_4分別表示 2011 年乾季、2011 年濕季、2012 年乾季與 2012 年濕季之效應值，抽樣自$N(0, \sigma_E^2)$；$B_{k(j)}$為第 j 環境下之第 k 區集效應，抽樣自$N(0, \sigma_B^2)$；ε_{ijk}為其他環境誤差，抽樣自$N(0, \sigma_\varepsilon^2)$。利用 R 套件 *lme4* version 1.1-12 (Bates et al. 2015) 之 *lmer* 函數以無截距項的模型估計出各品系的基因型效應，可直接代為 GWAS 及 GS 模型之應變數；

doi:10.6342/NTU201701806

而估計出的σ_E^2及σ_ε^2則在後續育種模擬世代推進中產生外表型之用。廣義遺傳率 (broad-sense heritability, h^2) 計算公式為

$$h^2 = \sigma_G^2/\sigma_P^2 \quad (3)$$

σ_G^2為遺傳變方，代入 327 品系校正外表型間之變方。σ_P^2為外表型變方，代入 327 品系在 2011D、2011W、2012D、2012W 四個環境下外表型區集平均值間之變方。

2.3 族群結構分析與全基因體關聯分析

為了解 327 品系之族群結構，以 *prcomp* 函數針對 38,639 SNP 降維，進行 PCA。另以套件 *rrBLUP* version 4.4 (Endelman 2011) 中 *GWAS* 函數進行全基因體關聯分析 (genome-wide association analysis, GWAS)，其混合模型為

$$y = X\beta + Zg + S\tau + \varepsilon \quad (4)$$

校正後的外表型資料為 y；β 可為環境因子或族群結構之固定型效應，因無明顯族群結構故無此項；g 代表每個品系的遺傳背景，視為隨機型效應，$\sigma_g^2 = K\sigma^2$，K 為品系間的共變異親緣關係矩陣 (kinship matrix)，預設以 *A.mat* 函數產生之實際累加性關係矩陣代入；τ 為固定型累加性 SNP 效應。參數設定 min.MAF 因已自行篩選故此設定 0，其餘參數 n.PC = 0、n.cor = 1、P3D = TRUE、plot=TRUE 等皆為預設，偽發現率 (false discovery rate, FDR) 為 0.05。

2.4 GS 模型之分子標誌

因 Spindel et al. (2015) 研究結果顯示此份水稻材料約每 50 kb (0.2 cM) 隨機抽取一個 SNP 共約 6-7,000 SNPs 即能使基因體選種有良好的表現，故從 38,639 個 SNP 每 50 kb 區間隨機抽取一個 SNP，並包含抽穗期性狀 GWAS 結果中第 3 和第 6 條染色體上最顯著的 SNP，共 5,264 個 SNP 使用於 GS 模型中並作為模擬族群基因型資料。其中第 3 和第 6 條染色體上最顯著的 SNP 與同條染色體上其餘顯著 SNP 之連鎖失衡 (linkage disequilibrium, LD) 皆小於 0.1，故以最顯著 SNP

doi:10.6342/NTU201701806

代表；LD 值以 *LDcorSV* version 1.3.1 (David Desrousseaux et al. 2013) *Measure.R2* 函數計算 r^2。為便於模擬育種流程世代推進之遺傳重組，將補值資料全部替換為主效對偶基因 (major allele) 1，經測試其 GS 模型之 RR-BLUP 與 BL 產量性狀預測準確度與補植結果差異不大。因染色體重組時需要遺傳距離 (genetic distance) 估計發生互換的機率，利用 cM Converter version 1.2.1 軟體 (Lorieux, http://mapdisto.free.fr/cMconverter/)，將分子標誌之物理距離 (physical distance) 轉換為遺傳距離；距離轉換基準為 IR64 × Azucena MSU7 CIAT 之遺傳圖譜。

2.5 建立預測模型

個體育種價估計值使用以下八種統計方法計算：RR-BLUP、BRR、BL、Bayes A、Bayes B、Bayes C、RKHS、RF 等，各方法均有適合之 R 套件：其中 RR-BLUP 利用 *rrBLUP* version 4.4，貝氏方法與 RKHS 利用 *BGLR* version 1.0.5 (Perez and de los Campos 2014)，RF 則利用 *randomForest* version 4.6-12 (Liaw and Wiener 2002)。

2.5.1 RR-BLUP

GS 預測模型的標準線性模型如式(5)：

$$y_i = \mu + \sum x_{ij}\beta_j + \varepsilon_i \tag{5}$$

y_i為品系 i 的外表型值，μ為外表型平均值，x_{ij}為第 i 品系第 j 個分子標誌的基因型值。β_j為第 j 個分子標誌的效應，設為從$N(0, \sigma_\beta^2)$中抽出的隨機型效應，其中分子標誌效應變方σ_β^2在 RR-BLUP 的假設下等於V_G/N_M (V_G為遺傳變方；N_M為分子標誌數量)。ε_i為殘差，取自$N(0, \sigma_\varepsilon^2)$。$x_{ij}$以 1、0、-1 構成基因型矩陣，代表雙倍體基因型 AA、AB、BB (Desta and Ortiz 2014)。利用 Endelman (2011) 發表之 *rrBLUP* 套件中的 *mixed.solve* 函數解混合模型估計β_j，其先以最大概度法 (maximum likelihood, ML) 或預設的受制最大概度法 (restricted maximum

doi:10.6342/NTU201701806

likelihood, REML) 估計兩個變方成分σ_β^2和σ_ε^2，再利用頻譜分解 (spectral decomposition) 算法修改β_j的常態最小平方估計式：

$$\hat{\beta} = (\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}'\mathbf{y} \quad (6)$$

X為基因型矩陣，**I**為單位矩陣，**y**是外表型向量，脊參數 $\lambda = \sigma_\varepsilon^2/\sigma_\beta^2$，由於分母較分子小，使得分子標誌效應大幅度壓縮向 0。與一般最小平方法相較多了$\lambda\mathbf{I}$這項，使得$\mathbf{X}'\mathbf{X}$非奇異 (nonsingular) 反矩陣存在，並減少預測變數間的共線性 (Lorenz et al. 2011)。

2.5.2 BRR

BRR 模型在式(5)的基礎下，另設分子標誌效應變方σ_β^2來自一個具有尺度參數的反卡方事前分布，效應事前分布為常態分布，分子標誌效應與超參數 (hyper-parameter) 的聯合分布 (join distribution) 如式(7)：

$$p(\beta, \sigma_\beta^2) = \left\{\prod\prod N(\beta_j|0, \sigma_\beta^2)\right\}\chi^{-2}(\sigma_\beta^2|df_\beta, S_\beta) \quad (7)$$

其中超參數df_β與S_β的值在 *BGLR* 套件中可自行設定或使用套件內設的規則給出預設值，本研究參數設定依照預設，自由度為 5，而尺度參數須符合變方分割參數 R2 表示分子標誌效應變方可解釋外表型變方的比例，預設為 0.5，即預測變數與殘差各解釋 50%應變數變異 (Perez and de los Campos 2014)。貝氏方法無法以統計方式取得參數估計值，故而發展出馬可夫鏈蒙地卡羅法 (Markov chain Monte Carlo, MCMC)，此套件藉由吉布斯抽樣法 (Gibbs sampling) 從事後分布中重複抽樣取得近似參數值的分布，再計算出該分布相應的摘要統計值 (summary statistics)，估計分子標誌效應 (Lorenz et al. 2011)。參考 Mohammadi et al. (2015) 發表的 *PopVar* 套件設定，在交叉驗證中抽樣次數 (iteration) 及生樣本數 (burn-in) 分別設為 1500 和 500 以提高計算效率；在基因體選種正式建模時則分別設為 12000 和 3000。此法計算成本高相當耗時，(Perez and de los Campos 2014) 發表的套件結合 R 內建函數與另外編寫的 C 和 Fortran 程序，以改善電腦運算效率。

doi:10.6342/NTU201701806

2.5.3 Bayesian LASSO

BL 模型的分子標誌效應之條件事前分布為常態分布如式(8)：

$$p(\beta|\tau^2,\sigma_\varepsilon^2)=\prod\prod N\left(\beta_j|0,\tau_j^2\sigma_\varepsilon^2\right) \quad (8)$$

σ_ε^2取自反卡方分布，τ_j^2為分子標誌變方參數，具分子標誌專一性，使模型針對不同大小效應估計值壓縮，其來自指數事前分布，以式(9)表示：

$$p(\tau^2|\lambda)=\prod\prod Exp\left(\tau_j^2|\frac{\lambda^2}{2}\right) \quad (9)$$

在 *BGLR* 套件中，正則化參數 (regularization parameter) λ 有三種事前分布可選擇，分別是固定值、$\lambda^2\sim Gamma(r,s)$或$\frac{\lambda}{\max}\sim Beta(p_0,\pi_0)$，本研究使用預設的 γ 分布 (gamma distribution)，s 為 1.1，與 BRR 同樣以內定規則求尺度參數解 (Perez and de los Campos 2014; Perez et al. 2010)。綜合以上假設，最後分子標誌效應的邊際分布為雙指數分布，根據效應估計值大小有不同程度的壓縮，效應越大壓縮幅度越小 (Perez et al. 2010)。同樣以吉布斯抽樣法估計分子標誌效應，設定同 BRR 所述。

2.5.4 BayesA、BayesB 與 BayesC

BayesA 為 Meueissen et al. (2001) 提出以貝氏分析放寬 RR-BLUP 遺傳效應均勻分布在基因體中的假設之方法，每個分子標誌效應 j 取自常態分布$N\left(0,\sigma_{\beta_j}^2\right)$，使每個分子標誌以不同程度壓縮至 0 (Lorenz et al. 2011)。其變方參數抽樣自反卡方分布，超參數之尺度參數來自 γ 分布，最後分子標誌效應邊際分布為尺度化 t 分布，聯合分布以式(10)表示 (Perez and de los Campos 2014)：

$$p\left(\beta,\sigma_\beta^2,S_\beta\right)=\left\{\prod N\left(\beta_j|0,\sigma_{\beta_j}^2\right)\chi^{-2}\left(\sigma_{\beta_j}^2|df_\beta,S_\beta\right)\right\}G\left(S_\beta|r,s\right) \quad (10)$$

BayesB 在 BayesA 的基礎下另外設定分子標誌完全沒有效應的機率π，其來自 β 分布 (beta distribution)，聯合分布如式(11)(Perez and de los Campos 2014)：

doi:10.6342/NTU201701806

$$p(\beta,\sigma_\beta^2,\pi)=\left\{\prod\left[\pi N(\beta_j|0,\sigma_\beta^2)+(1-\pi)1(\beta_j=0)\right]\chi^{-2}\left(\sigma_{\beta_j}^2|df_\beta,S_\beta\right)\right\}B(\pi|p_0,\pi_0)\times G(S_\beta|r,s) \quad (11)$$

BayesC 修改自 BayesB，假定全部存在之分子標誌效應事前變方相等，故效應邊際分布與 BRR 相同為常態分布，聯合分布如式(12)(Lorenz et al. 2011; Perez and de los Campos 2014)：

$$p(\beta,\sigma_\beta^2,\pi)=\left\{\prod\left[\pi N(\beta_j|0,\sigma_\beta^2)+(1-\pi)1(\beta_j=0)\right]\right\}\times\chi^{-2}(\sigma_\beta^2|df_\beta,S_\beta)B(\pi|p_0,\pi_0) \quad (12)$$

本研究使用 *BGLR* 套件計算，參數皆使用預設值，自由度df_β為 5，γ 分布的形狀參數 (shape parameter) *s* 為 1.1，在預設 R2 為 0.5 下以內建算法計算尺度參數 S_β，再求得速率參數 (rate parameter) *r*。β 分布之分子標誌具非 0 效應的事前機率 π_0預設為 0.5，事前計數 (prior counts)量p_0為 10 (Perez and de los Campos 2014)。同樣以吉布斯抽樣法估計分子標誌效應，設定同 BRR 所述。

2.5.5 RKHS

RKHS 的模型為非線性迴歸模型：

$$\begin{cases} y_i=\mu+u_i+\varepsilon_i \text{ with} \\ p(\mu,u,\varepsilon)\propto N(u|0,\boldsymbol{K}\sigma_u^2)N(\varepsilon|0,\boldsymbol{I}\sigma_\varepsilon^2) \end{cases} \quad (13)$$

核函數 $\boldsymbol{K}$ 將分子標誌資料轉換成個體間的平方歐基里德距離(squared-Euclidean distance)，式子為：

$$\boldsymbol{K}(x_i,x_{i'})=exp\left\{-h\times\frac{\sum_{j=1}^{p}(x_{ij}-x_{i'j})^2}{p}\right\} \quad (14)$$

其中x_i代表品系 *i* 基因型，x_{ij}為第 *i* 品系第 *j* 個分子標誌的基因型值，共有 *p* 個分子標誌。*h* 為帶寬參數 (bandwidth parameter) 或稱為平滑參數 (smoothing parameter)，影響輸入空間中距離與特徵空間中距離的關係，在此設為 0.5，表示

doi:10.6342/NTU201701806

兩者間相似；若為 2.5 則相關性小 (Lorenz et al. 2011; Perez and de los Campos 2014)。另外特徵空間中的個體效應變方事前分布為反卡方分布$\sigma_u^2 \sim \chi^{-2}(df, S)$，同樣以預設的自由度為 5 及分子標誌效應變方解釋外表型變方 50%之條件求得尺度參數 S，並以吉布斯抽樣法估計效應，設定同 BRR 所述 (Perez and de los Campos 2014)。

2.5.6 Random Forests

RF 是一種整體學習方法，並非單一的迴歸模型，而是利用自助重抽總合法 (bootstrap aggregating, bagging) 建立新的訓練資料，再隨機抽取一部份分子標誌資料來建造決策樹，本研究使用 Liaw 和 Wiener (2002) 發表的 *randomForest* 套件。參數設定依照預設，節點大小最小值為 5，即至少由 5 個變數組成；每個節點以包含m_{try}個變數的隨機預測變數子集合來運算確立分支的模型，預設$m_{try} = \frac{p}{3}$，p 為分子標誌數目；設定建造的決策樹數目n_{tree}為 1000 (Liaw and Wiener 2002)。

2.6 十折交叉驗證

八種統計方法最後藉由分子標誌資料預測該個體之育種價估計值 (genomic estimated breeding value, GEBV)，計算 GEBV 與校正後外表型資料間之皮爾森相關係數 (Pearson's correlation coefficient) 作為預測準確度 (r)，以此評估統計模型估計的準確性，相關係數越高則表示預測準確度越高。為選出各性狀預測準確度最佳的 GS 統計模型，利用十折交叉驗證法 (10-fold cross-validation)，將 327 品系隨機分成 10 組，除 3 組為 32 筆資料外皆為 33 筆資料，輪流取其中 9 組建立預測模型，再以餘下的一組資料代入模型得到 GEBV 計算預測準確度，最後取平均值代表此統計模型 327 品系資料的預測準確度。每種統計方法皆重複 1000 次，並利用 *set.seed* 函數設定亂數種子使全部統計方法每次重複具相同分組資

doi:10.6342/NTU201701806

料。最後以 Tukey's test 多重比較檢定各統計方法間是否具有顯著差異 ($\alpha = 0.05$)。

2.7 雙親本雜交

參考 Spindel et al. (2015) 發表的水稻育種流程，本研究規劃出如圖二及圖三之雙親本雜交流程圖，雙親本雜交後選拔個體推進至 F_6 再選出最佳品系，目標性狀為產量。每種雜交組合於 F_2 及 F_6 世代分別以 GS 與外表型選種 (phenotypic selection, PS) 兩種方法選拔產量性狀而分成兩種路徑，其餘世代皆以 PS 推進世代，藉此比較 GS 與 PS 兩種選拔方法的遺傳增進效果 (圖二)。每代模擬個數及選拔強度參考實際育種計畫，以譜系法推進世代 (圖三)。GS 路徑為每一種雜交組合產生 1000 個 F_2，選出產量 GEBV 前 100 個體自交，每株 F_2 產生 30 株 F_3，再計算 100 個家系模擬之產量外表型平均值，選出前 50 之家系並自交該家系表現最佳個體推進至 F_4。接著不斷選出 50 個家系中產量外表型最高個體自交推進至 F_6，再計算各家系產量 GEBV 平均值，選出平均值最高品系及該品系中產量 GEBV 表現最佳之單株；PS 路徑即藉由產量 GEBV 選拔部分皆改為模擬之產量外表型為選拔依據 (圖三)。親本選擇分別以 RR-BLUP 預測之產量 GEBV 前十名品系及校正外表型產量前十名品系中選出 5 個作為親本，其中有 6 個品系交集，共 14 個品系 (表二)。產量及抽穗期穩定度為第二層親本選拔標準，穩定度依據數據本身界定，在一固定環境，校正外表型及 GEBV 表現優良之前十名品系大多集中在當環境前二十名，故以二十名為穩定度分界線。最終選出親本為 A1271、A1274、B1024、B1026 及 B1027 共 5 個品系 (表三)，採半互交雜交 (half-diallel hybrids) 產生 10 種雜交組合 (表四)。

一開始假設親本皆為同型結合，子代基因體模擬使用 Yamamoto et al. (2017) 發表之 R code 完成，其 bin 大小為 0.1 cM，每條染色體發生重組次數為隨機變數，抽樣自期望值 λ 為該條染色體遺傳圖譜長度 (單位為 Morgan) 之卜瓦松分布 (Poisson distribution)，並假設重組發生位置為均勻分布 (uniform distribution)，忽

doi:10.6342/NTU201701806

略重組干擾。每個分子標誌之基因型依據最近 bin 之單倍型 (haplotype) 決定。模擬產生子代基因型資料代入 327 品系建立之預測模型中預測性狀 GEBV，此為基因型值，再加上每個世代隨機抽出之相同環境效應值及隨機抽出之個體殘差值，成為外表型值，作為 PS 之選拔依據。

2.8 輪迴選種

參考 Yamamoto et al. (2016) 發表的番茄輪迴選種流程及 Bernardo (2010) 發表的自交作物輪迴選種選拔個數，本研究規劃出圖四之水稻輪迴選種流程，目標為同時改良產量與抽穗期兩個性狀，希望族群產量越來越高、抽穗期越來越早。首先從 327 品系中選出產量前四名與抽穗期前四名之品系相互雜交，共 16 個雜交組合，每個組合產生 1 株 F_1 再自交產生 24 株 F_2，共 384 株 F_2 組成 Cycle 1 族群。接著同樣從 Cycle 1 族群中選出產量前四名與抽穗期前四名之個體相互雜交，每個組合產生 2 株 F_1 再自交產生 12 株 F_2，維持相同族群大小，共 384 株 F_2 組成 Cycle 2 族群，重複此步驟至產生 Cycle 5 族群為止。族群大小設定為 384 是為配合 96 孔盤之倍數且同時大於 327 品系數目，Cycle 2 起 F_1 選拔株數為 2 株則因 F_1 間基因型不再相同。起始親本選擇分為 GEBV 及 BLUE 兩種選拔方式，GEBV 即 GS 模型預測 327 品系之性狀表現，BLUE 為校正外表型，以 BLUE 代表真實外表型，希望藉此比較以 GEBV 或外表型選拔初始親本之輪迴選種族群改良效果差異。兩種方式選出之四個產量優良親本皆相同，而抽穗期則有三個親本不同 (表五)。

doi:10.6342/NTU201701806

三、結果

3.1 外表型資料

首先觀察 327 品系於各環境下之性狀表現，發現乾濕季明顯影響性狀表現趨勢，乾季較濕季抽穗早、株高矮、產量高；而年份效應會干擾季節的影響，2012 年株高較 2011 年高、2012 年濕季產量較 2011 年濕季高 (圖五)。以變方分析評估影響性狀表現之因子，結果顯示基因型、環境、區集，及基因型與環境之交感皆為顯著，只有產量性狀之環境主效應較不顯著 (表六)，故後續以線性混合模型校正外表型資料。表七為外表型資料校正前於四個環境下以及校正後性狀間相關係數表，抽穗期與株高為正相關，抽穗期與產量為負相關，株高與產量為弱負相關，表示這 327 優良育種品系之高產品種可能已具有抽穗早和株高矮的特性。校正後外表型值與校正前四個環境資料相關性高，抽穗期與株高校正值與各季之相關係數皆大於 0.8，產量則大於 0.7，顯示此校正外表型具代表性 (表八)。廣義遺傳率最高為 0.528 (抽穗期)，株高與產量遺傳率較低，分別為 0.382 和 0.353，可與圖五呼應，株高與產量較抽穗期易受環境效應影響 (表八)。

3.2 族群結構分析與全基因體關聯分析

以 327 品系×38,639 個 SNP 之資料所進行 PCA 之第一主成分解釋 5.48%變異，第二主成分解釋 4.14%變異 (圖六)，散佈圖看不出明顯次族群結構，驗證 Spindel et al. (2015) 之品系篩選結果，確實已剔除非 *indica* 品系。同樣以 327 品系 38,639 個 SNP 進行 GWAS 分析瞭解性狀遺傳結構，結果顯示抽穗期性狀有許多 SNP 之$-\log(p)$觀測值大於期望值，且於第 3、6 條染色體上具有顯著 SNP，尤其第 3 條染色體上顯著 SNP 較多；株高與產量性狀皆無顯著 SNP，但株高性狀有 SNP 之$-\log(p)$觀測值大於期望值，調降閾值FDR $=$ 0.1時於第 4 條染色體上有一顯著 SNP，而產量性狀FDR $=$ 0.1時仍無顯著 SNP，甚至有少許 SNP 之$-\log(p)$

doi:10.6342/NTU201701806

觀測值小於期望值 (圖七)。由此可推測抽穗期性狀由少數效應較強基因座控制；株高與產量性狀由大量微效基因座 (minor QTL) 控制，其中株高有幾個基因座效應較大。此推測又與外表型資料結果呼應，抽穗期遺傳率較高，可能因由少數效應較強基因座控制，較不易受環境效應影響；株高與產量反之。

3.3 統計模型比較

以 1000 次十折交叉驗證評估 8 種統計方法對三種性狀之預測準確度加以 Tukey's test 進行多重比較檢定 ($\alpha = 0.05$) 之結果 (表九，圖八) 顯示本研究材料之抽穗期最適用 GS 模型為 BayesB，平均預測準確度高達 0.700；株高於 RR-BLUP、BRR、BayesA、BayesB、BayesC 等預測準確度較高模型之間無顯著差異，故選擇最簡易快速之 RR-BLUP 方法於後續模擬中使用，平均預測準確度為 0.486；產量之預測準確最高模型為 RR-BLUP 及 BayesA，同樣於後續模擬中選擇使用 RR-BLUP 方法，平均預測準確度為 0.476。另外發現相同性狀不同統計方法之標準差大小相近，抽穗期標準差最小為 0.010-0.015，株高與產量分別為 0.022-0.023 及 0.020-0.021，表示這些統計方法以此份材料建立預測模型之預測準確度相當穩定 (表九)。

3.4 雙親本雜交之 GS 與 PS 比較

依據基因型表現及穩定度所選出之 5 個雜交親本為原始 327 品系中產量較高、抽穗期較早，及株高較矮之品系 (圖九)。模擬共 10 個雜交組合之 GS 與 PS 雙親本雜交選出 F_6 產量優良品系，比較選拔品系與 F_2 族群表現，發現不論 GS 或 PS，F_6 優良品系產量相較於 F_2 族群皆有顯著提升 (Welch's *t*-test $p < 0.05/21$)，但非選拔性狀之抽穗期與株高則不一定有改善，尤其 PS (表十至表十二)；另外優良品系內變異明顯較家系間變異小 (表十至表十二)。10 個雜交組合之 F_2 族群性狀表現多集中於兩親本間，符合預期，因模型只考慮累加性效應 (圖十)。比較 GS 與 PS 兩種選拔途徑選出之 F_6 優良品系各性狀表現，結果顯示，除第八個雜

doi:10.6342/NTU201701806

交組合外共有九個雜交組合之 GS 產量優於 PS，六個雜交組合中 GS 選拔之品系較 PS 早抽穗，八個雜交組合顯示 GS 選拔之品系株高較 PS 矮 (圖十至圖十二)。雖然只選拔產量性狀，但抽穗期與株高也有一定程度改善，應與親本同時具高產、抽穗早與株高矮特性有關。結果顯示在此模擬育種流程中，GS 產量選拔效率優於 PS。

3.5 輪迴選種

分別以 GEBV 及 BLUE 選拔初始親本進行五次循環 (cycle) 之輪迴選種，隨著循環數增加，兩種選拔初始族群方式之族群產量與抽穗期皆有顯著改進；以 BLUE 選拔初始親本之輪迴選種情境，株高雖非目標性狀，觀察其隨循環數增加漸次下降，但 GEBV 選拔族群僅在 Cycle 1 明顯株高降低，之後無顯著差異 (圖十三，表十三)。總的說來，使用 GEBV 進行親本選拔之後代產量優於 BLUE (Welch's *t*-test, $p < 0.0001$)，BLUE 選拔較 GEBV 選拔親本所產生之後代早抽穗且較矮 (圖十四)。進一步觀察輪迴選種各循環 F_2 族群相對 327 品系之標準化遺傳增進變化 (抽穗期與株高取絕對值)，發現兩初始族群之產量性狀走勢相近，GEBV 一開始即高於 BLUE，後期差距縮小；抽穗期一開始 GEBV 遺傳增進幅度較 BLUE 大，但於 Cycle 5 被 BLUE 超越；GEBV 株高性狀之遺傳增進從 Cycle 1 後就沒有明顯變化，僅在 Cycle 5 稍高，而 BLUE 之遺傳增進於 Cycle 2 大幅提升，又於 Cycle 4、5 穩定提升 (圖十五)。計算輪迴選種各循環 F_2 世代 5,264 個 SNP 之對偶基因 1 頻度，結果顯示兩種初始族群對偶基因頻度變化趨勢相同，皆於 Cycle 2 大量固定 (頻度為 1) 或小部分由族群中移除 (頻度為 0)，後續被固定基因型亦不斷增加 (圖十六)；對照各循環族群之性狀變異，發現族群標準差不斷下降，抽穗期至 Cycle 2、產量與株高至 Cycle 3 後即無太大差異 (表十三)，符合對偶基因 1 頻度變化結果，顯示 GS 使得輪迴選種於短期內快速累積有利對偶基因。

doi:10.6342/NTU201701806

四、討論

4.1 統計模型預測準確度

4.1.1 影響預測準確度因子

影響預測準確度因子很多，包括訓練族群大小、訓練族群與驗證族群相關性、統計模型、分子標誌數量與分布、外表型調查、性狀遺傳結構與遺傳率等 (Desta and Ortiz 2014; Lorenz et al. 2011; Zhao et al. 2015)。本研究 1000 次十折交叉驗證之訓練族群和驗證族群來自相同 327 品系，品系間無次族群結構，每次重複族群大小相同；黑小麥、番茄及大麥研究指出訓練族群和驗證族群的相關性越高則預測準確度越好 (Duangjit et al. 2016; Wurschum et al. 2017)，且比族群大小重要，族群小但與選拔族群關係緊密的訓練族群，較數量大、多樣性高但與選拔族群相關性低的訓練族群有更好的預測能力 (Lorenz and Smith 2015)。8 種統計模型三種性狀 1000 次預測準確度變方範圍 0.01-0.023 相當小 (表九)，或許即因本研究材料族群結構簡單且品系間相關性高，使得預測準確度穩定且變化不大。分子標誌數量與分布係依照 Spindel et al. (2015) 之結果從 38,639 個 SNP 每 50 kb 區間隨機抽取一個 SNP，共 5,264 個 SNP，與作者以 50 kb 為抽樣區間共抽出 7,142 個 SNP 相比少了近 2,000 個分子標誌。但作者以 120 kb 為間隔抽出 3,076 個 SNP 之預測準確度與 7,142 個 SNP 差異不大，故推測利用本研究所抽出之 5,264 SNP 應仍在最佳準確度範圍內。另外分子標誌均勻分布基因體或隨機分布對預測準確度影響則與分子標誌數量有關，數量越大則兩者差異不大，均勻分布稍優於隨機分布，尤其 BL 模型差異較大；分子標誌數量越少則隨機抽樣之準確度下降較均勻分布劇烈。但是在另一篇大豆研究中卻有不同結果，產量性狀均勻抽樣之分子標誌預測準確度反倒較隨機抽樣稍低 (Ma et al. 2016)。造成兩種分子標誌抽樣方式預測準確度高低不同的原因，可能為原始分子標誌分布在目標性狀 QTL 附近情

doi:10.6342/NTU201701806

況不同，Ma et al. 已先挑選靠近 QTL 的 SNP，因此隨機抽樣可能比等距抽樣更有效地選出與性狀 QTL 連鎖之 SNP，使得預測準確度較高。

4.1.2 性狀遺傳結構與統計模型關係

性狀之遺傳結構將影響統計模型預測準確度，一般認為 RR-BLUP 和 BRR 適合由大量 QTL 控制之複雜性狀；BL、BayesA、BayesB、BayesC 適合由少數 QTL 控制之簡單性狀；RKHS 適合具非累加性效應之性狀；RF 適合具基因上位性之性狀 (Desta and Ortiz 2014; Lorenz et al. 2011; Onogi et al. 2015)。發表本研究族群資料之團隊同時也有做 GWAS 分析 (Begum et al. 2015)，性狀遺傳結構判斷結果與以上 GS 適用模型之認知相互符合。其結果顯示抽穗期由少數 QTL 控制，其中第 3 條染色體上有一強效 QTL 於乾濕季之外表型變異解釋量 (phenotypic variance explained, PVE) 高達 40%以上；株高於第 2、3、6、8 條染色體上具顯著 SNP，僅第 3 條染色體上 SNP 之 PVE 接近 15%，其餘 PVE 約 6%上下，顯示株高由一定數量 QTL 控制；產量則僅有濕季於第 11 條染色體上有顯著 SNP，PVE 小於 10%，應由大量微效 QTL 控制。本研究 GWAS 結果於第 3、6 條染色體上發現與抽穗期性狀相關之顯著 SNP，其中第 3 條染色體最顯著之 SNP 於 Begum et al. (2015) 研究中同樣為最顯著 SNP，乾季之 PVE 高達 43%，濕季之 PVE 為 28%，此 SNP 也與株高性狀相關，乾季之 PVE 達 12%；第 6 條染色體最顯著之 SNP 在作者研究中同樣與抽穗期性狀相關，乾季之 PVE 為 2%。株高性狀調升閾值至FDR ＝ 0.1時僅於第 4 條染色體上有一顯著 SNP，但原作者在第 2、3、6、8 條染色體上顯著的 SNP 並未在本研究呈現顯著，可能因株高 QTL 之 PVE 不大或原著中顯著的 SNP 未被選取。本研究產量性狀無顯著 SNP，作者也僅於濕季在第 11 條染色體上有顯著 SNP，再次分析結果亦符合產量應受大量微效 QTL 控制。

本研究所使用族群基因型資料全為同型結合，不存在顯性效應，RKHS 優勢小，三種性狀之預測準確度皆為最差 (圖八)。各統計模型於抽穗期性狀之表現，BayesB、BayesA、RF 預測準確度較佳，RR-BLUP 和 BRR 僅比 RKHS 稍高，符

doi:10.6342/NTU201701806

合 GWAS 推測抽穗期由少數 QTL 控制故適合 BayesB 模型，且該性狀可能具上位性效應；株高性狀則除 RKHS 和 RF 外皆無顯著差異，對照 GWAS 結果株高性狀可能由一定數量 QTL 控制；產量顯然為大量 QTL 控制之複雜性狀，研究結果最適合模型為 RR-BLUP 和 BayesA，BL 僅比 RF 和 RKHS 稍好，BayesA 預測準確度高或許因為此模型只假設每個分子標誌具不同變方而沒有變數選擇仍包含全部分子標誌 (圖七至圖八)。

對照 Spindel et al. (2015) 和 Onogi et al. (2015) 的研究，前者以相同水稻族群不同交叉驗證資料組比較 RR-BLUP、BL、RKHS 和 RF 等 6 種方法對於抽穗期、株高及產量之預測準確度，後者則以日本 110 個水稻品系作為材料，比較包括 GBLUP (與 RR-BLUP 結果相同)、BL、RKHS 和 RF 在內共 9 種方法於抽穗期和株高等性狀之預測準確度。兩者結果皆顯示 RF 對於抽穗期有良好預測能力，前者顯示 RR-BLUP 適合預測產量；但 RKHS 和 RF 預測三種性狀能力皆同樣好，反而 BL 為四種方法中各性狀預測準確度最差，尤其抽穗期性狀 RR-BLUP 表現比 BL 好。Spindel et al. (2015) 以乾季資料作為驗證組估計預測準確度，抽穗期預測準確度最高為多元線性迴歸模型 (multiple linear regression, MLR) 之 0.627，株高為 RF 之 0.3411，產量為 RR-BLUP 之 0.3044；本研究則分別是 BayesB 之 0.7，RR-BLUP 之 0.486，RR-BLUP 之 0.476，相較之下準確度較高 (表九)。可能造成此結果差異原因包括不同校正外表型方式、交叉驗證資料組及分子標誌數量，許多因子影響統計模型預測準確度而非簡單通則。Onogi et al. (2015) 另外模擬不同訓練族群大小、QTL 數量、遺傳率和上位性效應存在與否等情境下各統計模型之預測準確度，結果顯示 BL 穩定度最佳，於各情境下準確度排名皆不錯，RF 只在訓練族群大小為 100 時表現良好，RKHS 於存在上位性效應情境表現佳，GBLUP 在訓練族群大與遺傳率高之情境下具優勢。顯然 BL 雖然穩定度佳，但在特定情境下仍須測試才知道最適模型。建議實際使用 GS 育種前先測試育種族群各性狀所適合統計模型，甚至測試模型最佳參數值，以便提高預測準確度。

doi:10.6342/NTU201701806

4.1.3 利用歷史資料建立預測模型

許多 GS 研究不會只使用單次區域試驗資料而會結合多年資料校正外表型再建立預測模型 (Habyarimana et al. 2017; Technow et al. 2014; You et al. 2016)，以更準確地評估各品系基因型值並提高預測準確度 (He et al. 2016; Lado et al. 2016)。又，結合單一年份多地點資料或任一年份、地點定義之下大環境 (maga-environment, ME) 資料可提升在不同環境下之模型預測準確率 (Lado et al. 2016)。本研究使用線性混合效應模型校正外表型，模型中將年份與季節效應結合成環境效應，不考慮基因型與環境間交感效應。計算抽穗期、株高和產量之廣義遺傳率分別為 0.528、0.382 和 0.353 (表八)，Spindel et al. (2015) 計算 2012 年乾季之遺傳率則分別為 0.4378、0.3546 和 0.3213，遺傳率之範圍及各性狀遺傳率相對大小類似。校正外表型與校正前四個環境資料相關性高皆大於 0.7 具代表性 (表八)，此校正方式應無問題。在測試各模型預測準確度時，曾分別以各別環境資料建立 RR-BLUP 與 BL 之預測模型，結果顯示使用校正外表型資料建立預測模型能夠提高預測準確度 (附圖一)。

4.2 雙親本雜交

4.2.1 探討雙親本雜交產量性狀 GS 效率較 PS 高之原因

本研究中所模擬的基因型值與外表型值差別只在殘差項，因此本研究 GS 與 PS 的比較為理想狀態，因為我們將 GS 模型預測值視同真實基因型值而未考慮準確度非 100%的問題。實際影響 GS 與 PS 相對效率因子為遺傳率：產量遺傳率為 0.353，327 品系校正外表型標準差為 515.71，經 RR-BLUP 壓縮後 GEBV 標準差為 336.47，雙親本雜交後代變異更低，而殘差標準差為 663.23，極易掩蓋基因型表現。遺傳率低之性狀外表型選拔不易，因此模擬結果 GS 產量優於 PS，僅有一雜交組合沒有差異 (圖十)。考慮 GS 預測準確度問題，理所當然準確度越高越好，但預測準確度與遺傳率相關性高 (Duangjit et al. 2016; You et al. 2016)，模擬

doi:10.6342/NTU201701806

研究顯示遺傳率越高則 GS 模型預測準確度越高 (Bernardo and Yu 2007)，因此以預測準確度除以遺傳率 (r/h^2) 評估 GS 對 PS 之相對效率 (Dekkers 2007; You et al. 2016; Ziyomo and Bernardo 2013)。本研究抽穗期、株高與產量最佳模型之平均相對效率分別為 1.33、1.27 和 1.35，皆大於 1，表示 GS 相對 PS 效率較高。有研究顯示 GS 遺傳增進優勢會隨著遺傳率提高而減少 (Rajsic et al. 2016; You et al. 2016)，考慮經濟效益 GS 在遺傳率小於 0.25 之性狀較有利 (Rajsic et al. 2016)。實際大麥資料比較產量、赤霉病 (Fusarium head blight, FHB) 感病性和脫氧雪腐鐮刀菌烯醇 (deoxynivalenol, DON) 濃度等性狀，性狀遺傳率為 0.54-0.82，單一世代 GS 預測準確度 0.32-0.99，而 GS 相對 PS 遺傳增進率最低為 0.16 但許多大於 1，顯示 GS 表現可比 PS 更好 (Sallam and Smith 2016)。即使 GS 相對效率較低，若每年利用 GS 加速世代推進能比 PS 選拔更多次，則 GS 單位時間內育種效率較高，如溫帶玉米育種 (Massman et al. 2013)。

4.2.2 實際育種計畫之應用

同上節一開始所言，本研究結果為模擬顯示之理想狀態，不等同實際情形，因此本研究結果具有一定限度。因能力所限，基因效應只考慮累加性效應而沒有考慮顯性效應和上位性效應、校正外表型時結合年份和季節資料簡化環境變異、不考慮基因型與環境交感影響，並且在世代推進時直接將 GS 預測之 GEBV 視為基因型值。雖然本研究結果鼓勵使用 GS，但是在這樣簡化情境之下，若要實際應用在育種計畫中有許多事項須得留意，比如訓練族群與育種族群關係、訓練族群大小、使用分子標誌數量等，以提升最重要的 GS 預測能力，才可能較 PS 選拔效率高。在建立預測模型時，可使用更多地點年份之歷史資料校正外表型提高預測能力 (Wang et al. 2014; Zhao et al. 2015)。

建立預測模型之訓練族群組成很重要，與育種族群相關性越高則預測準確度越高 (Duangjit et al. 2016; Wurschum et al. 2017)，因此若能使用育種族群建立預測模型效果最好，或是加入過去育種循環資料校正預測模型 (Auinger et al. 2016)，

doi:10.6342/NTU201701806

訓練族群與育種族群間需有足夠共同祖先否則預測能力將大幅下降 (Gowda et al. 2014; Zhao et al. 2015)。訓練族群間外表型變異越大越好，表示 QTL 散佈在族群中能更好估計分子標誌效應 (Marulanda et al. 2015)。訓練族群大小主要是增加模型穩定度 (Zhao et al. 2015)，不如與育種族群相關性影響大 (Lorenz and Smith 2015)。育種過程中 GS 模型準確度會下降且可能喪失在起使族群中頻度小之有利對偶基因 (Goddard 2009; Lorenz et al. 2011; Storlie and Charmet 2013; Zhao et al. 2015)。目前已發表之 GS 研究多以短期育種計畫為例，若要長期使用 GS 育種必須重新校正模型才能更穩定 (Zhao et al. 2015)。

分子標誌密度相對訓練族群組成對 GS 模型預測能力影響較小 (Zhao et al. 2015)，分子標誌數量適度即可，過多則無助益 (Duangjit et al. 2016; Gorjanc et al. 2017; Spindel et al. 2015)。目前有些研究使用 GBS 方法產生之 SNP 資料探討 GS 模型 (Ashraf et al. 2016; Dunckel et al. 2017; Spindel et al. 2015)，並認為利用此高通量基因型鑑定技術能夠得到大量分子標誌資料降低成本 (Poland et al. 2012; Sallam and Smith 2016)，但在 GBS 方法於實際育種流程中無法保證後代具有相同分子標誌資料，因此在實際世代推進研究中使用已開發之分子標誌套組或基因晶片為宜 (Spindel et al. 2015; Thomson 2014)。還有一些研究從經濟效益及遺傳增進角度探討最佳 GS 育種策略，在相同資源下比較雙單倍體 (doubled haploid, DH) 族群育種中各階段不同族群大小及只有 GS、GS 一次後接 PS 一次或兩次之選拔效率，結果顯示若 GS 預測準確度小於 0.5，則在 GS 選出優良 DH 品系後再進行一次區域試驗 PS 能夠最大提升一年選拔效率 (Longin et al. 2015; Marulanda et al. 2016)。

doi:10.6342/NTU201701806

4.3 輪迴選種

4.3.1 族群多樣性下降

本研究之輪迴選種模擬如同雙親本雜交模擬皆為理想情況，直接假定 GS 模型預測準確度 100%來觀察其育種潛力，因此並未再與 PS 比較選拔效率。結果顯示族群產量與抽穗期皆隨著循環數增加有顯著改進 (圖十三)，但族群變方不斷下降 (表十三)，對偶基因頻度於 Cycle 2 即大量固定 (圖十六)，顯然族群多樣性大幅降低。回頭檢視育種流程後發現，實際上進入 Cycle 2 的初始親本 (Cycle 0) 只有 5 個共 4 種雜交組合，皆為產量選拔出之前二名和三個依據抽穗期所選拔的品系，之後每次選拔幾乎有兩個個體來自相同雜交組合。為確認是否因為優良親本間親緣關係緊密造成性狀表現相似，我們藉由估計 Identity-by-state 值 (IBS = $\frac{\text{No. of same alleles}}{(\text{No. of shared loci})\times 2}$) 估算個體間的相似程度：327 *indica* 品系的 IBS 主要集中在 0.7 左右，*japonica* 品系間以及 *japonica* 與 *indica* 品系之 IBS 則集中在 0.65 左右，*japonica* 品系 IBS 超出預期較 *indica* 低，可能這些 *japoinca* 品系橫跨熱帶與溫帶次族群 (附圖二)。上述之 GEBV 和 BLUE 進入 Cycle 2 之四個雜交組合之親本兩兩間之 IBS 分別為 0.69、0.73、0.73、0.79 和 0.61、0.66、0.67、0.69。我們利用雙親本雜交模擬產生之 F_6 族群 50 個家系抽取 150 單株計算姊妹系 IBS，其數值皆大於 0.7，主要範圍為 0.8 以上，甚至有的雜交組合姊妹系之 IBS 高於 0.93 (附圖二)。此數據顯示 Cycle 0 選出之優良親本雜交組合間應該不是相近之姊妹系，只因這些品系的 GEBV 過於優異，造成所選拔的個體主要來自這些品系的後裔。另外發現產量前二名優良品系 B1024 與 B1027 之 IBS 高達 0.91 應為姐妹系，且只有這兩個品系之後裔進入 Cycle 2。本研究中輪迴選種每次循環從 384 個體中選出 8 個作為下一階段之親本，選拔強度相當高 (2.1%)，應是造成族群多樣性快速下降之主因。一篇 GS 對雙親本玉米族群多樣性影響之研究結果顯示當選拔強度為 50%時，GS 個體遺傳相似性與 PS 個體無顯著差異；但選拔強度提高至 25%以

doi:10.6342/NTU201701806

下後，GS 個體遺傳相似性稍微提升顯著大於 PS 個體 (Jacobson et al. 2015)。設定選拔強度為 10%，實際大麥資料比較 GS 相對 PS 選拔效率之研究顯示 GS 個體遺傳相似性顯著提升 (Sallam and Smith 2016)。Bernardo et al. (2006) 建議 MAS 於輪迴選種時選拔個體數約等於預計執行輪迴選種之循環數，依本研究結果每次選出個體幾乎至少有兩個來自相同雜交組合，或許選拔個數可調降為 6 個即有相似程度之選拔增進。GS 能夠較 PS 更快速增加有利對偶基因頻度 (Sallam and Smith 2016)，達到輪迴選種主要目的，短期育種計劃沒有問題；但若想要執行長期輪迴選種希望維持族群變異，選拔強度就不能太強，且須更新預測模型 (Zhao et al. 2015)。

4.3.2 初始親本選拔方式

因 Bernardo and Yu (2007) 前置實驗指出在基因體選種前的 Cycle 0 先以外表型選拔比以外表型結合分子標誌資料選拔效果更好，本研究分成 GEBV 和 BLUE 兩種初始親本選拔方式。兩者選出相同產量優良親本，只有抽穗期優良親本有差異。Cycle 5 結果顯示 GEBV 產量優於 BLUE，而抽穗期和株高則是 BLUE 優於 GEBV。推測 GEBV 產量較高是因抽穗期優良親本中之 B1025 品系產量 GEBV 也很好 (表二與表三)，但仔細觀察遺傳增進變化發現三個性狀於 Cycle 1 皆為 GEBV 高於 BLUE，之後差距縮小甚至抽穗期和株高 BLUE 超過 GEBV (圖十五)。再看對偶基因頻度變化，BLUE 於 0.2-0.8 間變動，幅度較 GEBV 為大 (圖十六)，BLUE 於 Cycle 2 之頻度中位數較 GEBV 低且之後主要頻度範圍也較大 (附圖三)。BLUE 進入 Cycle 2 的四個雜交組合之 IBS 也略低於 GEBV，顯示 BLUE 遺傳多樣性較 GEBV 高，多次選拔潛力優於 GEBV。Bernardo and Yu (2007) 也不清楚為何結合分子標誌資料選拔不如外表型選拔，並推測可能族群太小導致無法估得較正確之基因型值用以選拔，或是估得分子標誌效應後再回頭估計訓練族群之 GEBV 屬於後測 (post-dictive) 而非預測 (pre-dictive) 方式因此使得效益有限。雖然因沒有重複模擬無法確定以 BLUE 選拔初始親本造成之後裔抽

doi:10.6342/NTU201701806

穗期減少並非遺傳重組造成的偶然，但以外表型選拔能全面調查作物整體性狀表現，可先行淘汰目標性狀預測值優良但其他性狀不佳的個體 (Massman et al. 2013)。

4.4 雙親本雜交與輪迴選種

雙親本雜交與輪迴選種結果皆顯示非選拔性狀 (抽穗期或株高) 較 327 品系早和矮，但最顯著的遺傳增進發生在早世代選拔，之後的變化幅度較小，僅輪迴選種 BLUE 選拔初始親本族群隨著循環數增加株高不斷降低，從雙親本雜交選出之親本校正外表型判斷應是此族群產量高之品系同時具備早抽穗、株高矮特性之故 (圖九)。或許因此族群材料為進入區域產量試驗之品系，已為育種選出之優良品系，故具高產又早抽穗之違背一般生理原則特性。Yamamoto et al. (2016) 不只模擬輪迴選種，在每個 Cycle 另外模擬產生自交系 (inbred lines, ILs)，結果顯示自交系的性狀表現皆在其衍生之族群附近，隨著輪迴選種循環數增加才有明顯增進，否則只雜交一次與選拔一次的自交系表現無法達到育種目標，顯然如輪迴選種般多次雜交重組與選拔是必要的。本研究輪迴選種以 GEBV 選拔親本之 Cycle 5 平均產量和最大值分別為 5948 kg/ha 和 6195 kg/ha (表十三)，高於雙親本雜交 10 個組合各自 F_6 族群產量平均值 5641-5926 kg/ha 和 9 個組合之最大值 5936-6232 kg/ha (表十)，兩者同為經過 5 次選拔。雖然輪迴選種可能因族群多樣性下降導致產量增長幅度不大，但平均值仍略高於 F_6 族群，且優良個體表現較佳，顯然多次雜交重組仍有效改良族群表現並產生優良個體。玉米等異交作物常以輪迴選種改良族群再選出優良個體做雙親本雜交，水稻雖為自交作物但仍可透過輪迴選種獲得親本材料，若族群具有雄不稔基因可便利雜交工作的進行 (Grenier et al. 2015)，再藉由 GS 能加速世代推進、縮小種植規模並節省大量調查外表型工作。

doi:10.6342/NTU201701806

五、結論

本研究以模擬方式探討在世代推進的情形下 GS 在水稻育種計畫中之應用，證實在理想狀況下，基因體選種於雙親本雜交之產量性狀選拔效率優於外表型選種，且基因體選種在輪迴選種中能有效改良族群抽穗期與產量性狀表現、快速累積有利對偶基因頻度。未來若欲將 GS 應用在實際育種計畫中，須首先了解訓練族群與育種族群關係及目標性狀最適統計模型以提高預測準確度，可參考本研究之流程，先以 PCA 分析族群結構，接著以 GWAS 分析目標性狀之遺傳結構，再透過交叉驗證選出各性狀最佳統計模型。接著或可先以短期輪迴選種快速累積有利對偶基因頻度改良族群，從中選拔優良品系進行純化或選出優良親本進行雜交，於 F_2 早世代及 F_6 以 GS 方式節省大量人力且較精確選拔出優良品系。

doi:10.6342/NTU201701806

表一、327 品系在各年份、季節的品系數、區集數和各性狀缺值數

Table 1. The No. of lines, the No. of blocks, and the No. of missing data for 327 lines in each year and each season

Year	Season	lines with data	No. Blocks	No. FL missing	No. PH missing	No. Yield missing
2011	dry	326	3	0	0	4
2011	wet	326	3	1	0	1
2012	dry	325	2	0	0	5
2012	wet	323	3	2	0	3

doi:10.6342/NTU201701806

表二、GEBV 預測產量前十名品系及校正外表型產量 BLUE 前十名品系

Table 2. The top ten lines in yield of GEBV and BLUE

GEBV		BLUE	
Entry	Yield	Entry	Yield
B1024	6371.054	B1024	6247.8
B1027	6370.615	B1027	6091.709
B1026	6160.164	A1271	5989.436
A1271	6115.114	B1026	5914.709
B1019	6018.69	B1118	5894.527
M1476	6003.656	A1274	5816.527
A1338	5993.299	M1470	5812.963
M1470	5984.14	M1476	5808.073
B1025	5983.807	B1147	5752.982
M1405	5961.804	B1070	5685.436

灰底為兩者重疊之品系。

Lines overlapped in both estimation are highlighted in gray.

doi:10.6342/NTU201701806

表三、產量 GEBV 前十名品系及校正外表型產量 BLUE 前十名品系其產量及抽穗期平均值於各季原始排名前二十名之整理

Table 3. Summary for the GEBV-top 10 and the BLUE-top 10 which were among the top-20 in each trial

GEBV			BLUE		
Entry	Yield top 20th season	FL top 20th season	Entry	Yield top 20th season	FL top 20th season
B1024*	2011D, 2011W, 2012D, 2012W	2011W	B1024*	2011D, 2011W, 2012D, 2012W	2011W
B1027*	2011D, 2011W, 2012D, 2012W	2011D, 2011W, 2012W	B1027*	2011D, 2011W, 2012D, 2012W	2011D, 2011W, 2012W
B1026*	2011W, 2012D, 2012W	2011D, 2011W, 2012W	A1271*	2011D, 2011W, 2012D, 2012W	2012D
A1271*	2011D, 2011W, 2012D, 2012W	2012D	B1026*	2011W, 2012D, 2012W	2011D, 2011W, 2012W
B1019	2011W, 2012W	2011W, 2012W	B1118	2011W, 2012D, 2012W	2012D, 2012W
M1476	2011D, 2011W, 2012D	2011D, 2012D	A1274*	2011W, 2012D, 2012W	2011W, 2012D, 2012W
A1338	2011D, 2012D		M1470	2012D & 2012W NA	2012D & 2012W NA
M1470	2012D & 2012W NA	2012D & 2012W NA	M1476	2011D, 2011W, 2012D	2011D, 2012D
B1025	2011D, 2011W	2011W, 2012W	B1147	2011D, 2012D	
M1405	2011W	2012W	B1070	2012D, 2012W	

灰底為兩者重疊之品系。2011 與 2012 為年分，D 為乾季，W 為濕季。M1470 品系在 2012 年兩季缺值故不列入考慮。具*之品系為選出之親本。

Lines overlapped in both estimation are highlighted in gray. 2011 and 2012 are year of trial, D stands for dry season, W stands for wet season. M1470 line is not taken into consideration for its value was missing in both dry and wet seasons in 2012. The lines with the symbol * are the selected parental lines.

doi:10.6342/NTU201701806

表四、雙親本雜交之親本

Table 4. The parents of bi-parental crosses

	Parent 1	Parent 2
Cross 1	A1271	A1274
Cross 2	A1271	B1024
Cross 3	A1271	B1026
Cross 4	A1271	B1027
Cross 5	A1274	B1024
Cross 6	A1274	B1026
Cross 7	A1274	B1027
Cross 8	B1024	B1026
Cross 9	B1024	B1027
Cross 10	B1026	B1027

doi:10.6342/NTU201701806

表五、輪迴選種起始親本

Table 5. The initial parents (Cycle 0) for recurrent selection

GEBV		BLUE	
Yield	FL	Yield	FL
B1024	B1007	B1024	A1311
B1027	B1035	B1027	A1259
B1026	B1025	A1271	M1463
A1271	B1008	B1026	B1007

doi:10.6342/NTU201701806

表六、外表型變方分析表

Table 6. The ANOVA table of phenotype

Source	d.f.	Type III SS	MS	*F*	*P*
FL					
Line	326	24652.5	75.6211	24.4444	< 0.001
Environment	3	46.0663	15.3554	4.96362	0.00195
Block (E)	7	189.668	27.0954	8.75855	< 0.001
Line × Environment	970	16011	16.5061	5.33558	< 0.001
Residuals	2265	7007	3.0936		
PH					
Line	326	53821.4	165.096	5.04052	< 0.001
Environment	3	673.986	224.662	6.85911	< 0.001
Block (E)	7	1257.81	179.687	5.48597	< 0.001
Line × Environment	970	55770.9	57.4958	1.75539	< 0.001
Residuals	2265	74285.7	32.7972		
Yield					
Line	326	4.8×10^{8}	1.5×10^{6}	4.5049	< 0.001
Environment	3	2.1×10^{6}	714760	2.20257	0.08582
Block (E)	7	9.8×10^{6}	1.4×10^{6}	4.32246	< 0.001
Line × Environment	969	6.9×10^{8}	711583	2.19278	< 0.001
Residuals	2256	7.3×10^{8}	324512		

doi:10.6342/NTU201701806

表七、外表型資料性狀間相關係數

Table 7. The coefficients of correlation between phenotypes

	2011 Dry	2011 Wet	2012 Dry	2012 Wet	BLUE
r(FL, PH)	0.21	0.32	0.43	0.28	0.33
r(FL, Yield)	-0.31	-0.56	-0.25	-0.15	-0.42
r(PH, Yield)	-0.088	-0.26	-0.12	-0.088	-0.19

FL 為抽穗期、PH 為株高、Yield 為產量，BLUE 為校正外表型。樣本數介於 323 至 327 之間。

FL, flowering time; PH, plant height; BLUE, the adjusted phenotypic values. The sample size is between 323 and 327.

doi:10.6342/NTU201701806

表八、外表型資料校正前後之相關係數與廣義遺傳率 (h^2)

Table 8. The coefficients of correlation between trait values before and after adjustment, and the broad-sense heritability

	2011 Dry	2011 Wet	2012 Dry	2012 Wet	h^2
FL	0.88	0.92	0.91	0.88	0.528
PH	0.86	0.87	0.84	0.91	0.382
Yield	0.78	0.73	0.79	0.76	0.353

doi:10.6342/NTU201701806

表九、8 種統計方法於各性狀之預測準確度

Table 9. The prediction accuracy of eight statistical methods for each trait

	FL	PH	Yield
RR-BLUP	0.595±0.015 [e] (0.544-0.648)	0.486±0.022 [a] (0.407-0.550)	0.476±0.020 [a] (0.398-0.543)
BRR	0.594±0.015 [e] (0.542-0.641)	0.483±0.022 [ab] (0.402-0.548)	0.469±0.021 [de] (0.377-0.545)
BL	0.623±0.014 [d] (0.573-0.674)	0.482±0.022 [b] (0.401-0.543)	0.467±0.021 [e] (0.386-0.545)
BayesA	0.690±0.013 [b] (0.636-0.731)	0.485±0.022 [a] (0.406-0.546)	0.474±0.021 [ab] (0.385-0.537)
BayesB	0.700±0.010 [a] (0.662-0.741)	0.486±0.022 [a] (0.405-0.544)	0.473±0.021 [bc] (0.395-0.54)
BayesC	0.636±0.015 [c] (0.584-0.68)	0.484±0.022 [ab] (0.409-0.554)	0.470±0.021 [cd] (0.397-0.535)
RKHS	0.579±0.015 [f] (0.524-0.63)	0.473±0.023 [c] (0.376-0.543)	0.446±0.021 [f] (0.359-0.516)
RF	0.689±0.010 [b] (0.653-0.72)	0.481±0.023 [b] (0.407-0.551)	0.448±0.020 [f] (0.373-0.505)

平均值±標準差，括弧內為最小值及最大值，字母顯示 Tukey's test 的結果 ($\alpha = 0.05$)，字母不同代表具顯著差異。FL 為抽穗期、PH 為株高、Yield 為產量。

Mean ± standard deviation; minimum and maximum are provided between the brackets; the letters show Tukey's test results ($\alpha = 0.05$), different letters signify significant differences. FL, flowering time; PH, plant height.

doi:10.6342/NTU201701806

表十、GS 及 PS 之產量 GEBV 結果

Table 10. The Yield GEBV for GS and PS

Yield	mid-parent	F_2	F_2		F_6		F_6 best line	
			GS	PS	GS	PS	GS	PS
Cross 1	5496	5500 ± 71 [f] (5277 - 5732)	5623 ± 29 [c] (5591 - 5732)	5525 ± 70 [e] (5309 - 5673)	5641 ± 85 [b] (5393 - 5936)	5583 ± 80 [d] (5350 - 5797)	5863 ± 42 [a] (5802 - 5936)	5589 ± 9 [d] (5572 - 5606)
Cross 2	5717	5713 ± 98 [f] (5417 - 6006)	5889 ± 32 [d] (5845 - 6006)	5730 ± 87 [f] (5508 - 5962)	5926 ± 86 [c] (5699 - 6155)	5810 ± 108 [e] (5564 - 6042)	6106 ± 20 [a] (6065 - 6148)	5978 ± 21 [b] (5932 - 6015)
Cross 3	5611	5617 ± 110 [g] (5292 - 6065)	5815 ± 52 [d] (5752 - 6065)	5675 ± 110 [f] (5471 - 5940)	5844 ± 120 [c] (5566 - 6117)	5768 ± 119 [e] (5472 - 6124)	6073 ± 21 [a] (6035 - 6117)	5954 ± 22 [b] (5917 - 6012)
Cross 4	5717	5716 ± 111 [f] (5364 - 6040)	5906 ± 39 [d] (5858 - 6040)	5747 ± 117 [f] (5423 - 6040)	5923 ± 129 [c] (5496 - 6232)	5848 ± 119 [e] (5580 - 6221)	6199 ± 22 [a] (6155 - 6232)	6041 ± 15 [b] (6008 - 6068)
Cross 5	5624	5624 ± 106 [e] (5304 - 5967)	5806 ± 44 [c] (5757 - 5967)	5652 ± 99 [e] (5436 - 5907)	5821 ± 108 [c] (5551 - 6189)	5734 ± 121 [d] (5420 - 5996)	6127 ± 32 [a] (6061 - 6189)	5857 ± 38 [b] (5795 - 5942)
Cross 6	5519	5522 ± 110 [f] (5203 - 5940)	5719 ± 49 [d] (5664 - 5940)	5553 ± 108 [f] (5295 - 5806)	5764 ± 118 [c] (5498 - 6118)	5670 ± 125 [e] (5369 - 5966)	6070 ± 19 [a] (6033 - 6118)	5899 ± 37 [b] (5814 - 5966)
Cross 7	5624	5622 ± 116 [g] (5273 - 5925)	5817 ± 39 [d] (5770 - 5925)	5677 ± 98 [f] (5439 - 5863)	5879 ± 128 [c] (5520 - 6148)	5759 ± 126 [e] (5437 - 6088)	6088 ± 16 [a] (6061 - 6120)	5917 ± 30 [b] (5871 - 5981)
Cross 8	5739	5739 ± 57 [d] (5565 - 5907)	5840 ± 21 [b] (5813 - 5907)	5750 ± 58 [d] (5616 - 5906)	5846 ± 40 [b] (5715 - 5943)	5771 ± 75 [c] (5578 - 5972)	5914 ± 2 [a] (5909 - 5917)	5916 ± 19 [a] (5891 - 5945)
Cross 9	5845	5845 ± 53 [d] (5683 - 6031)	5931 ± 19 [b] (5910 - 6031)	5848 ± 55 [d] (5718 - 5969)	5924 ± 48 [b] (5789 - 6036)	5870 ± 75 [c] (5667 - 6034)	6031 ± 4 [a] (6027 - 6036)	5871 ± 3 [c] (5866 - 5875)
Cross 10	5739	5737 ± 66 [f] (5555 - 5949)	5855 ± 22 [c] (5827 - 5949)	5748 ± 69 [f] (5560 - 5878)	5870 ± 51 [b] (5751 - 5977)	5791 ± 81 [e] (5566 - 5968)	5970 ± 5 [a] (5962 - 5977)	5810 ± 9 [d] (5797 - 5832)

平均值±標準差，括弧內為最小值及最大值，同雜交組合不同字母具顯著差異 ($p < 0.05/21$)。

Mean ± standard deviation; minimum and maximum are provided between the brackets; different letters signify significant differences ($p < 0.05/21$).

doi:10.6342/NTU201701806

表十一、GS 及 PS 之抽穗期 GEBV 結果

Table 11. The flowering time GEBV for GS and PS

FL	mid-parent	F_2	F_2		F_6		F_6 best line	
			GS	PS	GS	PS	GS	PS
Cross 1	84.1	84.0 ± 0.5 ^{c} (82.5 - 85.3)	83.8 ± 0.4 ^{c} (82.7 - 84.9)	83.9 ± 0.5 ^{c} (82.6 – 85.0)	84.0 ± 0.7 ^{c} (82.9 - 85.8)	84.0 ± 0.7 ^{c} (82.5 - 85.6)	83.5 ± 0.3 ^{b} (83.1 - 83.9)	83.0 ± 0.0 ^{a} (83.0 - 83.1)
Cross 2	84.3	83.9 ± 0.7 ^{c} (81.9 – 86.0)	83.2 ± 0.5 ^{a} (82.0 - 84.4)	83.9 ± 0.7 ^{c} (82.5 - 85.5)	83.5 ± 0.7 ^{b} (81.7 - 85.7)	83.9 ± 0.7 ^{c} (81.9 - 85.7)	83.2 ± 0.0 ^{a} (83.1 - 83.3)	83.3 ± 0.1 ^{a} (83.1 - 83.5)
Cross 3	84.5	84.4 ± 0.7 ^{e} (82.1 - 86.9)	83.5 ± 0.6 ^{a} (82.1 - 85.1)	84.1 ± 0.7 ^{bcd} (82.3 - 85.8)	83.9 ± 0.8 ^{b} (82.0 - 86.1)	84.0 ± 0.6 ^{c} (82.5 - 85.7)	84.4 ± 0.2 ^{e} (84.1 - 84.7)	84.2 ± 0.2 ^{d} (83.9 - 84.4)
Cross 4	84.2	84.1 ± 0.6 ^{d} (81.8 - 85.9)	83.5 ± 0.5 ^{b} (81.8 - 84.9)	84.0 ± 0.7 ^{cd} (82.9 - 85.8)	83.7 ± 0.6 ^{b} (82.2 - 85.5)	83.8 ± 0.7 ^{c} (82.4 - 85.6)	82.4 ± 0.1 ^{a} (82.2 - 82.6)	84.1 ± 0.1 ^{d} (84.0 - 84.2)
Cross 5	83.1	83.2 ± 0.5 ^{c} (81.6 - 84.6)	83.0 ± 0.4 ^{b} (82.0 - 84.1)	83.1 ± 0.5 ^{bcd} (82.1 - 84.6)	83.1 ± 0.6 ^{b} (81.7 - 84.7)	83.2 ± 0.6 ^{d} (81.8 - 84.6)	82.4 ± 0.1 ^{a} (82.2 - 82.6)	84.1 ± 0.1 ^{e} (83.9 - 84.3)
Cross 6	83.4	83.4 ± 0.6 ^{d} (81.7 - 85.3)	83.1 ± 0.5 ^{bc} (81.8 - 84.5)	83.3 ± 0.6 ^{cd} (81.8 - 84.6)	83.1 ± 0.7 ^{b} (81.5 - 84.6)	83.4 ± 0.7 ^{d} (81.6 - 85.3)	82.9 ± 0.1 ^{a} (82.7 – 83.0)	83.7 ± 0.2 ^{e} (83.3 – 84.0)
Cross 7	83.0	83.3 ± 0.5 ^{c} (81.7 - 84.8)	83.2 ± 0.5 ^{bc} (82.0 - 84.4)	83.3 ± 0.5 ^{c} (82.0 - 84.4)	83.1 ± 0.6 ^{b} (81.2 - 85.1)	83.0 ± 0.7 ^{a} (81.3 - 84.8)	83.5 ± 0.1 ^{d} (83.4 - 83.8)	83.3 ± 0.1 ^{c} (83.2 - 83.5)
Cross 8	83.6	83.7 ± 0.2 ^{d} (83.1 - 84.4)	83.6 ± 0.2 ^{b} (83.1 - 84.1)	83.7 ± 0.2 ^{cd} (83.1 - 84.3)	83.5 ± 0.3 ^{b} (82.9 - 84.1)	83.7 ± 0.3 ^{c} (82.8 - 84.4)	83.4 ± 0.0 ^{a} (83.3 - 83.4)	83.7 ± 0.0 ^{d} (83.7 - 83.8)
Cross 9	83.2	83.0 ± 0.2 ^{b} (82.2 - 83.8)	82.9 ± 0.2 ^{a} (82.3 - 83.5)	83.0 ± 0.2 ^{b} (82.5 - 83.6)	83.2 ± 0.3 ^{c} (82.4 – 84.0)	83.3 ± 0.3 ^{d} (82.6 - 84.3)	83.0 ± 0.0 ^{b} (83.0 – 83.0)	83.6 ± 0.0 ^{e} (83.6 - 83.6)
Cross 10	83.5	83.7 ± 0.3 ^{c} (82.8 - 84.6)	83.4 ± 0.2 ^{a} (82.8 - 83.9)	83.7 ± 0.4 ^{c} (83.0 - 84.6)	83.4 ± 0.4 ^{a} (82.6 - 84.7)	83.5 ± 0.5 ^{b} (82.4 - 84.6)	83.4 ± 0.0 ^{a} (83.3 - 83.4)	83.8 ± 0.0 ^{c} (83.7 - 83.9)

平均值±標準差，括弧內為最小值及最大值，同雜交組合不同字母具顯著差異 ($p < 0.05/21$)。

Mean ± standard deviation; minimum and maximum are provided between the brackets; different letters signify significant differences ($p < 0.05/21$).

doi:10.6342/NTU201701806

表十二、GS 及 PS 之株高 GEBV 結果

Table 12. The plant height GEBV for GS and PS

PH	mid-parent	F_2	F_2		F_6		F_6 best line	
			GS	PS	GS	PS	GS	PS
Cross 1	105.2	105.2 ± 1.4 ^{c} (101.2 - 109.7)	104.1 ± 1.1 ^{ab} (101.2 - 107.3)	105.0 ± 1.3 ^{c} (102.2 - 108.6)	103.8 ± 1.6 ^{a} (99.9 - 107.5)	104.5 ± 2.0 ^{b} (100.7 - 110.5)	104.3 ± 0.9 ^{ab} (102.9 - 105.5)	105.5 ± 0.1 ^{d} (105.2 - 105.7)
Cross 2	102.7	102.7 ± 1.4 ^{b} (98.7 - 107.8)	102.4 ± 1.5 ^{ab} (98.8 - 105.8)	102.6 ± 1.5 ^{ab} (98.9 - 107.1)	102.6 ± 2.0 ^{b} (98.0 - 107.1)	102.4 ± 2.1 ^{a} (98.0 - 108.4)	102.4 ± 0.5 ^{ab} (101.5 - 103.3)	103.8 ± 0.4 ^{c} (103.2 - 104.5)
Cross 3	104.2	104.2 ± 1.5 ^{a} (99.7 - 109.5)	103.9 ± 1.4 ^{a} (100.1 - 107.6)	104.1 ± 1.5 ^{a} (99.9 - 109.5)	104.2 ± 2.0 ^{a} (99.2 – 108.0)	104.0 ± 2.1 ^{a} (98.2 - 109.2)	105.9 ± 0.6 ^{c} (104.6 – 107.0)	104.9 ± 0.2 ^{b} (104.5 - 105.2)
Cross 4	103.0	103.0 ± 1.6 ^{b} (98.4 - 107.9)	103.1 ± 1.5 ^{bc} (98.6 - 106.6)	103.0 ± 1.5 ^{abc} (99.8 - 106.2)	103.3 ± 1.7 ^{c} (98.9 - 108.3)	103.3 ± 2.1 ^{c} (98.6 - 109.8)	102.6 ± 0.2 ^{a} (102.2 – 103.0)	104.3 ± 0.4 ^{d} (103.7 - 105.2)
Cross 5	102.6	102.6 ± 1.4 ^{ab} (97.5 - 107.5)	102.9 ± 1.4 ^{b} (99.9 - 106.2)	102.9 ± 1.3 ^{ab} (99.7 - 106.3)	102.4 ± 1.8 ^{a} (96.6 - 108.4)	102.5 ± 1.7 ^{ab} (96.1 - 106.5)	103.5 ± 0.3 ^{c} (103.0 - 104.1)	104.4 ± 0.2 ^{d} (104.2 - 104.8)
Cross 6	104.1	104.1 ± 1.3 ^{a} (100.2 – 108.0)	104.8 ± 1.2 ^{bc} (101.2 - 107.2)	104.4 ± 1.4 ^{ab} (100.2 - 107.7)	104.9 ± 1.5 ^{c} (100.4 - 108.3)	104.8 ± 1.6 ^{bc} (100.3 - 108.9)	106.0 ± 0.4 ^{d} (105.2 - 106.7)	105.7 ± 0.5 ^{d} (104.7 - 106.8)
Cross 7	103.0	103.1 ± 1.4 ^{c} (98.8 - 107.7)	102.9 ± 1.4 ^{bc} (99.4 - 107.7)	102.8 ± 1.4 ^{bc} (98.8 - 107.2)	102.9 ± 1.8 ^{c} (98.8 - 107.8)	102.5 ± 1.8 ^{ab} (97.7 - 106.3)	102.4 ± 0.2 ^{a} (101.9 - 102.8)	104.9 ± 0.5 ^{d} (103.9 - 105.6)
Cross 8	101.6	101.6 ± 0.8 ^{e} (99.0 - 103.9)	101.0 ± 0.8 ^{bc} (99.0 - 103.2)	101.5 ± 0.8 ^{de} (100.1 - 103.2)	101.0 ± 1.2 ^{b} (98.0 - 103.7)	101.3 ± 1.1 ^{cd} (98.9 - 103.6)	102.1 ± 0.1 ^{f} (102.0 - 102.3)	100.8 ± 0.1 ^{a} (100.6 – 101.0)
Cross 9	100.4	100.5 ± 0.8 ^{a} (98.1 – 103.0)	100.8 ± 0.8 ^{bc} (98.8 – 103.0)	100.5 ± 0.8 ^{abc} (98.6 - 102.3)	100.6 ± 0.8 ^{c} (98.9 - 102.5)	100.6 ± 1.1 ^{abc} (97.7 - 103.7)	100.5 ± 0.0 ^{b} (100.5 - 100.6)	101.3 ± 0.1 ^{d} (101.2 - 101.4)
Cross 10	101.9	101.9 ± 0.6 ^{d} (100.2 - 104)	101.3 ± 0.6 ^{b} (100.2 - 102.7)	101.9 ± 0.7 ^{d} (100.4 – 104.0)	101.2 ± 0.8 ^{b} (99.5 - 103.1)	101.6 ± 0.9 ^{c} (99.4 - 103.6)	100.1 ± 0.1 ^{a} (99.9 - 100.3)	102.0 ± 0.1 ^{d} (101.7 - 102.3)

平均值±標準差，括弧內為最小值及最大值，同雜交組合不同字母具顯著差異 ($p < 0.05/21$)。

Mean ± standard deviation; minimum and maximum are provided between the brackets; different letters signify significant differences ($p < 0.05/21$).

doi:10.6342/NTU201701806

表十三、輪迴選種各循環 F_2 族群性狀 GEBV

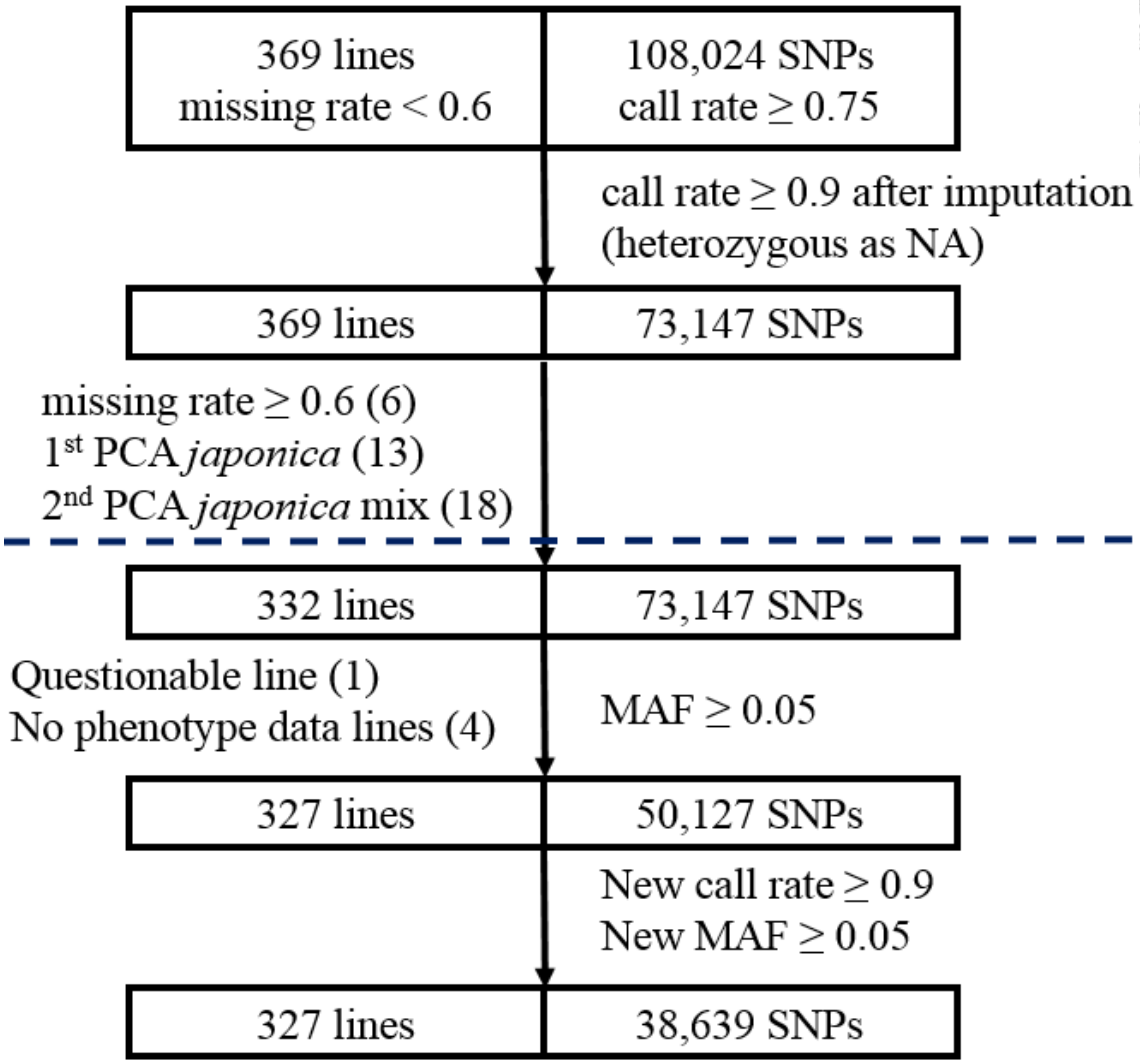
Table 13. GEBV for F_2 population at each cycle of recurrent selection

		G0	G1	G2	G3	G4	G5
Yield	GEBV	4770 ± 366 ^{f}	5410 ± 190 ^{e}	5698 ± 148 ^{d}	5765 ± 109 ^{c}	5842 ± 99 ^{b}	5948 ± 103 ^{a}
		(3480 - 5845)	(4880 - 5834)	(5241 - 6009)	(5466 - 6031)	(5513 - 6088)	(5617 - 6195)
	BLUE	4770 ± 366 ^{f}	5268 ± 160 ^{e}	5537 ± 114 ^{d}	5697 ± 106 ^{c}	5790 ± 98 ^{b}	5862 ± 90 ^{a}
		(3480 - 5845)	(4792 - 5699)	(5146 - 5862)	(5417 - 6057)	(5548 - 6053)	(5607 - 6186)
FL	GEBV	88.7 ± 3.2 ^{f}	83.3 ± 0.7 ^{e}	82.2 ± 0.5 ^{d}	81.9 ± 0.5 ^{c}	81.2 ± 0.4 ^{b}	81.0 ± 0.5 ^{a}
		(82.0 – 96.0)	(81.6 - 85.2)	(80.5 - 83.6)	(80.4 – 83.0)	(80.0 - 82.2)	(79.7 - 82.1)
	BLUE	88.7 ± 3.2 ^{f}	84.2 ± 1.4 ^{e}	82.9 ± 1.2 ^{d}	81.7 ± 0.6 ^{c}	81.1 ± 0.5 ^{b}	80.5 ± 0.5 ^{a}
		(82.0 - 96)	(81.4 - 89.3)	(80.2 - 87.9)	(79.9 - 83.4)	(79.9 - 82.9)	(79.3 - 82.1)
PH	GEBV	109.9 ± 5.0 ^{c}	103.4 ± 3.5 ^{ab}	103.3 ± 1.8 ^{ab}	103.5 ± 1.2 ^{ab}	103.5 ± 1.1 ^{b}	103.3 ± 1.1 ^{a}
		(95.0 - 127.1)	(94.5 - 110.3)	(99.5 - 109.3)	(100.6 - 106.9)	(99.9 - 107.3)	(100.4 - 106)
	BLUE	109.9 ± 5.0 ^{f}	103.8 ± 3.2 ^{e}	99.3 ± 2.0 ^{d}	99.1 ± 1.7 ^{c}	98.4 ± 1.6 ^{b}	98.0 ± 1.6 ^{a}
		(95.0 - 127.1)	(94.4 - 110.6)	(93.9 - 104.9)	(94.0 - 104.2)	(93.9 - 102.8)	(94.1 - 102.2)

平均值±標準差，括弧內為最小值及最大值，不同字母具顯著差異 ($p < 0.05/15$)。

Mean ± standard deviation; minimum and maximum are provided between the brackets; different letters signify significant differences ($p < 0.05/21$).

doi:10.6342/NTU201701806



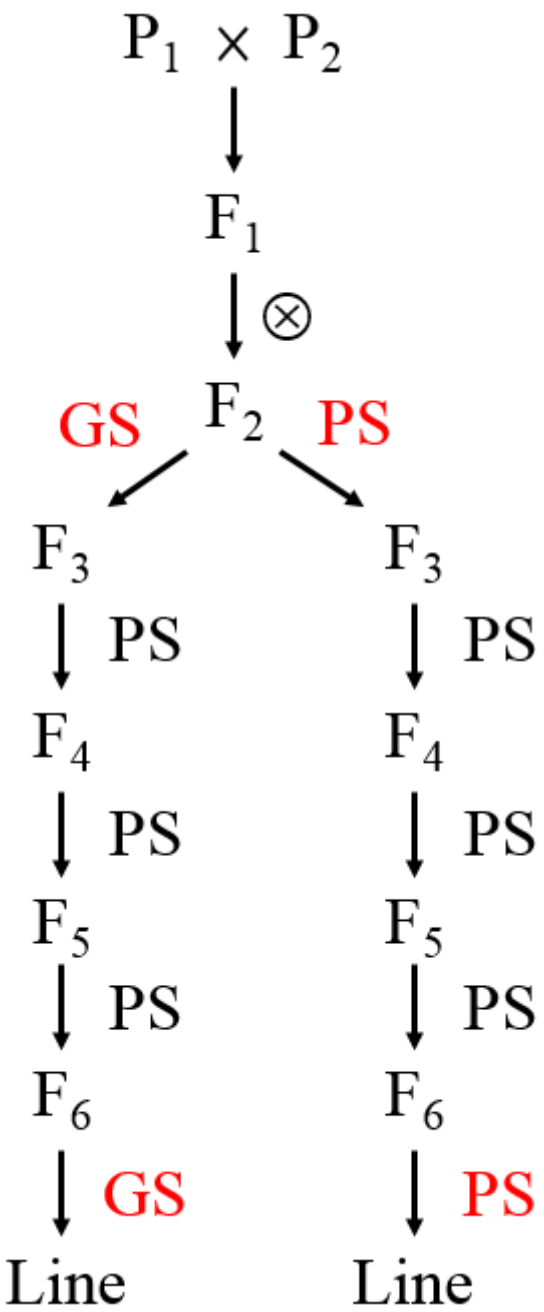
圖一、外表型資料與基因型資料篩選

虛線上方為 Spindel et al. (2015) 篩選流程；下方為本研究再次篩選條件。

Figure 1. The filtering of the phenotypic data and the genotypic data

Above the dash line is the filtering process of Spindel et al. (2015); below are the additional filtering steps conducted in this study.

doi:10.6342/NTU201701806



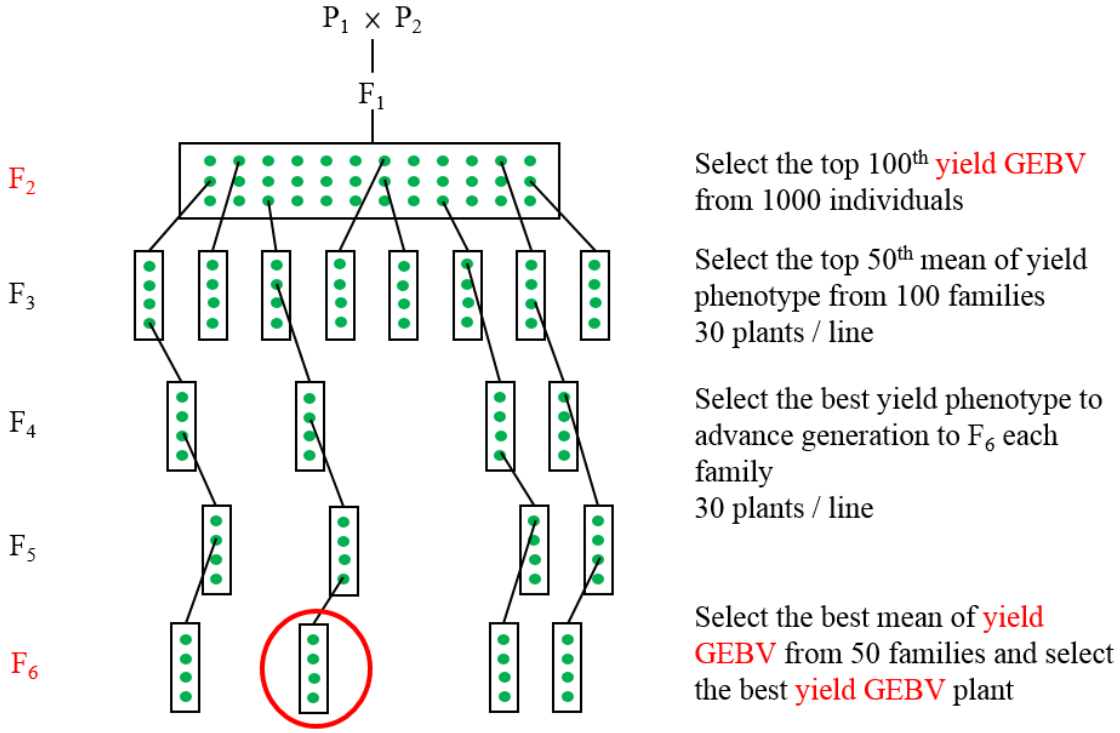
圖二、雙親本雜交流程

GS 為基因體選種，PS 為外表型選種，於 F_2 及 F_6 世代以不同方法選拔產量性狀而分成兩種路徑，其餘世代皆以 PS 進行選拔，藉此比較 GS 與 PS 兩種選拔方法的遺傳增進效果。

Figure 2 The bi-parental cross scheme

GS, genomic selection; PS, phenotypic selection. Two paths are divided from F_2 generation in order to compare the genetic improvement between GS and PS.

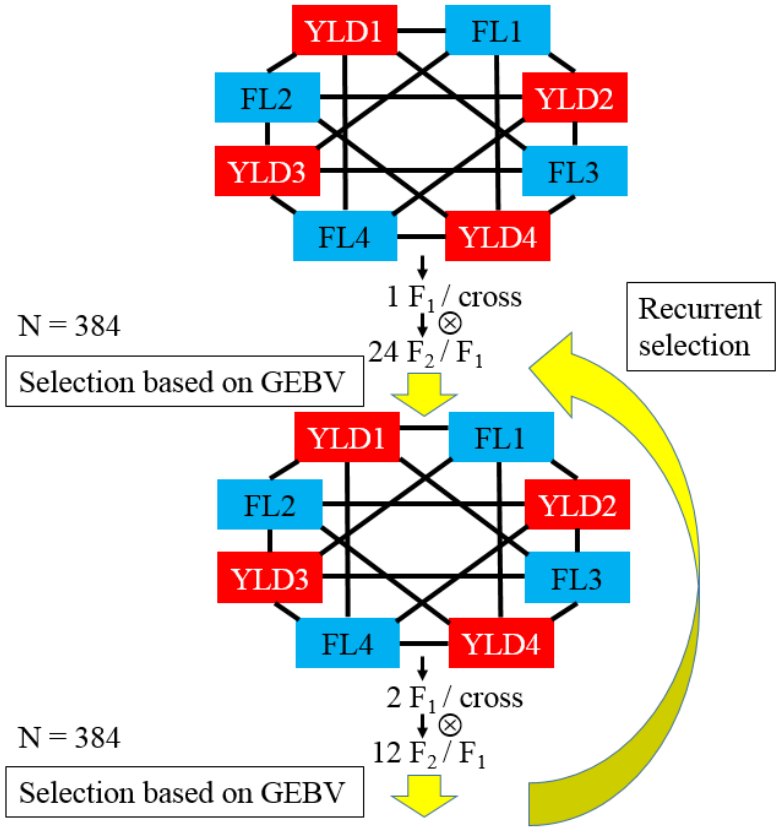
doi:10.6342/NTU201701806



圖三、雙親本雜交流程之 GS 路徑示意圖，以譜系法推進世代

Figure 3. The bi-parental cross scheme for GS by modified pedigree method

doi:10.6342/NTU201701806



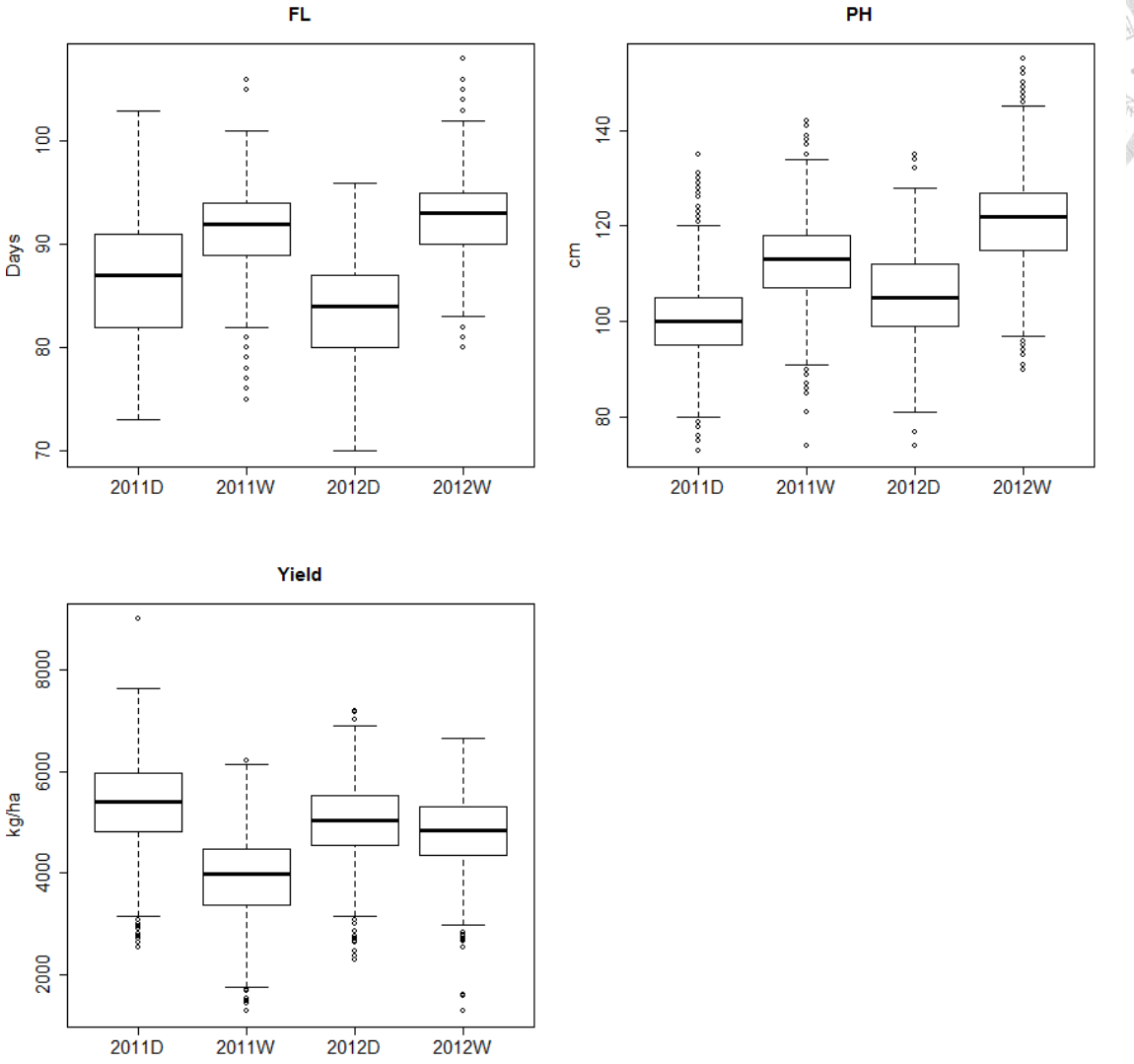
圖四、輪迴選種流程

首先從 327 品系中選出產量前四名與抽穗期前四名之品系相互雜交，共 16 個雜交組合，每個組合產生 1 株 F_1 再自交產生 24 株 F_2，共 384 株 F_2 組成 Cycle 1 族群。接著同樣從 Cycle 1 族群中選出產量前四名與抽穗期前四名之個體相互雜交，每個組合產生 2 株 F_1 再自交產生 12 株 F_2，維持相同族群大小，共 384 株 F_2 組成 Cycle 2 族群，重複此步驟至產生 Cycle 5 族群為止。

Figure 4. The recurrent selection scheme

From the initial 327 lines, we selected the top four lines in yield and the top four lines in flowering time, and then cross them to produce 16 crossing combinations. Each combination produces one F_1 and then self to produce 24 F_2 to form a total of 384 F_2 in Cycle 1. From Cycle 1, we selected the top four individuals in yield and the top four individuals in flowering time, and then cross them to produce 16 crossing combinations. Each combination produces two F_1 and then self to produce 12 F_2. A total of 384 F_2 forms Cycle 2. The process was iterate till cycle 5.

doi:10.6342/NTU201701806



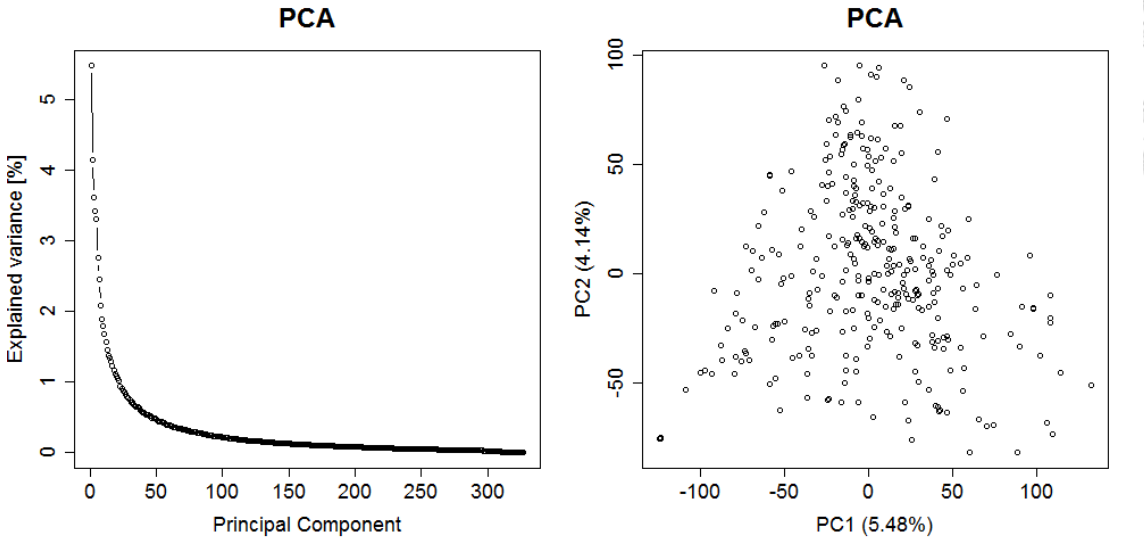
圖五、327 品系各性狀於四個環境下之表現

FL 為抽穗期、PH 為株高、Yield 為產量。

Figure 5. The performance of the 327 lines in four trials

FL, flowering time; PH, plant height.

doi:10.6342/NTU201701806



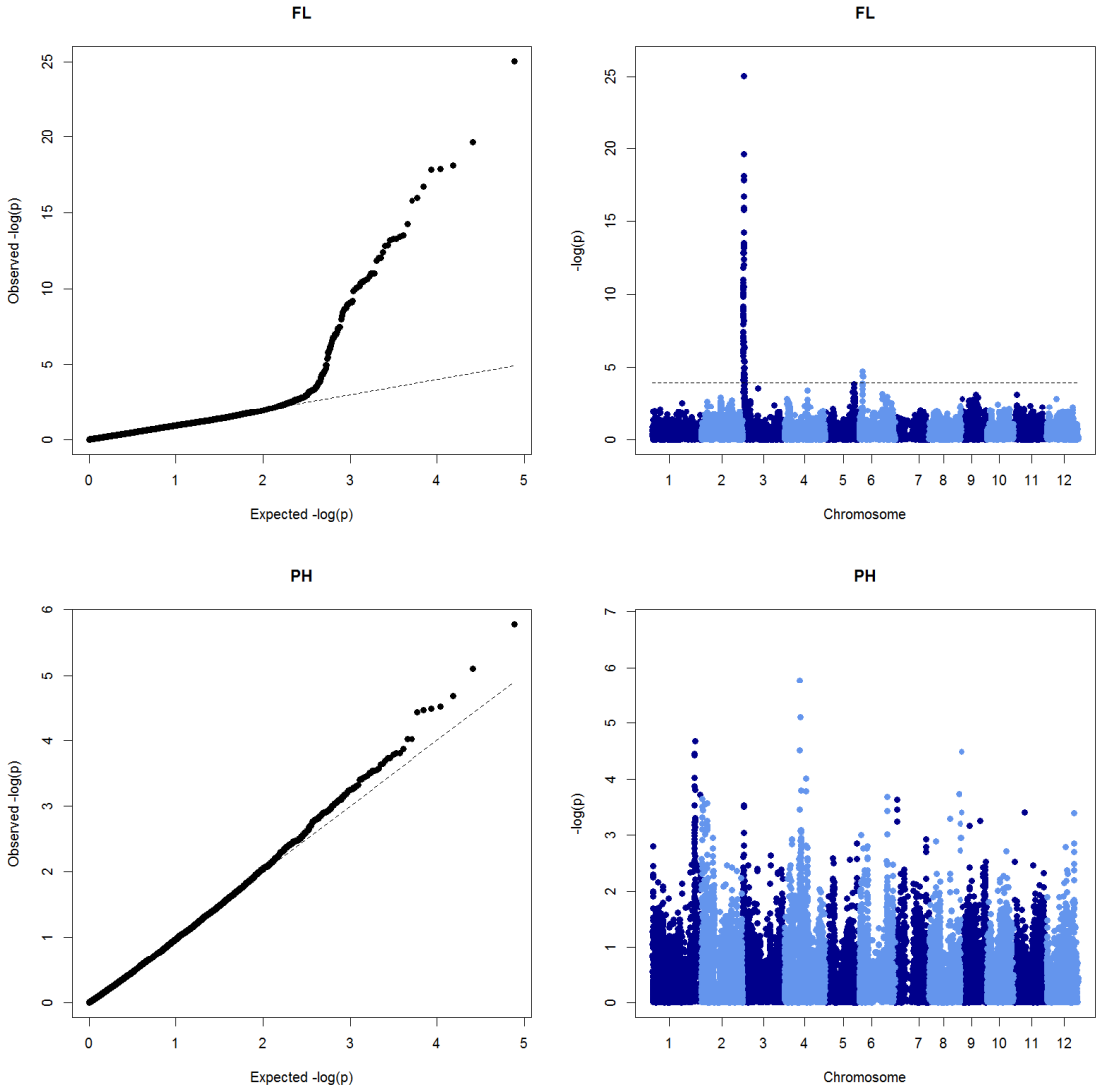
圖六、以 PCA 針對 327 品系結合 38,639 個 SNP 所得之族群結構分析圖

左為各成分解釋變異比例，右為第一、二主成分所畫散佈圖。

Figure 6. Applying PCA to 327 lines × 38,639 SNPs for a population structure summary analysis

The left part is the variance proportion explained by each component, and the right part is the distribution of the 327 lines based on the first and the second components.

doi:10.6342/NTU201701806



圖七、327 品系和 38,639 個 SNP 之 GWAS 結果

QQ plot 在左、Manhattan plot 在右；

虛線為 FDR = 0.05，無虛線表示無顯著 SNP。FL 為抽穗期、PH 為株高、Yield 為產量。

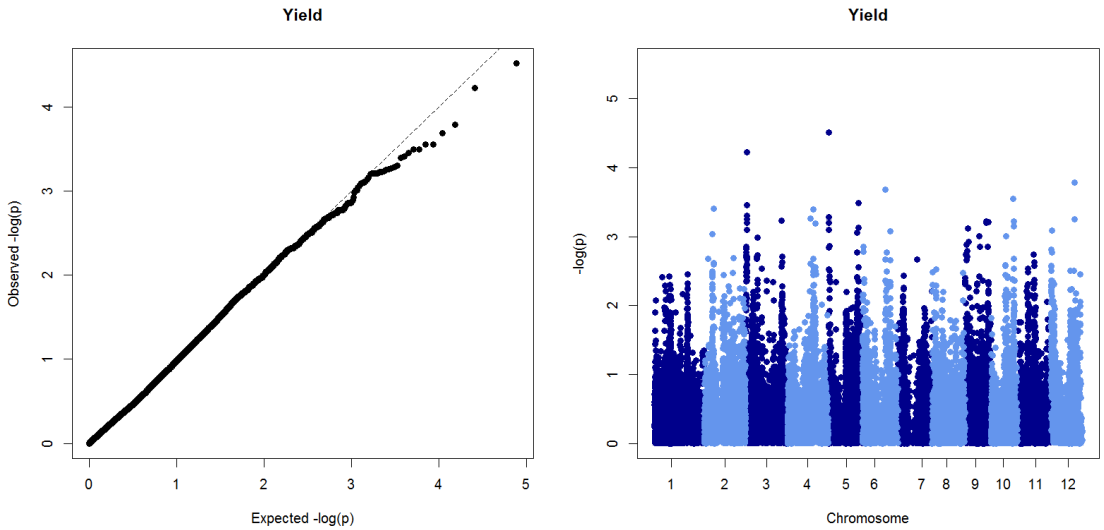
Figure 7. GWAS results for 327 lines × 38,639 SNPs

QQ plot is at the left-hand side and Manhattan plot at the right-hand side.

The dotted line is signifies FDR = 0.05, and no dotted line indicates no significant SNP.

FL, flowering time; PH, plant height.

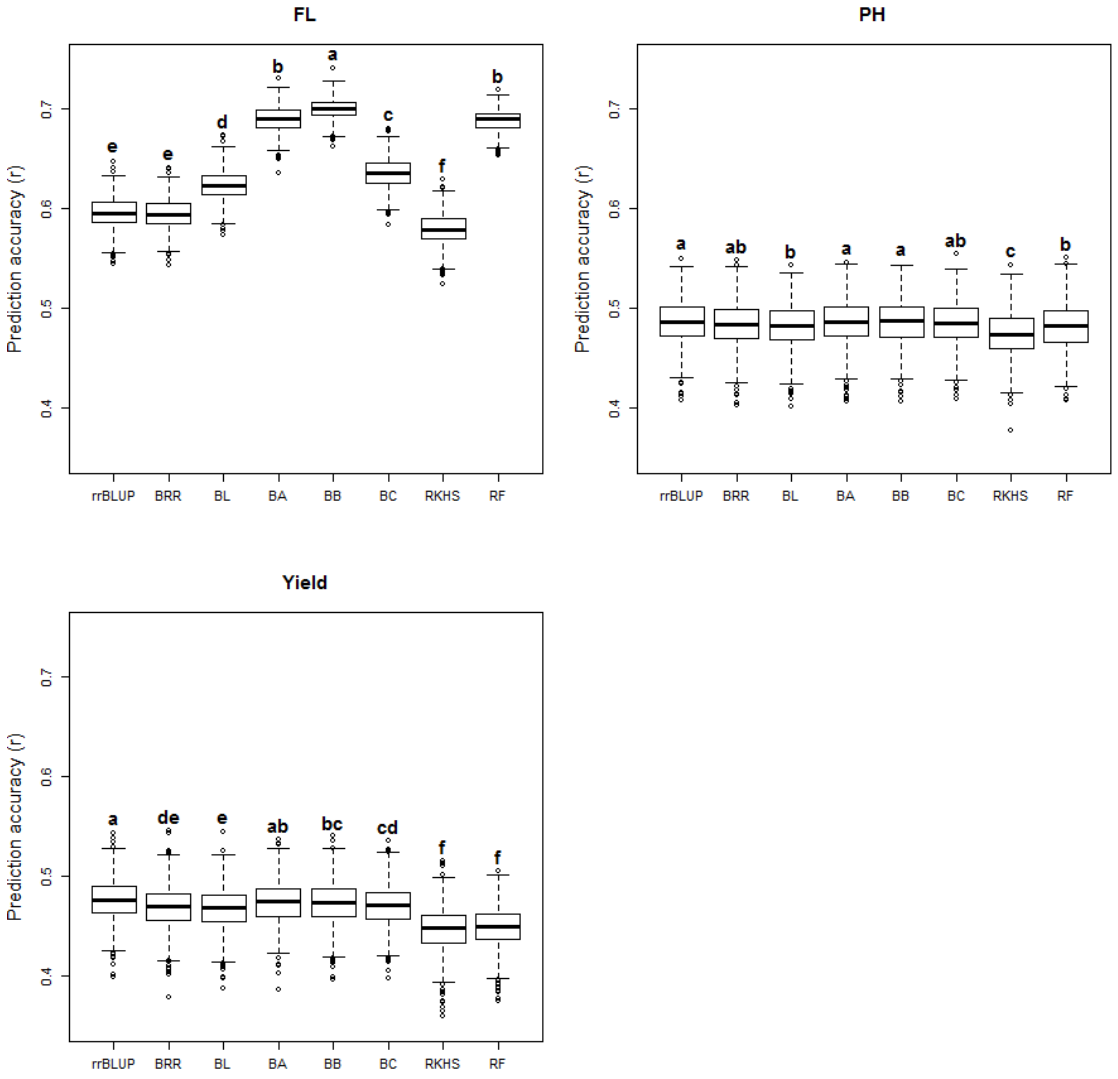
doi:10.6342/NTU201701806



圖七、327 品系和 38,639 個 SNP 之 GWAS 結果 (續)

Figure 7. GWAS results for 327 lines × 38,639 SNPs (continued)

doi:10.6342/NTU201701806

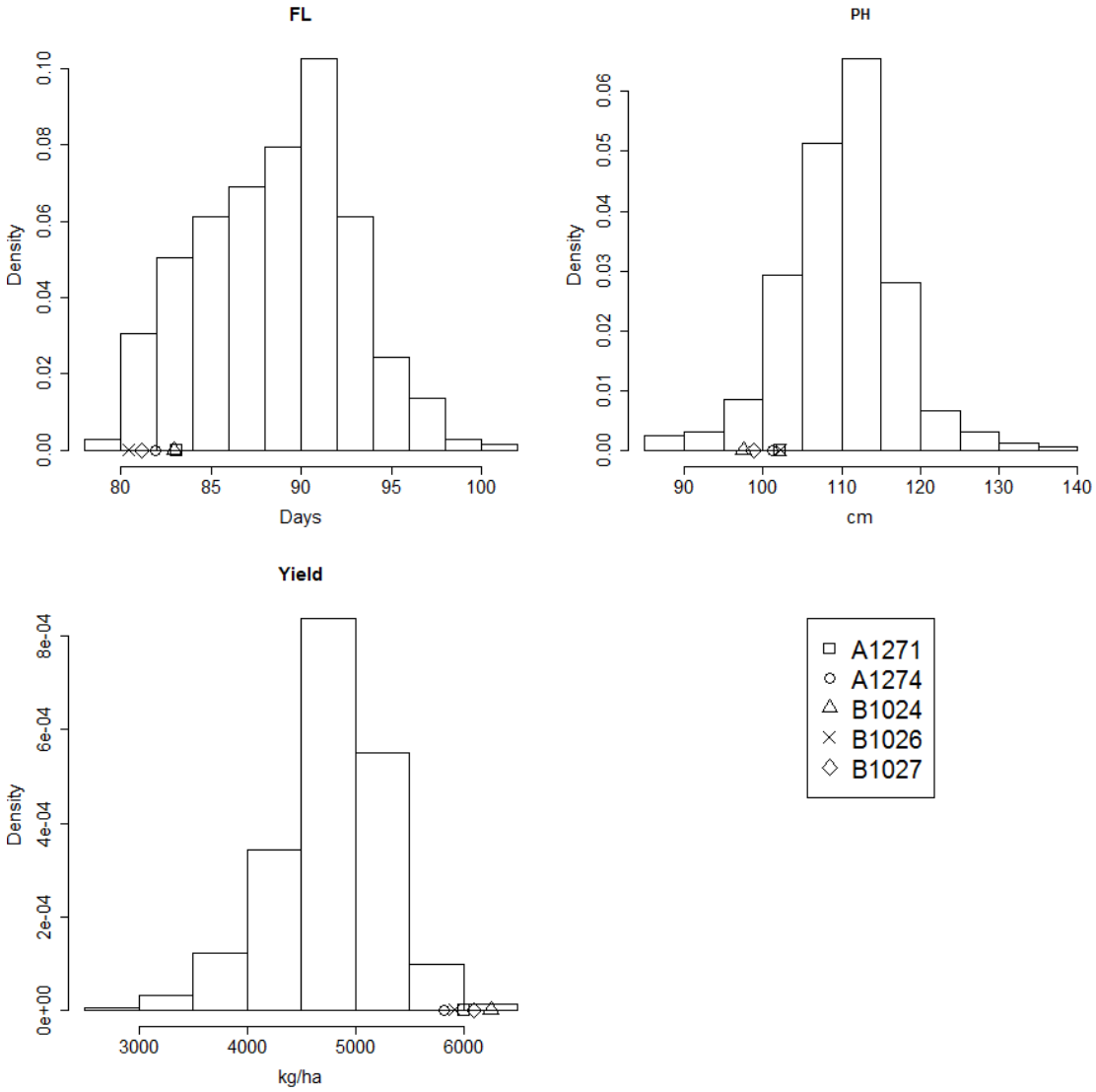


圖八、八種統計方法在各性狀的預測準確度比較圖

每個方法以十折交叉驗證重複 1000 次，再以 Tukey's test 多重比較檢定各統計方法間是否具有顯著差異。FL 為抽穗期、PH 為株高、Yield 為產量。

Figure 8. The prediction accuracy comparison of eight statistical methods for each trait Each method was 1000 times cross-validated (10-fold cross-validation), and then Tukey's test was used to determine whether there was a significant difference between the statistical methods. FL, flowering time; PH, plant height.

doi:10.6342/NTU201701806



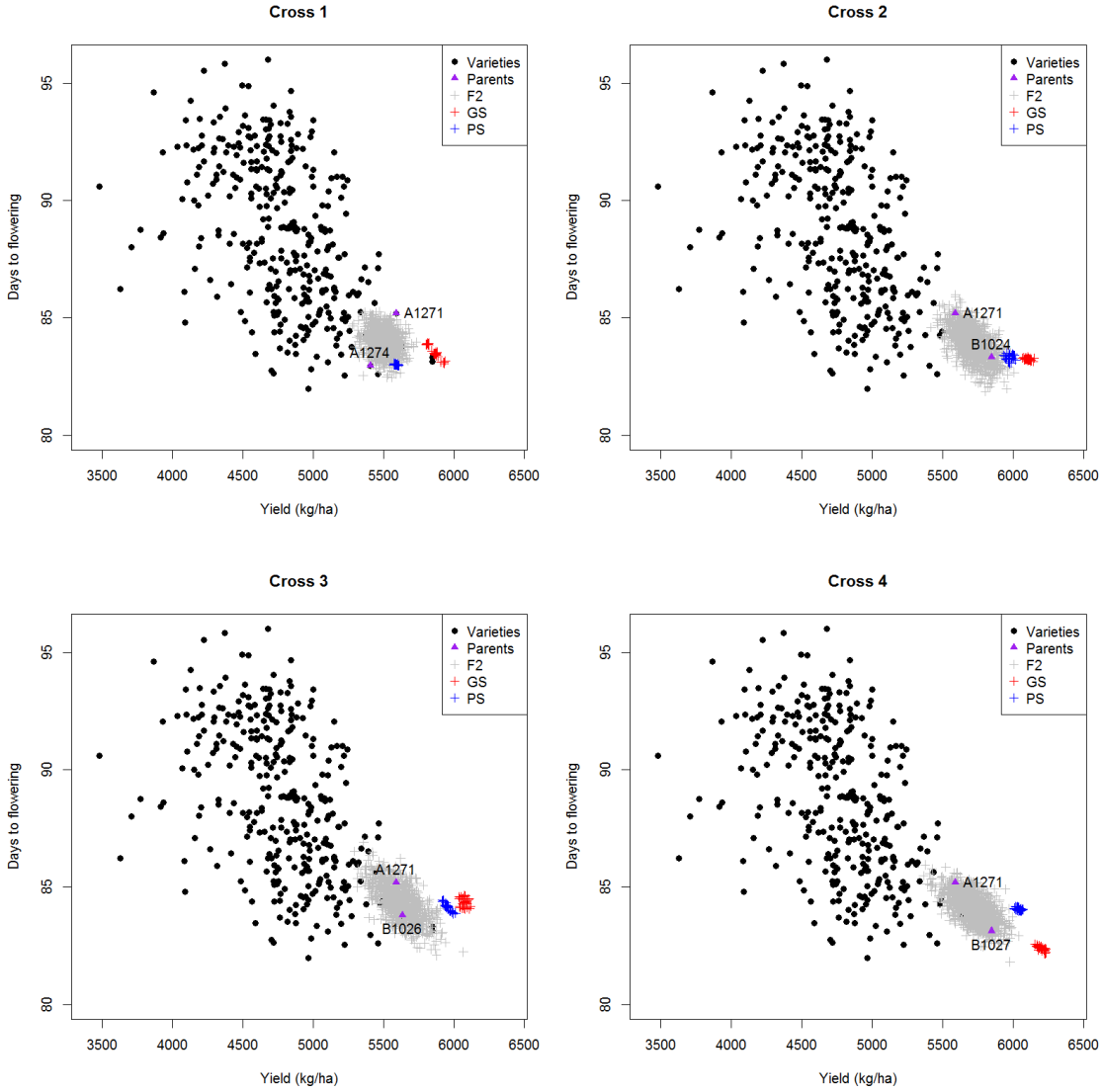
圖九、327 品系校正外表型分布圖

五個親本之表現亦標於圖上；FL 為抽穗期、PH 為株高、Yield 為產量。

Figure 9. The distribution of the adjusted phenotypes of 327 lines

The performances of the five selected parents are marked on the figure. FL, flowering time; PH, plant height.

doi:10.6342/NTU201701806



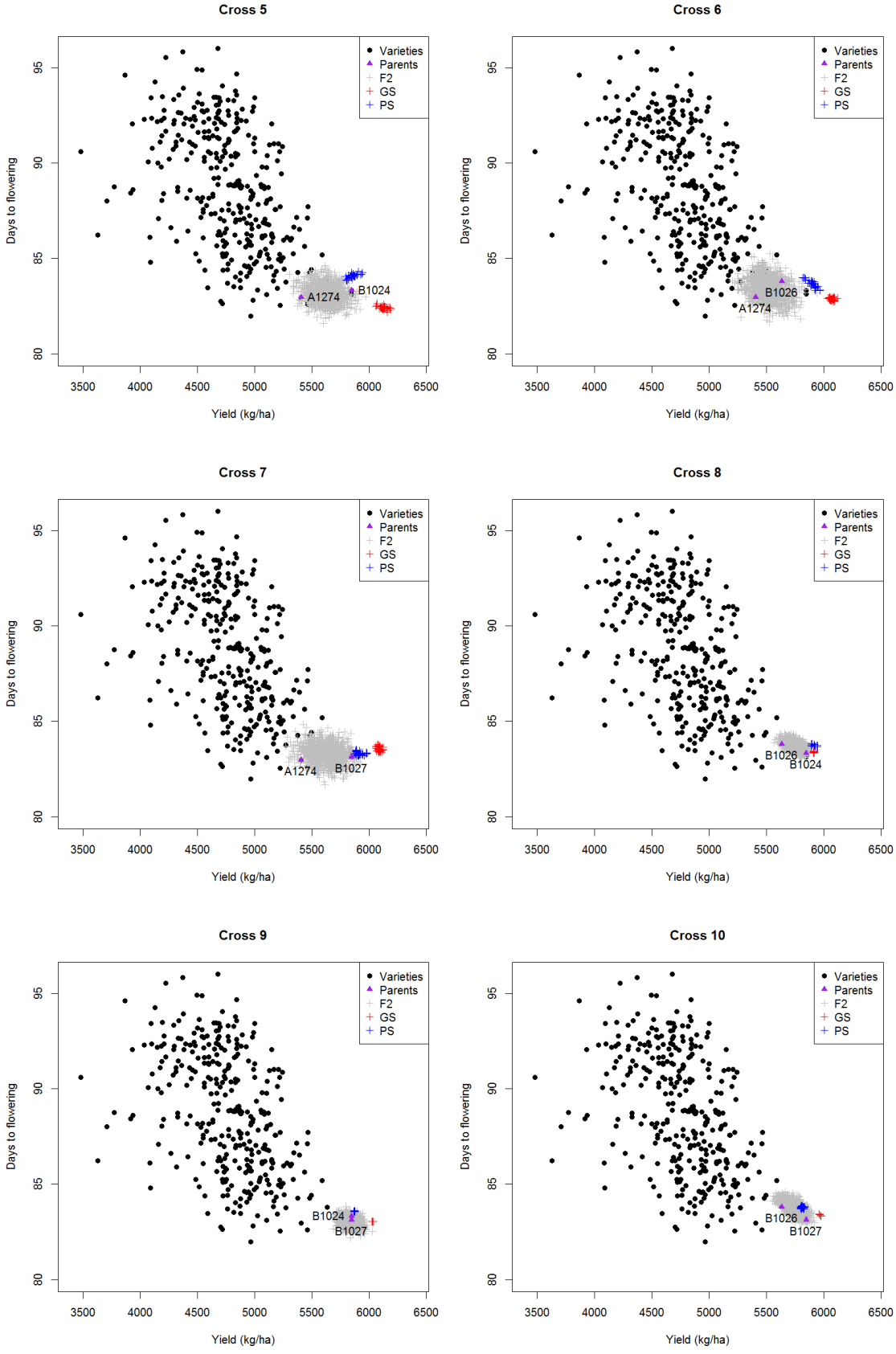
圖十、10 個雜交組合之產量 GEBV 與抽穗期 GEBV 散佈圖

黑點為 327 品系，紫三角為親本，灰十字為 F_2 族群，紅十字為 GS 選出 F_6 優良品系，藍十字為 PS 選出 F_6 優良品系。

Figure 10. Yield and days to flowering GEBV in 10 bi-parental crosses

Black spot represent the 327 initial lines, purple triangles for the parents, gray crosses for the F_2 group, red crosses for F_6 best line selected from GS, and blue crosses for F_6 best line selected from PS.

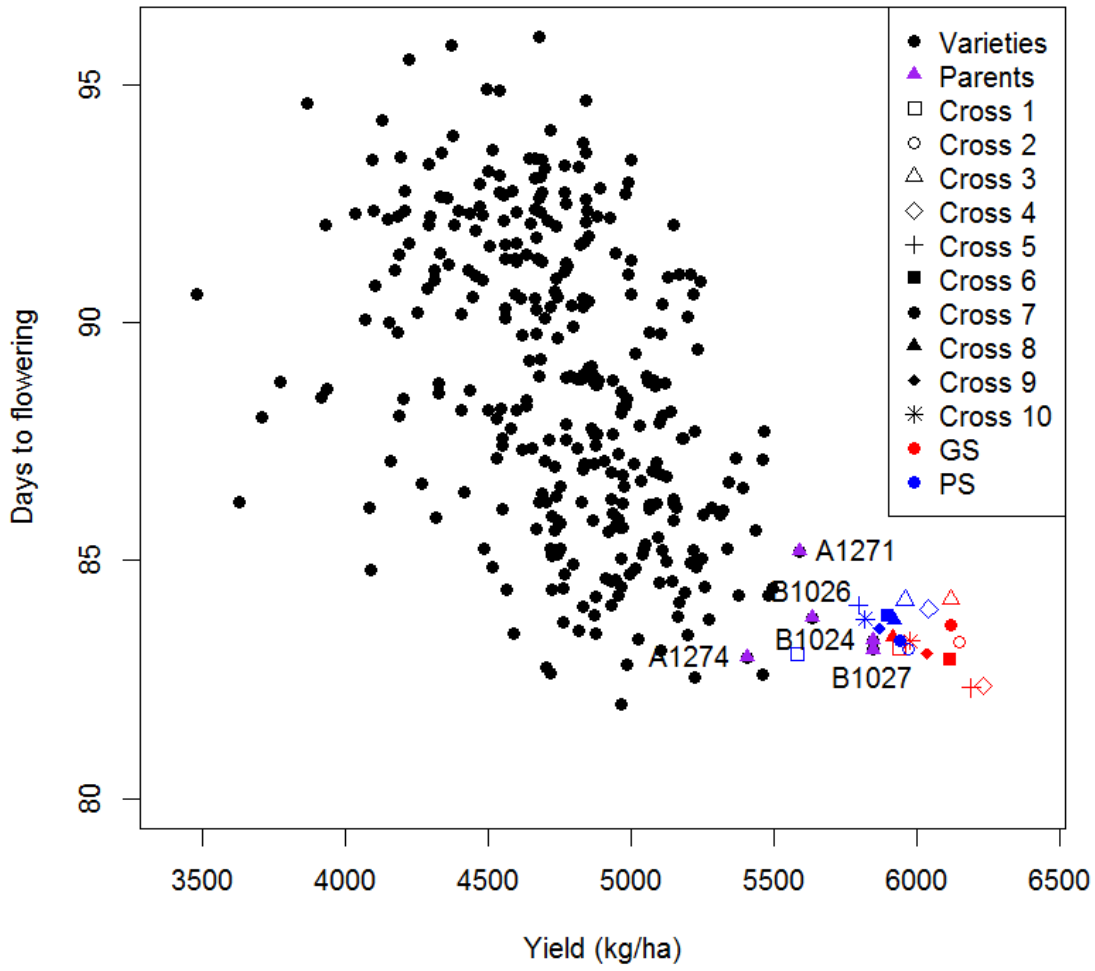
doi:10.6342/NTU201701806



圖十、10 個雜交組合之產量 GEBV 與抽穗期 GEBV 散佈圖 (續)

Figure 10. Yield and days to flowering GEBV in 10 bi-parental crosses (continued)

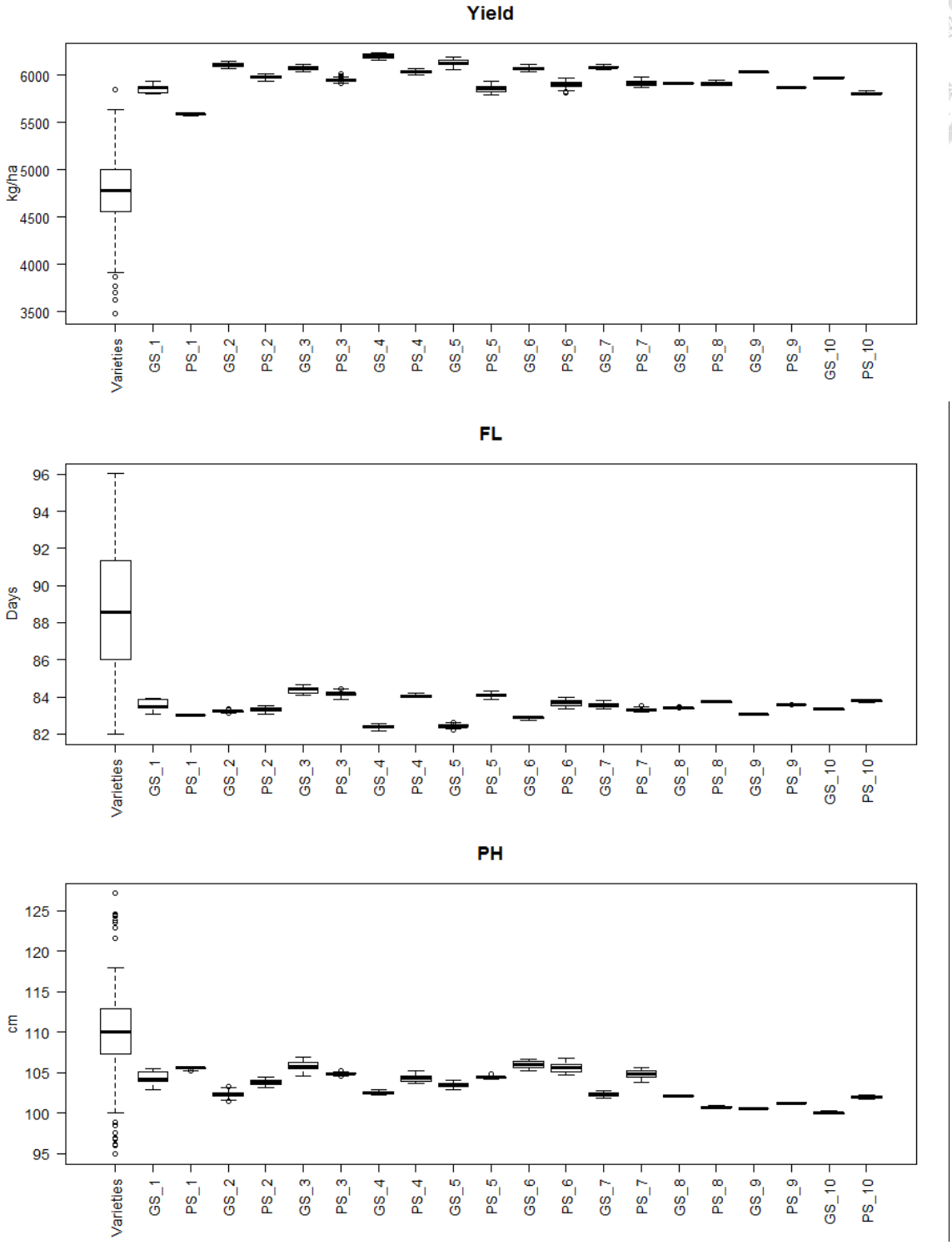
doi:10.6342/NTU201701806



圖十一、10 個雜交組合選出 F_6 優良品系之表現最佳個體之產量 GEBV 與抽穗期 GEBV 散佈圖

Figure 11. Yield and flowering time GEBV of the best individual of the F_6 best line selected from the ten crossing combinations

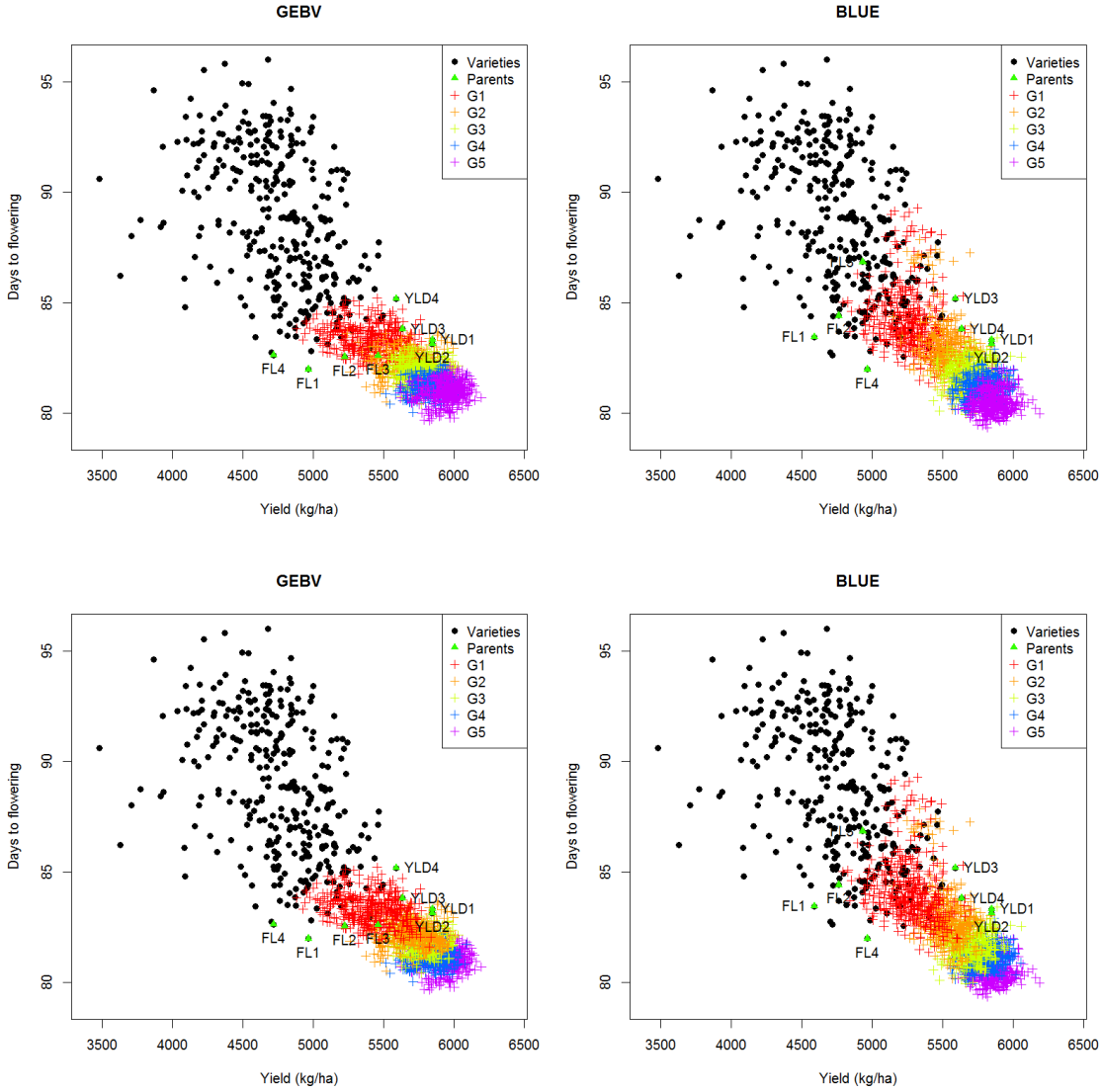
doi:10.6342/NTU201701806



圖十二、327 品系與 10 個雜交組合 GS 與 PS 選出 F_6 優良品系性狀之 GEBV 分布

Figure 12. Distribution for the 327 lines and the F_6 best lines selected from 10 crossing combinations

doi:10.6342/NTU201701806



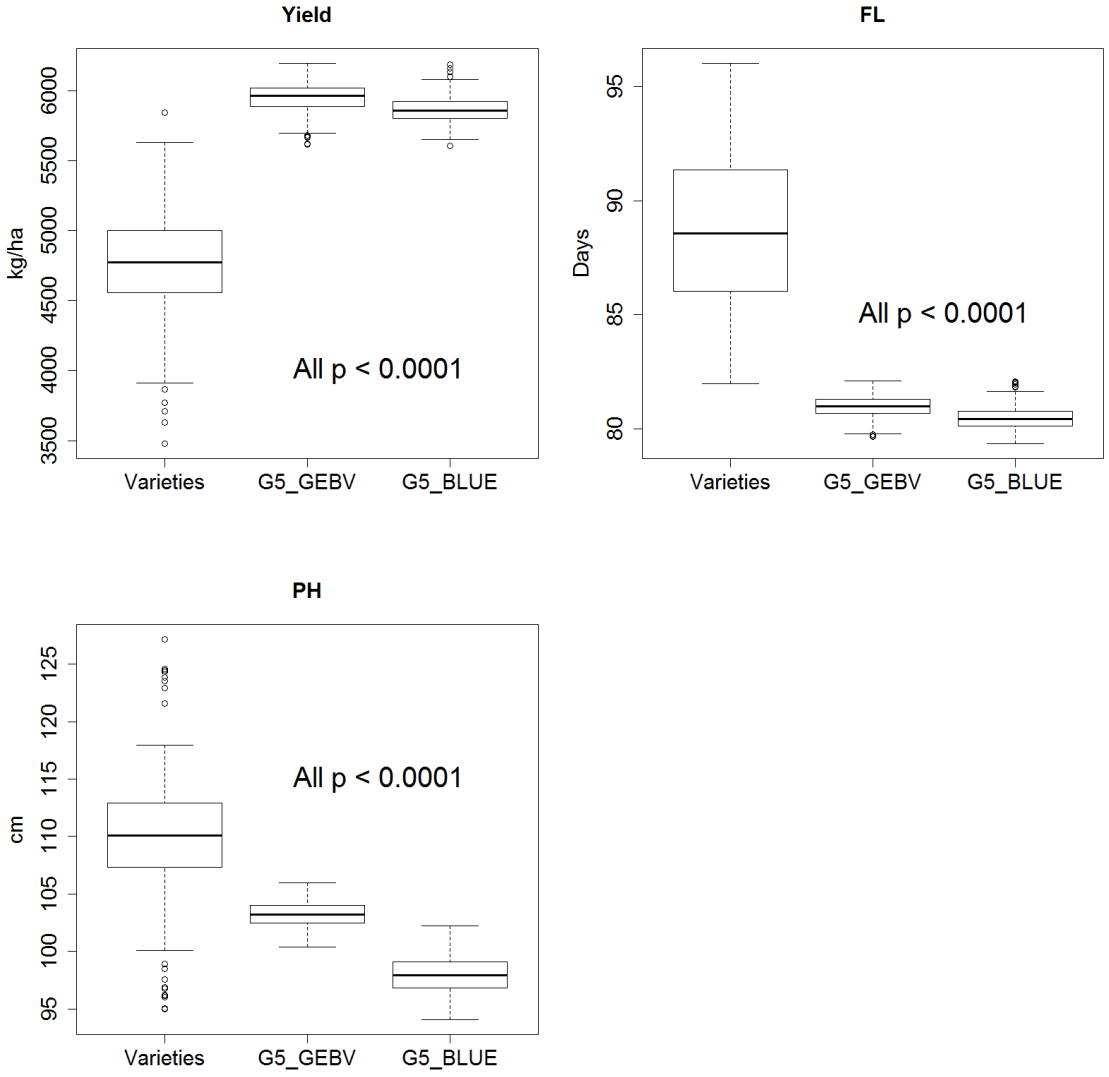
圖十三、輪迴選種五次循環 F_2 族群之產量 GEBV 與抽穗期 GEBV 散佈圖

左方為 GEBV 選出初始親本，右方為校正外表型 BLUE 選出初始親本。黑點為 327 品系，綠三角為初始親本，紅至紫十字依序為第一至第五次循環之 F_2 世代。

Figure 13. Yield and flowering time GEBV of the F2 population at each cycle of the recurrent selection

Results from GEBV-selected parents are on the left panel, results from the BLUE-selected parents are on the right panel. Black spots represent the 327 initial lines, the green triangles for the initial parents, red to purple crosses in order for the first to the fifth cycle of F_2 generation.

doi:10.6342/NTU201701806



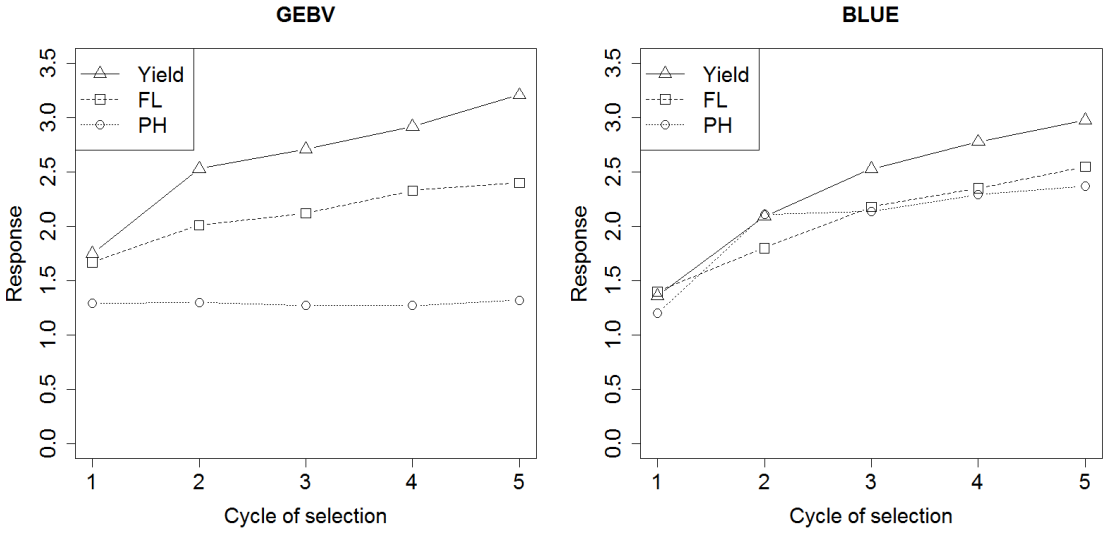
圖十四、輪迴選種第五次循環 F_2 族群性狀 GEBV 盒鬚圖

由左至右依序為 327 品系、GEBV 選拔初始親本之 Cycle 5 數值、以 BLUE 選拔初始親本之 Cycle 5 數值。經 Welch's *t*-test 兩兩比較，皆具顯著差異 ($p < 0.0001$)。Yield 為產量、FL 為抽穗期、PH 為株高。

Figure 14. Boxplot of F_2 populations at Cycle 5 of the recurrent selection

From left to right: the initial 327 lines,Cycle 5 of GEBV-selected initial parents, and Cycle 5 of BLUE-selected initial parents. Welch's *t*-tests show significant differences between any two groups ($p < 0.0001$). FL, flowering time; PH, plant height.

doi:10.6342/NTU201701806



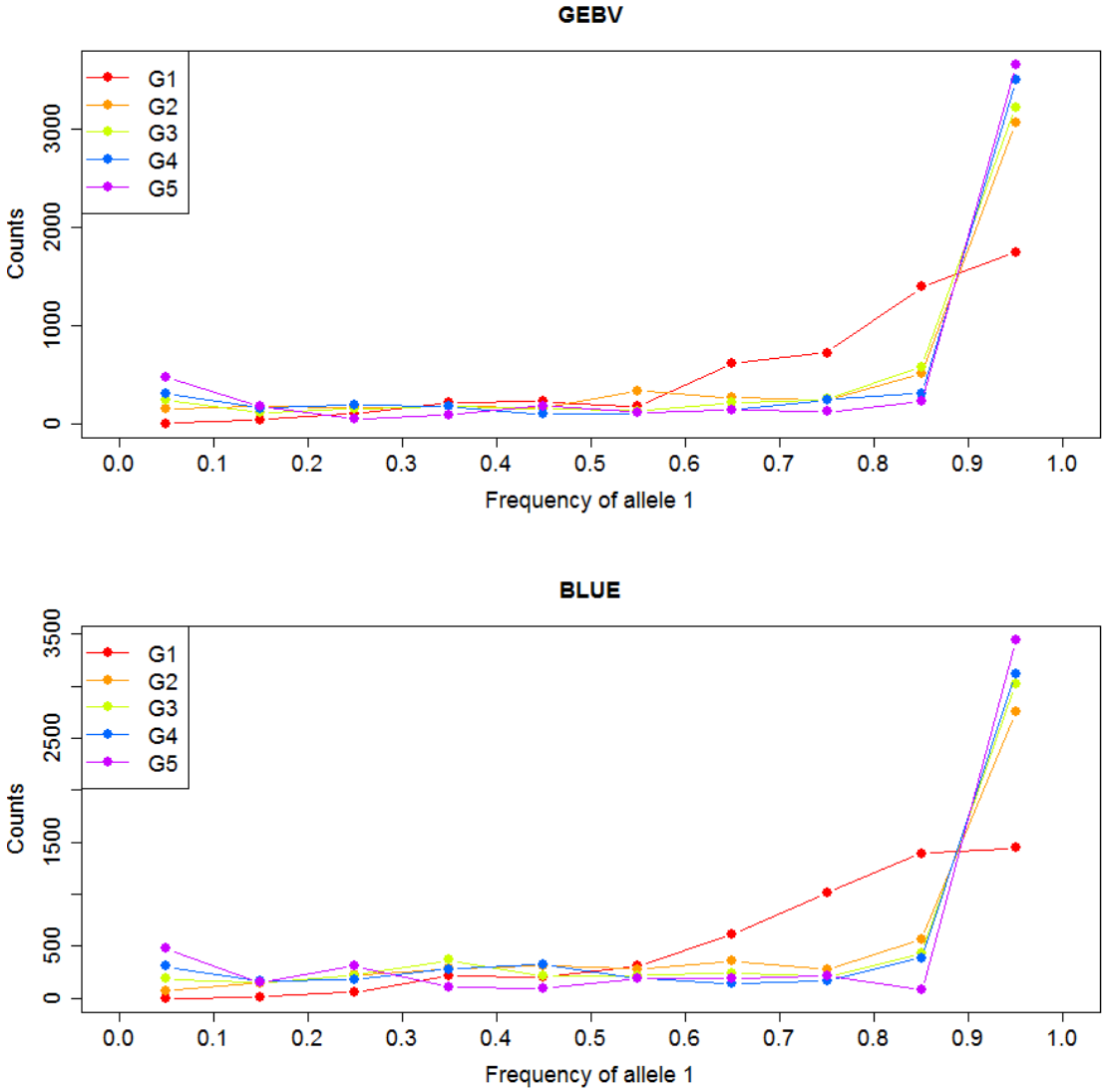
圖十五、輪迴選種各循環 F_2 族群相對 327 品系之標準化遺傳增進變化

抽穗期與株高為取絕對值結果。Yield 為產量、FL 為抽穗期、PH 為株高。

Figure 15. The standardized selection response for recurrent selection at each cycle's F_2 population relative to 327 lines

Absolute values are used for flowering time (FL) and plant height (PH).

doi:10.6342/NTU201701806



圖十六、5,264 個 SNP 之對偶基因 1 頻度於輪迴選種各循環 F_2 族群之變化

上圖顯示 GEBV 選拔初始親本之頻度變化，下圖顯示 BLUE 選拔初始親本之頻度變化，紅色至紫色依序為 Cycle 1 至 Cycle 5 之 F_2 世代。此圖由直方圖而來，x 軸為對偶基因 1 頻度，從 0 至 1 以組距 0.1 分為 10 組，y 軸為次數。

Figure 16. Frequency of allele 1 across the five cycles of recurrent selection

The upper figure shows the frequency evolution for cycles from GEBV-selected parents; the lower figure shows the frequency evolution for cycles from BLUE-selected parents. The red to purple curves show frequencies from Cycle 1 to Cycle 5. This diagram comes from the histogram, The x-axis is the frequency of allele 1, divided into 10 groups from 0 to 1; the y-axis is the number of counts at each frequency class.

doi:10.6342/NTU201701806

參考文獻

Ashraf B, Edriss V, Akdemir D, Autrique E, Bonnett D, Crossa J, Janss L, Singh R, Jannink J-L (2016) Genomic prediction using phenotypes from pedigreed lines with no marker data. Crop Sci 56:957-964

Asoro FG, Newell MA, Beavis WD, Scott MP, Jannink J-L (2011) Accuracy and training population design for genomic selection on quantitative traits in elite north american oats. Plant Genome 4:132-144

Auinger HJ, Schonleben M, Lehermeier C, Schmidt M, Korzun V, Geiger HH, Piepho HP, Gordillo A, Wilde P, Bauer E, Schon CC (2016) Model training across multiple breeding cycles significantly improves genomic prediction accuracy in rye (*Secale cereale* L.). Theor Appl Genet 129:2043-2053

Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67:1-48

Begum H, Spindel JE, Lalusin A, Borromeo T, Gregorio G, Hernandez J, Virk P, Collard B, McCouch SR (2015) Genome-wide association mapping for yield and other agronomic traits in an elite breeding population of tropical rice (*Oryza sativa*). PLoS One 10:e0119873

Bernardo R (2008) Molecular markers and selection for complex traits in plants: learning from the last 20 years. Crop Sci 48:1649-1664

Bernardo R (2010) Genomewide selection with minimal crossing in self-pollinated crops. Crop Sci 50:624-627

Bernardo R, Moreau L, Charcosset A (2006) Number and fitness of selected individuals in marker-assisted and phenotypic recurrent selection. Crop Sci 46:1972-1980

Bernardo R, Yu J (2007) Prospects for genomewide selection for quantitative traits in maize. Crop Sci 47:1082-1090

Beyene Y, Semagn K, Mugo S, Tarekegne A, Babu R, Meisel B, Sehabiague P, Makumbi D, Magorokosho C, Oikeh S, Gakunga J, Vargas M, Olsen M, Prasanna BM, Banziger M, Crossa J (2015) Genetic gains in grain yield through genomic selection in eight bi-parental maize populations under drought stress. Crop Sci 55:154-163

doi:10.6342/NTU201701806

Breiman L (2001) Random forests. Machine Learning 45:5-32

Combs E, Bernardo R (2013) Genomewide selection to introgress semidwarf maize germplasm into U.S. corn belt inbreds. Crop Sci 53:1427-1436

Crossa J, Perez P, Hickey J, Burgueno J, Ornella L, Ceron-Rojas J, Zhang X, Dreisigacker S, Babu R, Li Y, Bonnett D, Mathews K (2014) Genomic prediction in CIMMYT maize and wheat breeding programs. Heredity 112:48-60

David Desrousseaux, Florian Sandron, Aurelie Siberchicot, Christine Cierco-Ayrolles, Mangin B (2013) LDcorSV: linkage disequilibrium corrected by the structure and the relatedness

Dekkers JC (2007) Prediction of response to marker-assisted and genomic selection using selection index theory. J Anim Breed Genet 124:331-341

Desta ZA, Ortiz R (2014) Genomic selection: genome-wide prediction in plant improvement. Trends Plant Sci 19:592-601

Duangjit J, Causse M, Sauvage C (2016) Efficiency of genomic selection for tomato fruit quality. Mol Breed 36:29

Dunckel S, Crossa J, Wu SY, Bonnett D, Poland J (2017) Genomic selection for increased yield in synthetic-derived wheat. Crop Sci 57:713-725

Fox J, Weisberg S (2011) An R Companion to Applied Regression, 2nd edn. Thousand Oaks CA: Sage

Goddard M (2009) Genomic selection: prediction of accuracy and maximisation of long term response. Genetica 136:245-257

Gorjanc G, Battagin M, Dumasy J-F, Antolin R, Gaynor RC, Hickey JM (2017) Prospects for cost-effective genomic selection via accurate within-family imputation. Crop Sci 57:216-228

Gowda M, Zhao Y, Wurschum T, Longin CF, Miedaner T, Ebmeyer E, Schachschneider R, Kazman E, Schacht J, Martinant JP, Mette MF, Reif JC (2014) Relatedness severely impacts accuracy of marker-assisted selection for disease resistance in hybrid wheat. Heredity 112:552-561

doi:10.6342/NTU201701806

Grenier C, Cao TV, Ospina Y, Quintero C, Chatel MH, Tohme J, Courtois B, Ahmadi N (2015) Accuracy of genomic selection in a rice synthetic population developed for recurrent selection breeding. PLoS One 10:e0136594

Habyarimana E, Parisi B, Mandolino G, Wehling P (2017) Genomic prediction for yields, processing and nutritional quality traits in cultivated potato (*Solanum tuberosum* L.). Plant Breeding 136:245-252

He S, Schulthess AW, Mirdita V, Zhao Y, Korzun V, Bothe R, Ebmeyer E, Reif JC, Jiang Y (2016) Genomic selection in a commercial winter wheat population. Theor Appl Genet 129:641-651

Heffner EL, Jannink J-L, Sorrells ME (2011) Genomic selection accuracy using multifamily prediction models in a wheat breeding program. Plant Genome 4:65-75

Heffner EL, Lorenz AJ, Jannink JL, Sorrells ME (2010) Plant breeding with genomic selection: gain per unit time and cost. Crop Sci 50:1681-1690

Heffner EL, Sorrells ME, Jannink J-L (2009) Genomic selection for crop improvement. Crop Sci 49:1-12

Jacobson A, Lian L, Zhong S, Bernardo R (2015) Minimal loss of genetic diversity after genomewide selection within biparental maize populations. Crop Sci 55:783-789

Lado B, Barrios PG, Quincke M, Silva P, Gutiérrez L (2016) Modeling genotype × environment interaction for genomic selection with unbalanced data from a wheat breeding Program. Crop Sci 56:2165-2179

Liaw A, Wiener M (2002) Classification and regression by randomForest. R News 2:18-22

Lin Z, Cogan NO, Pembleton LW, Spangenberg GC, Forster JW, Hayes BJ, Daetwyler HD (2016) Genetic gain and inbreeding from genomic selection in a simulated commercial breeding program for perennial ryegrass. Plant Genome 9:1-12

Longin CF, Mi X, Wurschum T (2015) Genomic selection in wheat: optimum allocation of test resources and comparison of breeding strategies for line and hybrid breeding. Theor Appl Genet 128:1297-1306

doi:10.6342/NTU201701806

Lorenz AJ, Chao SM, Asoro FG, Heffner EL, Hayashi T, Iwata H, Smith KP, Sorrells ME, Jannink JL (2011) Genomic selection in plant breeding: knowledge and prospects. Adv Agron 110:77-123

Lorenz AJ, Smith KP (2015) Adding genetically distant individuals to training populations reduces genomic prediction accuracy in barley. Crop Sci 55:2657-2667

Lorenz AJ, Smith KP, Jannink JL (2012) Potential and optimization of genomic selection for fusarium head blight resistance in six-row barley. Crop Sci 52:1609-1621

Lorenzana RE, Bernardo R (2009) Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theor Appl Genet 120:151-161

Ma Y, Reif JC, Jiang Y, Wen Z, Wang D, Liu Z, Guo Y, Wei S, Wang S, Yang C, Wang H, Yang C, Lu W, Xu R, Zhou R, Wang R, Sun Z, Chen H, Zhang W, Wu J, Hu G, Liu C, Luan X, Fu Y, Guo T, Han T, Zhang M, Sun B, Zhang L, Chen W, Wu C, Sun S, Yuan B, Zhou X, Han D, Yan H, Li W, Qiu L (2016) Potential of marker selection to increase prediction accuracy of genomic selection in soybean (*Glycine max* L.). Mol Breed 36:113

Marulanda JJ, Melchinger AE, Würschum T, Pillen K (2015) Genomic selection in biparental populations: assessment of parameters for optimum estimation set design. Plant Breeding 134:623-630

Marulanda JJ, Mi X, Melchinger AE, Xu JL, Wurschum T, Longin CF (2016) Optimum breeding strategies using genomic selection for hybrid breeding in wheat, maize, rye, barley, rice and triticale. Theor Appl Genet 129:1901-1913

Massman JM, Jung H-JG, Bernardo R (2013) Genomewide selection versus marker-assisted recurrent selection to improve grain yield and stover-quality traits for cellulosic ethanol in maize. Crop Sci 53:58-66

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics Society of America 157:1819-1829

Mohammadi M, Tiede T, Smith KP (2015) PopVar: a genome-wide procedure for predicting genetic variance and correlated response in biparental breeding populations. Crop Sci 55:2068-2077

doi:10.6342/NTU201701806

Moreau L, Charcosset A, Gallais A (2004) Experimental evaluation of several cycles of marker-assisted selection in maize. Euphytica 137:111-118

Onogi A, Ideta O, Inoshita Y, Ebana K, Yoshioka T, Yamasaki M, Iwata H (2015) Exploring the areas of applicability of whole-genome prediction methods for Asian rice (*Oryza sativa* L.). Theor Appl Genet 128:41-53

Park T, Casella G (2008) The Bayesian lasso. Journal of the American Statistical Association 103:681-686

Perez P, de los Campos G (2014) Genome-wide regression and prediction with the BGLR statistical package. Genetics 198:483-495

Perez P, de Los Campos G, Crossa J, Gianola D (2010) Genomic-enabled prediction based on molecular markers and pedigree using the Bayesian linear regression package in R. Plant Genome 3:106-116

Poland J, Endelman J, Dawson J, Rutkoski J, Wu S, Manes Y, Dreisigacker S, Crossa J, Sánchez-Villeda H, Sorrells M, Jannink J-L (2012) Genomic selection in wheat breeding using genotyping-by-sequencing. Plant Genome 5:103-113

R Core Team (2016). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rajsic P, Weersink A, Navabi A, Peter Pauls K (2016) Economics of genomic selection: the role of prediction accuracy and relative genotyping costs. Euphytica 210:259-276

Rutkoski J, Singh RP, Huerta-Espino J, Bhavani S, Poland J, Jannink JL, Sorrells ME (2015) Genetic gain from phenotypic and genomic selection for quantitative resistance to stem rust of wheat. Plant Genome 8:1-10

Sallam AH, Smith KP (2016) Genomic selection performs similarly to phenotypic selection in barley. Crop Sci 56:2871-2881

Schaeffer LR (2006) Strategy for applying genome-wide selection in dairy cattle. J Anim Breed Genet 123:218-223

Spindel J, Begum H, Akdemir D, Virk P, Collard B, Redona E, Atlin G, Jannink JL, McCouch SR (2015) Genomic selection and association mapping in rice (*Oryza sativa*): effect of trait genetic architecture, training population composition, marker

doi:10.6342/NTU201701806

number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. PLoS Genet 11:e1004982

Storlie E, Charmet G (2013) Genomic selection accuracy using historical data generated in a wheat breeding program. Plant Genome 6

Taylor JF, Taylor KH, Decker JE (2016) Holsteins are the genomic selection poster cows. Proc Natl Acad Sci U S A 113:7690-7692

Technow F, Schrag TA, Schipprack W, Bauer E, Simianer H, Melchinger AE (2014) Genome properties and prospects of genomic prediction of hybrid performance in a breeding program of maize. Genetics 197:1343-1355

Thomson MJ (2014) High-throughput SNP genotyping to accelerate crop improvement. Plant Breeding and Biotechnology 2:195-212

Wang Y, Mette MF, Miedaner T, Gottwald M, Wilde P, Reif JC, Zhao Y (2014) The accuracy of prediction of genomic selection in elite hybrid rye populations surpasses the accuracy of marker-assisted selection and is equally augmented by multiple field evaluation locations and test years. BMC Genomics 15:556

Wong CK, Bernardo R (2008) Genomewide selection in oil palm: increasing selection gain per unit time and cost with small populations. Theor Appl Genet 116:815-824

Wurschum T, Maurer HP, Weissmann S, Hahn V, Leiser WL (2017) Accuracy of within- and among-family genomic prediction in triticale. Plant Breeding 136:230-236

Yamamoto E, Matsunaga H, Onogi A, Kajiya-Kanegae H, Minamikawa M, Suzuki A, Shirasawa K, Hirakawa H, Nunome T, Yamaguchi H, Miyatake K, Ohyama A, Iwata H, Fukuoka H (2016) A simulation-based breeding design that uses whole-genome prediction in tomato. Sci Rep 6:19454

Yamamoto E, Matsunaga H, Onogi A, Ohyama A, Miyatake K, Yamaguchi H, Nunome T, Iwata H, Fukuoka H (2017) Efficiency of genomic selection for breeding population design and phenotype prediction in tomato. Heredity 118:202-209

You FM, Booker HM, Duguid SD, Jia G, Cloutier S (2016) Accuracy of genomic selection in biparental populations of flax (*Linum usitatissimum* L.). Crop Journal 4:290-303

doi:10.6342/NTU201701806

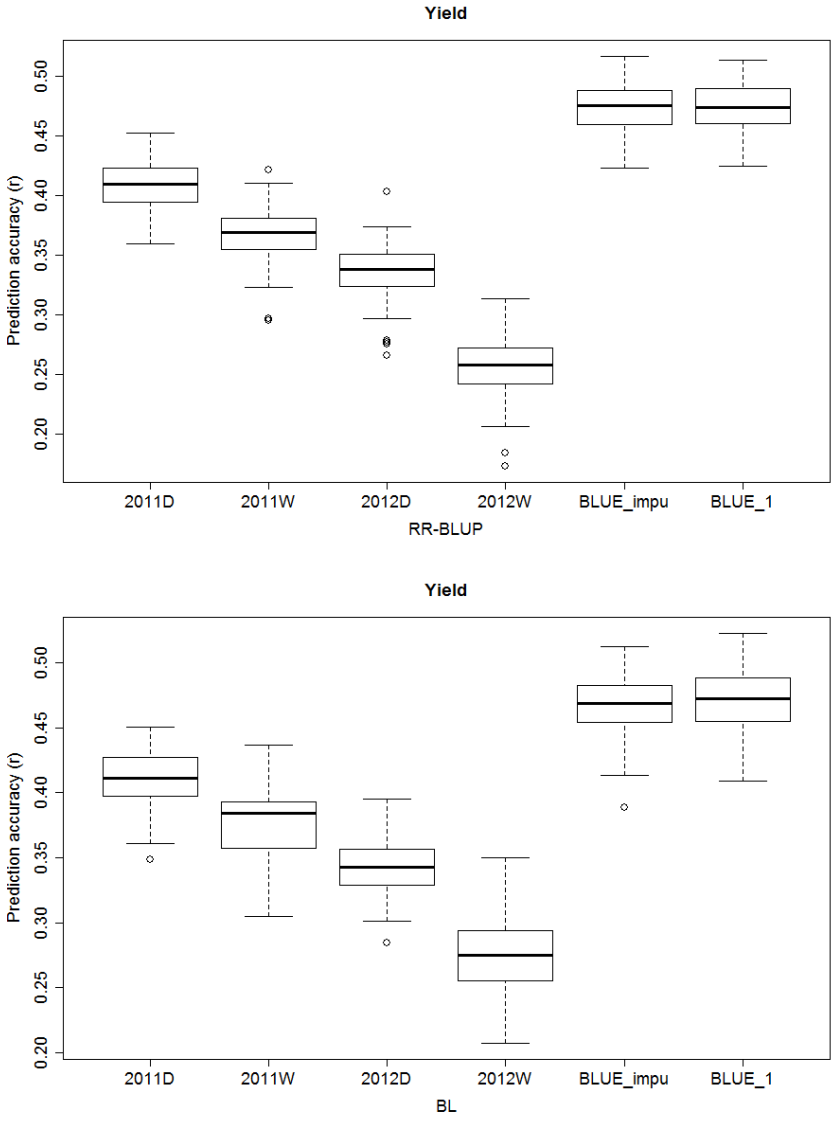
Zhao Y, Mette MF, Reif JC, Ordon F (2015) Genomic selection in hybrid breeding. Plant Breeding 134:1-10

Zhong S, Dekkers JC, Fernando RL, Jannink JL (2009) Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. Genetics 182:355-364

Ziyomo C, Bernardo R (2013) Drought tolerance in maize: indirect selection through secondary traits versus genomewide selection. Crop Sci 53:1269-1275

doi:10.6342/NTU201701806

附錄



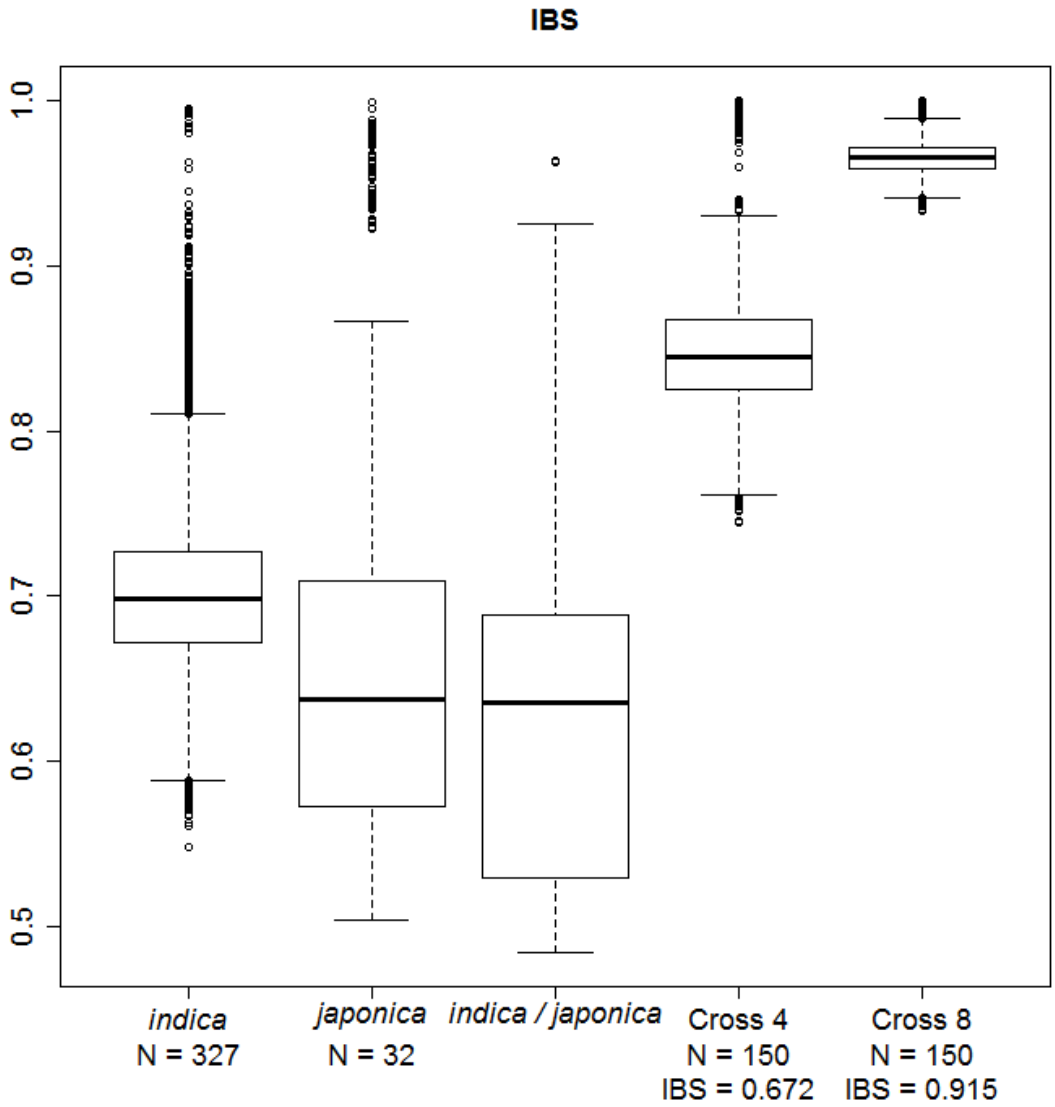
附圖一、測試建立預測模型資料之預測準確度

上圖為 RR-BLUP 模型，下圖為 BL 模型，100 次十折交叉驗證結果。各盒鬚圖為使用各季資料或 BLUE 進行模型訓練。各季資料及 BLUE_impu 使用 Spindel et al. (2015) 利用 TASSEL 3.0 FastImputationBitFixedWindow 擴充套件替補缺值後之基因型資料 (imputed genotypic data)；BLUE_1 則以主要對偶基因取代替補資料。

Fig. S2 Preliminary test to establish the prediction model

The upper panel shows the RR-BLUP model, the lower panel shows the BL model; both from 100 times ten-fold cross-validation results. Each season or BLUE were used for model training. For original seasonal data and BLUE_impu, genotypic data imputed using TASSEL 3.0 FastImputationBitFixedWindow plugin (Spindel et al. 2015) were used; for BLUE_1, imputed data were replaced by major allele, the allele 1.

doi:10.6342/NTU201701806



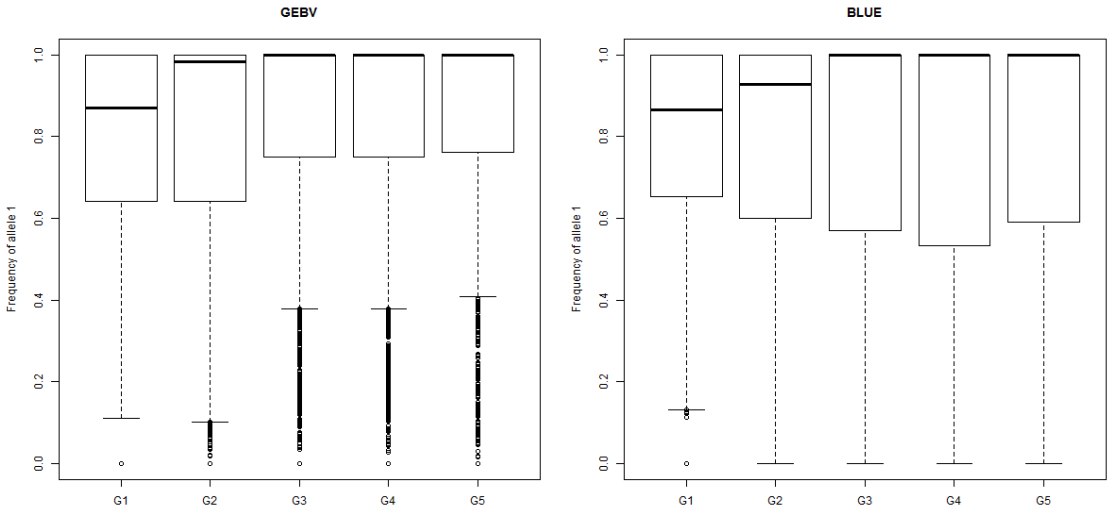
附圖二、*indica* 品系、*japonica* 品系、*indica* 與 *japonica* 品系間以及雙親本雜交 F_6 族群之 IBS 值分布範圍

N 為族群大小，F_6 族群分別為親本 IBS 最低與最高之兩個雜交組合。

Fig. S3 The distribution of IBS for *indica* lines, *japonica* lines, *indica* and *japonica*, and F_6 populations derived from bi-parental cross

N is population size, and the two F_6 populations are selected for their parents showed minimum and maximum IBS across the 10 bi-parental populations.

doi:10.6342/NTU201701806



附圖三、5,264 個 SNP 之對偶基因 1 頻度於輪迴選種各循環 F_2 世代變化

Fig. S4 The change of frequency of allele 1 for 5,264 SNPs along recurrent selection

doi:10.6342/NTU201701806