

# 以雙限制酶切位點標定法開發甘藍品種鑑別 SNP 套件<sup>1</sup>

林延諭<sup>2\*</sup>、蕭政弘<sup>3</sup>、王群山<sup>4</sup>、張心怡<sup>5</sup>

## 摘 要

本研究以甘藍為對象，運用次世代定序技術中的 ddRADseq 進行基因體分析。通過對 112 個育種材料的定序與分析，成功識別 745,196 個變異位點，其中包括 648,810 個 SNP 位點。經嚴格的條件篩選，保留 3,565 個高品質 SNP 位點，並藉由這些 SNP 之基因型值，建構甘藍育種材料遺傳異質性及遺傳歧異性資料，為材料管理與育種提供豐富之參考訊息。由其中挑選 8 組可區分 109 個品系之 SNP 合成分析套組，於 24 個驗證材料中之可用性試驗結果顯示，本分析套組具高度的穩定性及可靠性。本研究不僅開發輔助甘藍育種之分子標誌套組，建立之甘藍 ddRADseq 試驗與分析流程可經濟且快速的獲得大量樣品的基因型值，可作為甘藍分子輔助育種及基因定位有利工具。

**關鍵字：**甘藍、單一核苷酸多型性、雙限制酶切位點標定法

## 前 言

甘藍(*Brassica oleracea* var. *capitata*)俗稱高麗菜，為臺灣最大宗的葉菜類蔬菜，年栽培面積約 8,000 公頃(行政院農業委員會農糧署，2021)，於臺灣可周年生產，然夏季較常發生產量低、球型外觀與品質不佳及頂燒症(高溫缺鈣生理障礙)等問題(蕭，2012)。為此，臺中區農業改良場重新引入耐熱種質資源，於 96 年及 102 年先後育成‘甘藍台中 1 號’、及‘甘藍台中 2 號’(蕭，2013)。並且於 111 年利用次世代定序系統，以基因體重定序開發甘藍雜交種子純度檢定 SNP 分子標誌，以有效控制雜交種子純度(林等人，2021)。近年，為改善甘藍黃葉病(*Fusarium yellow*；*Fusarium oxysporum*)，臺中場積極引入抗病種原，並利用分子標誌輔助將抗病基因導入優良自交系中，隨導人品系增加與回交、自交世代推進，育種材料規模也不斷增加，為提高管理效率，並於後續進一步利用分子標誌輔助回交選拔，掌握現有育種材料之遺傳背景有其必要性。

<sup>1</sup> 農業部臺中區農業改良場第 1087 號研究報告。

<sup>2</sup> 農業部臺中區農業改良場前助理研究員，現為農業試驗所助理研究員。

<sup>3</sup> 臺中區農業改良場研究員兼副場長。

<sup>4</sup> 臺灣大學農藝學系博士。

<sup>5</sup> 臺灣大學農藝學系研究助理。

\*通訊作者 email：ylin@tari.gov.tw

隨次世代定序技術發展，定序成本大幅下降，然而在依舊有限的定序資源下，對大量個體進行全基因體定序仍然是不切實際的。憑藉著過往 RFLP 與 AFLP 標誌的使用經驗，結合限制酶對基因體取樣與次世代定序的技術應運而生，Baird 等人(2008)發表基於單一限制酶切的限制酶切位點標法定序(restriction site-associated DNA sequencing, RADseq)，展示藉由限制酶之選擇可調控 SNP 標誌密度。隨後於 2012 年，基於雙限制酶切的 Two-Enzyme GBS (Poland *et al.*, 2012)及 ddRADseq (Peterson *et al.*, 2012)不約而同的被發表，除原作為降低基因體複雜度的 rare cutter 外，增加的 common cutter 與 Y-轉接子，可確保定序文庫中的片段皆為經雙限制酶切的序列，使其更具一致性與均勻。Two-Enzyme GBS 與 ddRADseq 除了定序文庫結構上些微的差異外，ddRADseq 更強調片段大小的篩選，藉由嚴謹的篩選，減少緊鄰的區塊被重複定序的機會，同時提高不同樣品於目標區間的定序量，進一步提高使用彈性與穩健性。

目前，十字花科薺苔屬作物已有薺薹(*Brassica rapa*)使用 *Pst*I 與 *Msp*I 限制酶組合進行 ddRADseq 並定位抗病 QTL (Laila *et al.*, 2019)，而芥菜(*Brassica juncea*)則使用 *Mse*I 及 *Sac*I 限制酶組合進行 ddRADseq 辨別 SNP 位點的紀錄(Sudan *et al.*, 2019)，甘藍則尚未有相關文獻發表。因此，本研究將由資料模擬方式選定甘藍雙限制酶組合，利用 112 個育種材料建立與驗證甘藍 ddRAD 試驗與資料分析流程，開發 SNP 分析套組供材料管理、雜交成功率檢測及後續分子輔助回交育種使用。

## 材料與方法

### 一、植物材料

本研究以 52 個具耐熱或抗黃葉病及園藝性狀優良之甘藍自交系、39 個抗病育種材料、7 個重要雜交一代品種及 14 個抗病雜交一代品種作為材料，取嫩葉 0.05 g 萃取 DNA。

### 二、植物核酸萃取

植株 DNA 抽取過程採用自動化核酸萃取儀(Smart LabAssist-32, 臺灣圓點奈米技術股份有限公司, 臺灣)，與試劑組(TANBead Plant DNA Auto Plate, 臺灣圓點奈米技術開發有限公司, 臺灣)。將秤重之葉片置於 2 mL 厚壁試管中，加入鋼珠及 800  $\mu$ L Lysis buffer 震盪 1 min，靜置於室溫 5 min。以 8,000 g 離心 1 min，將上清液注入已預先分注試劑之 96 孔盤，放入自動化核酸萃取儀，依操作手冊選擇(BIO-W4-BBTO)程式，完成 DNA 萃取。

### 三、定序文庫建置

- (一) DNA 定量：取 2  $\mu$ L 樣品 DNA 以螢光計(Invitrogen Qubit 4, Thermo Fisher Scientific Inc., USA) 定量，將樣品稀釋至 30 ng/ $\mu$ L，稀釋後以 Quant-iT™ PicoGreen™ (Thermo Fisher Scientific Inc., USA)檢測樣品濃度，濃度偏差較大之樣品需重新稀釋，以確保投入量相同。
- (二)雙重限制酶切：取 20  $\mu$ L DNA 以 *Pst*I、*Taq*I (New England BioLabs, USA) 於 37°C 30 min 進

行酵素酶切，升溫至 65°C 30 min 中止反應。

(三)轉接子黏合：以 T4 DNA Ligase (New England Biolabs, USA)於 16°C 反應隔夜黏合轉接子，後升溫至 65°C 20 min 中止反應，並緩慢降溫至 23°C，反應溶液以 0.8x 體積之 AMPure XP (Bechman Coulter Life Sciences, USA) 純化並回溶於 40  $\mu$ L QH<sub>2</sub>O 中。

(四)聚合酶連鎖反應與 dual index 導入：以經過雙限制酶剪切並黏合轉接子的 DNA 作為模板，將帶有 Nextera XT dual index 序列與 illumina 定序所需的序列引子組以 PCR 的方式加入定序文庫，以 1 U Phusion HF SNA Polymerases (Thermo Fisher Scientific Inc., USA)、200  $\mu$ M dNTPs 進行 2 steps PCR，反應先於 98 °C 反應 30 s 後以 98 °C 10 s、72 °C 30 s 進行 20 個循環、於 72 °C 反應 10 min 後於 4 °C 保存。PCR 擴增產物再次以 0.8x AMPure XP 純化。

(五)樣品混合及片段篩選：各樣品的 PCR 產物以 Quant-iT™ PicoGreen™ PicoGreen 定量後等量混合，並以 BluePippin (Sage Science, USA)針對 250-600 bp 片段回收，並以 1x AMPure XP 濃縮完成定序文庫建製。定序文庫委託基龍米克斯公司(基龍米克斯生物科技股份有限公司，臺灣)以 Illumina HiSeq4000 進行 150 bp 雙邊定序。

#### 四、SNP 探勘與基因型判讀

定序資料以 AdapterRemoval (Schubert *et al.*, 2016)去除短讀序列中屬於 illumina 定序結構之序列。修剪後的讀序以 Bowtie2 (Langmead & Salzberg, 2012)將短讀序列比對(alignment)至參考序列 Ensembl 資料庫中 *Brassic oleracea* 之參考基因組序列(release-42)。比對後的讀序以 samtools (Danecek *et al.*, 2016) 依照物理圖譜位置排序並壓縮轉換為 BAM 檔案，再以 GATK (Genome Analysis ToolKit)(Danecek *et al.*, 2011)的最佳操作流程(best practice)以綜合參試樣品(joint calling)方式探勘基因體變異位點並判讀基因型。利用全部樣本的 BAM 檔案以 GATK 的 HaplotypeCaller 執行初次變異位點探勘並將產出的 VCF 檔案視為 dbSNP 作為 BQSR(base quality score recalibration)的共變因子(covariate)，建立讀序品質校正模型，最後將參試樣品的 BAM 檔案以讀序品質校正模型修正讀序品質，校正後的 BAM 檔案，再進入變異位點探勘與基因型判讀(Variant Calling)。

變異位點探勘與基因型判讀流程依 gVCF 流程進行，以 GATK 的 HaplotypeCaller 指令將校正過 BAM 檔案壓縮成 gVCF 檔案，再以 genotypeGVCFs 指令合併參試樣品的 gVCF 文件，探勘變異位點並判讀基因型產生最終 VCF 檔案，完成生物資訊分析流程。

#### 五、分子標誌開發與樣品檢測

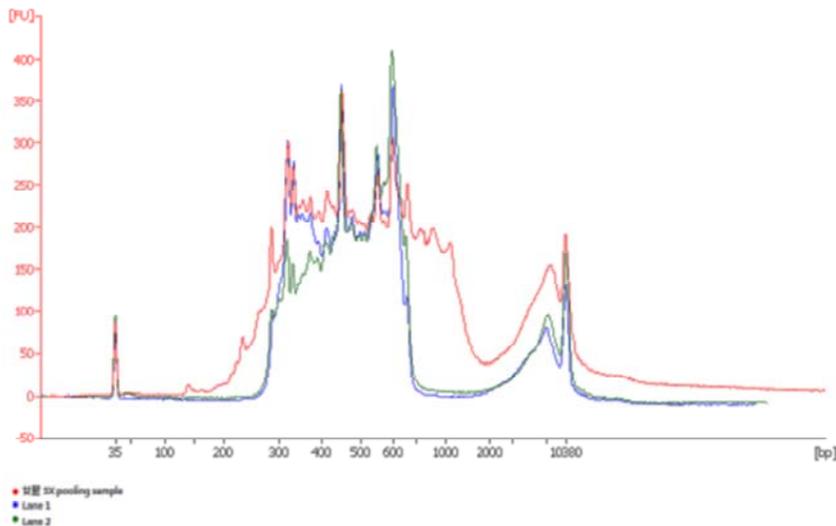
針對每個篩選出的 SNP 位點，檢查其上下游序列中是否存在鄰近變異，以確保專一性，並僅保留位於模擬限制酶預測片段內的 SNPs。於 ThermoFisher 網站以 Custom TaqMan Assay Design Tool 設計 TaqMan 分析套組，並委託金萬林公司(金萬林企業股份有限公司，臺灣)合成。TaqMan 分析套組之即時聚合酶連鎖反應(real-time PCR)，反應體積為 10  $\mu$ l，包含 2.5  $\mu$ l ddH<sub>2</sub>O、5  $\mu$ l TaqMan GTXpress Master Mix(Thermo Fisher Scientific, USA)、0.5  $\mu$ l TaqMan SNP Genotyping Assay

20X(Thermo Fisher Scientific, USA)、2  $\mu$ l template DNA，樣品混合後以即時聚合酶鏈鎖反應儀(QuantStudio 3 Real-Time PCR Systems, Thermo Fisher)進行反應，反應先以 60°C 30s 進行螢光背景預讀、接著以 95°C 20s 活化聚合酶、循環階段為 95°C 1s 及 60°C 20s 共 40 個循環，最終以 60°C 30s 進行反應結束，於每次循環及反應結束前偵測螢光值。

## 結果與討論

### 一、ddRAD 定序文庫製備與定序：

以臺中場育種人員選定 52 個具有耐熱、抗黃葉病及園藝性狀優良之甘藍自交系、39 個抗病育種材料、7 個重要雜交一代品種及 14 個抗病雜交一代品種為核心材料。為提高定序資源使用效率，本研究在預備試驗中模擬分析不同限制酶組合所產生片段的大小、數量及其分布(資料未呈現)，並依據所得結果，選定 *Pst*I 與 *Taq*I 作為限制酶組合。於雙重酶切簡化基因體後，導入 Nextera XT Index Kit 條碼系統後，混合成定序文庫，再進行片段長度篩選，經 bioanalyzer 分析結果顯示定序文庫主要片段大小為 300-600 bp(圖一)，符合本試驗預期之設定。經 Illumina HiSeq 4000 平台定序後，所有參試樣品共計獲得 489 百萬筆讀序，平均每樣品有 4.4 百萬筆讀序，所有樣品之平均序列品質分數(Phred score)為 38.2，約 91.6%序列 Q score > 30 顯示多數序列具有良好定序品質。

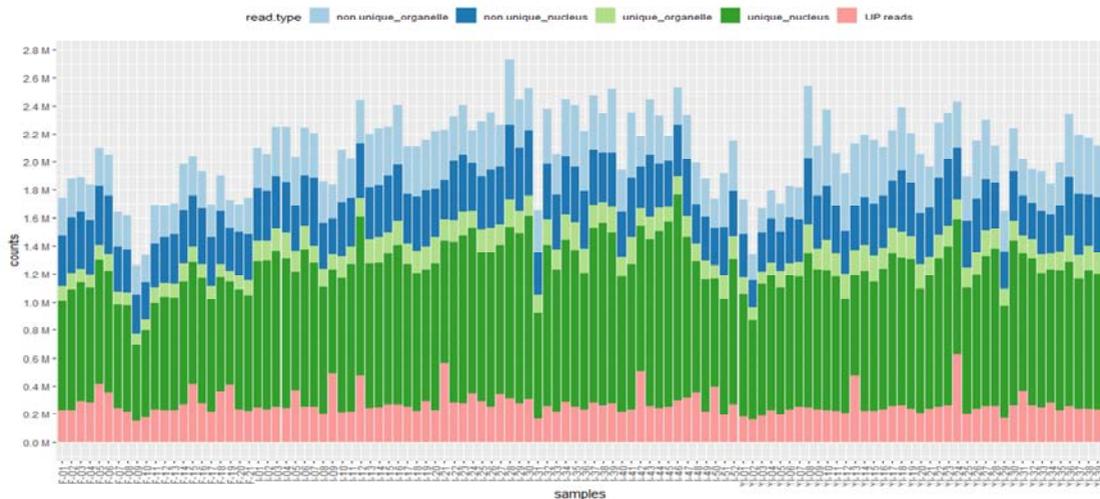


圖一、甘藍簡化基因體文庫以 Bioanalyzer 分析之結果，紅色為原始文庫，藍色及綠色為使用 BluePippin 篩選回收後之定序文庫。

Fig. 1. Bioanalyzer analysis results of the simplified genome library for cabbage, depicting the original library in red and the sequenced libraries after BluePippin size selection and recovery in blue and green, respectively.

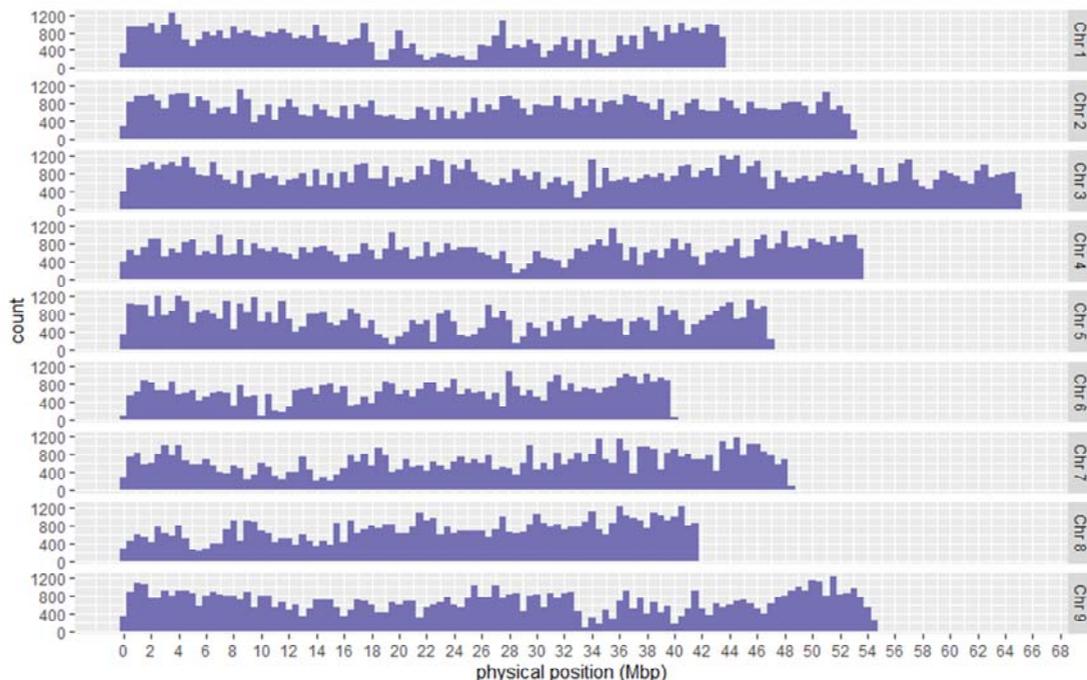
## 二、序列資料處理與 SNP 辨識：

原始序列於移除轉接子序列後，便與參考序列進行比對，在全部 235,222,316 筆序列中，有 205,213,007 筆序列可成功比對至參考序列，比對成功率為 87.2%，各樣品成功比對序列介於 73.6-90.7%間，雖然低於林等人(2021)甘藍基因體重定序之 97.8%，但高於同樣使用 ddRADseq 的 Laila 等人(2019)於蕓苔 68.7%，顯示本限制酶組合與建庫流程，於甘藍中具有良好應用潛力。在可成功比對的序列中，有 62.3%序列對應至單一之基因座，各樣品平均值介於 54.4-73.7%(圖二)。變異點辨識部分，總計辨識出 745,196 個變異點，其中有 648,810 個 SNP 位點。與林等人(2021)之甘藍基因體重定序結果相較，其由 209.8 M 對序列中辨識出 7.5 M 個變異點，單位定序量之變異位點高於本試驗；然進一步計算樣品平均定序量後，本試驗平均每百萬筆序列可辨識出 345,855 個變異位點，高於該研究的 107,296 個變異位點，顯示 ddRADseq 於相同定序資源下可容納更多定序樣品外，還具有相當優異的變異位點辨識比例。推測此結果與限制酶組合有關，Fellers (2008) 於比較 9 種限制酶對小麥基因體酶切之影響，挑選包含 *Pst*I 在內 4 個對甲基化區域敏感的限制進一步檢視其於小麥、玉米及菸草定序結果之影響，結果顯示 *Pst*I 限制酶可顯著減少重複序列比例並提高基因富集效率。



圖二、甘藍 ddRADseq 序列比對結果，各樣品序列數及分類。

Fig. 2. ddRADseq alignment results for cabbage, showing the number and classification of sequences for each sample.

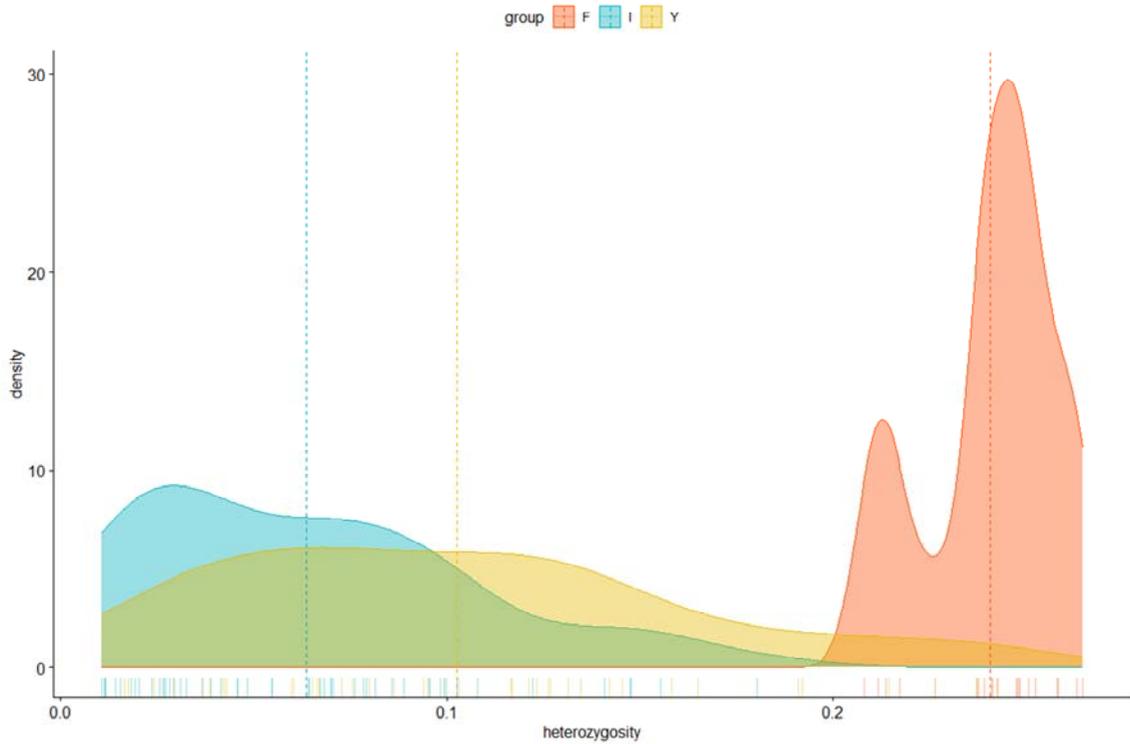


圖三、SNP 位點於各染色體分布情形。

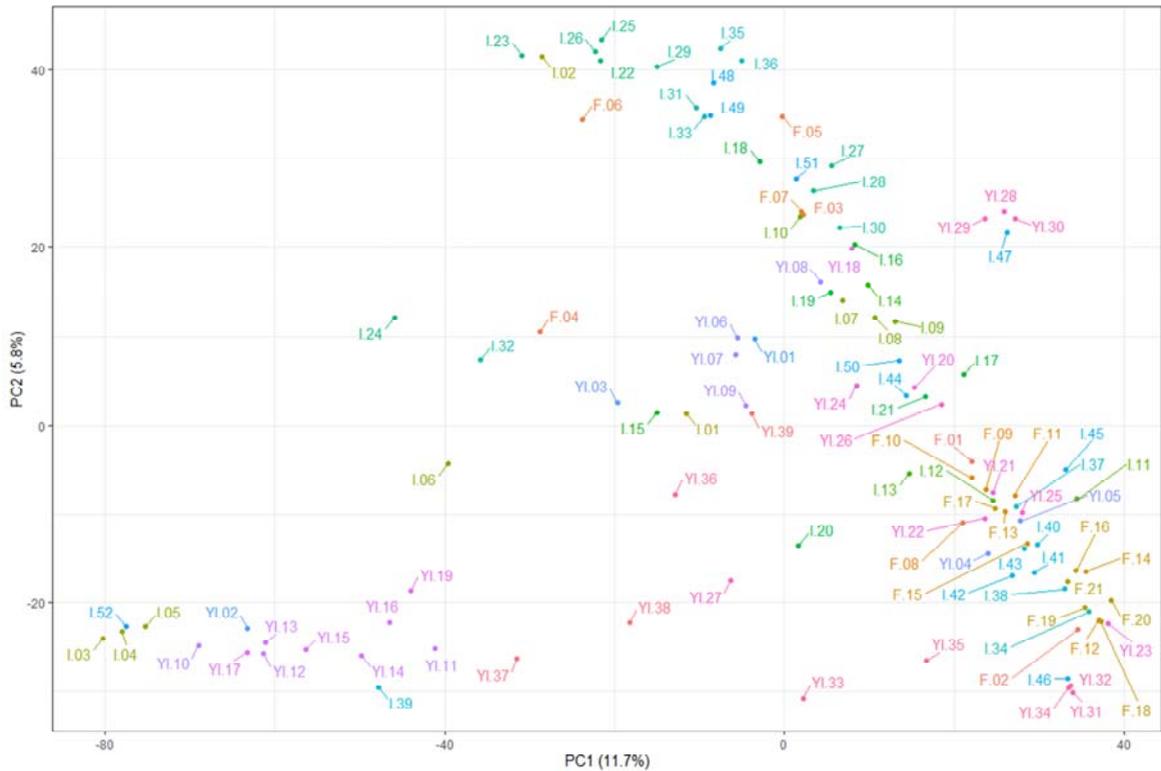
Fig. 3. Distribution of SNP positions across chromosomes.

### 三、樣品遺傳異質性與遺傳歧異分析：

SNP 經過 QUAL、QD、缺值率及多型性篩選，並剔除前後 50 bp 內有其他變異點之 SNP 後，平均每樣品有 18,356 個基因型資料可用於後續遺傳分析。遺傳異質性分析部分，可將樣品分為  $F_1$  與自交系兩類。 $F_1$  群體包含臺灣主要栽培品種與抗黃葉病之商業品種，其 SNP 異質結合率介於 20.8-26.4%。自交系的異質結合率則隨自交世代不同，有較大差異，變化幅度介於 1.1-23.7%，其中臺中場已育成品種之親本異質結合率較低，介於 1.1-1.2% 間，顯示種原維持狀態良好(圖四)。為使新品種具有更高之一致性與穩定性，將挑選同質結合率較高之單株進行自交純化，作為親本材料。藉由主成分分析結果(圖五)，以視覺化之方式提供育種人員檢視導入抗病基因之品系與原品系之差異程度或選定雜交組合等，可協助擬定後續育種策略。



圖四、112 個 cabbage 樣品遺傳異質結合率密度分布：雜交一代(F，紅色)、自交系(I，藍色)及導入抗病基因之自交系(Y，黃色)，其中 I 及 Y 包含不同自交世代品系。  
Fig. 4. Density plot of genetic heterogeneity in 112 cabbage samples:  $F_1$  hybrids (F, red), inbred lines (I, blue), and disease-resistant inbred lines (Y, yellow), where both I and Y encompass multiple selfing generations.



圖五、112個cabbage樣品遺傳歧異之PCA分析結果。

Fig. 5. PCA analysis results depicting genetic diversity among the 112 cabbage lines.

#### 四、SNP 分析套組開發：

為進一步篩選以 ddRAD 定序獲得之變異位點資訊，由使用 Perl 語言與 R 分析軟體，對 VCF 檔進行解析，設定篩選條件。577,617 個變異位點中，QUAL 大於 1000 的計有 125,800 組、QD 大於 10 的計有 542,008 組，同時滿足這兩個條件的高質量變異位點計有 123,3141 組。通過 is.SNV 及 singleALTOOnly 標籤篩選的計有 497,950 SNPs，同時滿足以上條件的計有 104,700 筆。因本計畫規劃開育種材料品系鑑別分析套組，為加快高鑑別力標誌挑選，讀入所有 112 個育種材料的基因型值，其中有多型性的有 96,759 筆，缺值率小於 25%的有 55,962 筆，再增加基因型品質分數需達 30 以上之條件後，計有 40,363 筆 SNPs。為減少後續合成分析套組時因鄰近變異位點導致專一性不足的問題，排除於鄰近位置有其他變異的 SNPs，為再一步提高變異位點可信度，以甘藍參考序列模擬限制酶，建立預測目標片段物件，僅保留位於預測片段的 SNPs，最終獲得 7,451 筆 SNPs。以 Perl 程式碼摘取序列資料後，以 Custom TaqMan Assay Design Tool 進行引子設計，設計成功率為 92.2%，共獲得 6,870 個分析套組資訊。透過設定不同的 PIC (Polymorphism Information Content) 閾值，觀察篩選出之 SNP 數量變化，最終將 PIC 值達 0.4 以上作為篩選條件，獲得 3,565 筆 SNPs (表一)。

接著，為確保挑選的 SNP 組合對廣泛樣品的鑑別力，排除親緣關係過於接近的 I-11、I-12、I-13 與 I-TN1-M，保留能夠區分其餘 109 個甘藍育種材料的 SNP 組合，並使用 IGV (Integrative Genomics Viewer) 進行視覺化驗證，最終選出 8 個 SNP 作為分析套組(表二)。此階段的分析套組僅經由 ThermoFisher 軟體計算評估，仍需實際驗證 TaqMan Assay 是否能穩定判讀基因型結果。如果上述 8 個 SNP 中有任何一個無法穩定判讀，會導致部分樣品無法區分，此時須進一步挑選可彌補識別能力的額外 SNP。為了確保驗證樣品的代表性，根據基因型表挑選候選 SNP 位點，並計算遺傳距離(Manhattan distance)，將差異在 7 個 SNP 以上的品系組挑出來，最終選出 24 個品系作為驗證族群。結果顯示 8 組分析套組皆能穩定判讀基因型，且與預期基因型一致(圖六)。

表一、設定之篩選條件下各染色體上變異位點數量

Table 1. Number of variants on each chromosome under the set filtering conditions

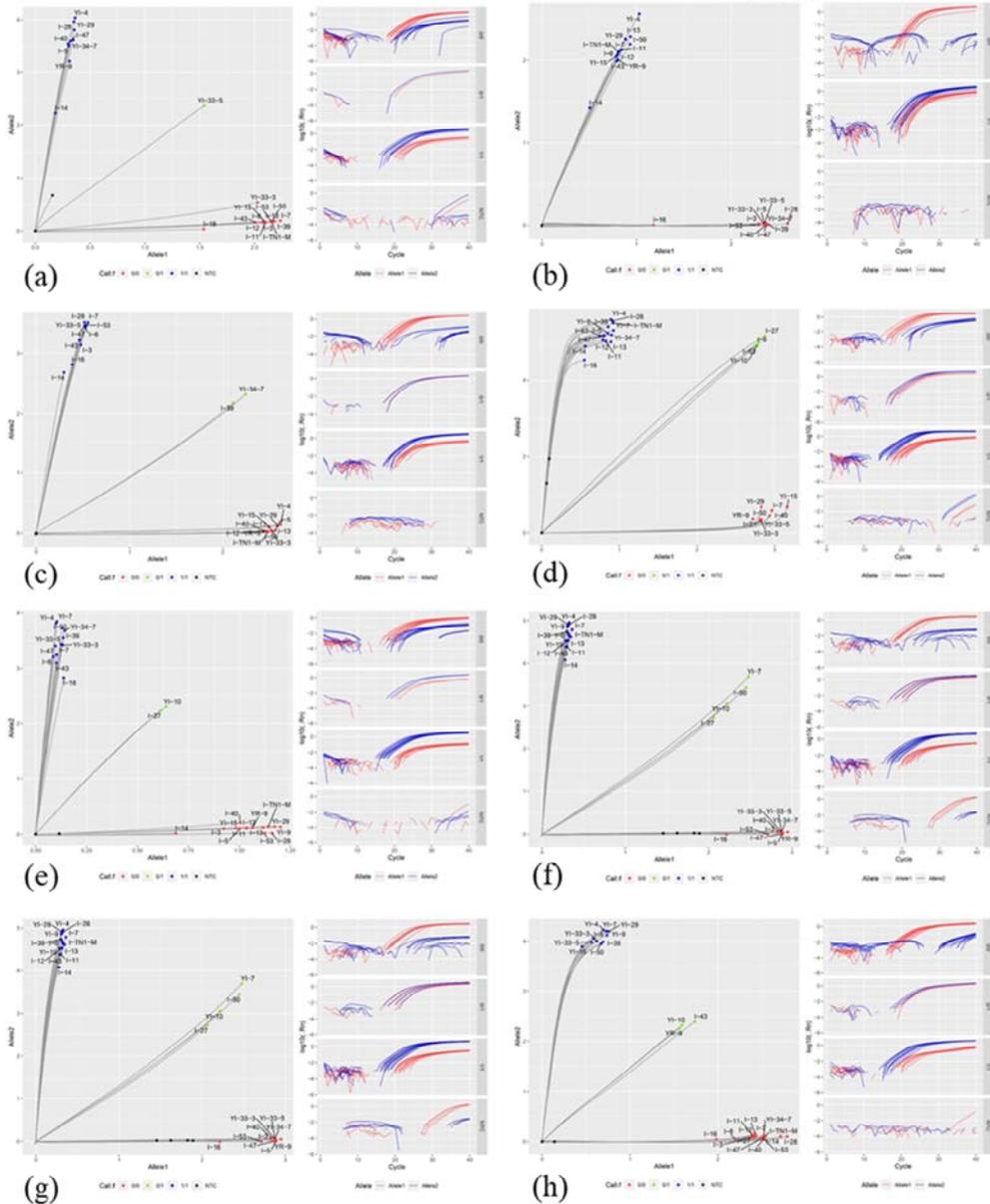
	Chr1	Chr2	Chr3	Chr4	Chr5	Chr6	Chr7	Chr8	Chr9	Total
All Variants	51,844	62,066	92,487	66,228	59,859	51,705	68,240	56,229	68,283	577,617
QUAL>1000	11,580	13,027	20,632	14,019	12,616	11,617	14,643	12,351	15,259	125,800
QD>10	48,396	57,841	87,098	62,231	56,017	48,743	64,573	52,409	64,087	542,008
HighQ	11,352	12,706	20,275	13,711	12,342	11,425	14,351	12,077	15,028	123,314
isSNP	44,636	53,805	79,108	57,269	51,464	44,591	59,224	48,184	59,083	497,950
HighQ.SNP	9,597	10,892	17,011	11,714	10,487	9,687	12,249	10,214	12,810	104,700
Polymorphism	8,884	10,118	15,851	10,888	9,763	8,847	10,940	9,520	11,948	96,759
Missing<0.25	5,162	5,675	9,013	6,285	5,941	5,115	6,255	5,500	7,016	55,962
GQ.poly	5,106	5,610	8,870	6,200	5,870	5,043	6,151	5,439	6,947	55,236
GQ.missing	3,845	4,004	6,806	4,443	4,373	3,665	4,503	3,990	4,971	40,600
GQ.missing.poly	3,822	3,986	6,756	4,421	4,345	3,641	4,469	3,971	4,952	40,363
in silico frags	689	691	1,256	873	882	705	751	754	850	7,451
Success in assay design	642	634	1,156	808	818	649	690	682	791	6,870
PIC>0.4	283	320	585	460	480	346	347	366	378	3,565

表二、本研究開發之 TaqMan Assay 引子與探針序列

Table 2. TaqMan Assay primer/probe sequences developed in this study

Assay Name <sup>1</sup>	Primer/Probe	Sequence (5'→3')
2:47705934_A/G	Forward	GCCTGGGTGATTTTATCATACTCTTC
	Reverse	CCTTTTCTACTTCAGAACCATACATTACT
	Probe 1 (VIC)	CTTTGGAGCTACTGCTC
	Probe 2 (FAM)	TTGGAGCTGCTGCTC
5:19077016_T/G	Forward	ATTCTTCTCTCTCAAATTTTCAGACCAA
	Reverse	CTCAGCTCAAAAATCAAGAACAATGGA
	Probe 1 (VIC)	CTTTAATGTACCACTGATATAA
	Probe 2 (FAM)	AATGTACCCCTGATATAA
6:22509740_T/A	Forward	CTGCAGTGGGAATGATGCA
	Reverse	CCTTCTACTACTACAGCCCAAGCAA
	Probe 1 (VIC)	AACAACCAGATTTATTCTT
	Probe 2 (FAM)	ACAACCAGATATATTCTT
4:64210791_G/A	Forward	CAAGAAAGCATTGGTTCACGTTATCC
	Reverse	GCTAGTGTGGTACATGGATTGCT
	Probe 1 (VIC)	CCTTCCACCGGCATAC
	Probe 2 (FAM)	CCTTCCACCAGCATAAC
7:31829522_T/C	Forward	AAACGGGTAAGCTGAGCATTATCTT
	Reverse	GCCAACTAATGTAAACCACCACCAT
	Probe 1 (VIC)	CTCACATCCCATGGAAG
	Probe 2 (FAM)	CTCACATCCCGTGAAG
8:34814345_G/A	Forward	ATCTCCACAACCACATCTTGAAACT
	Reverse	GCGTGTAGAGAGCTGCTTCTT
	Probe 1 (VIC)	TCCAAGAGAAGGATCTAG
	Probe 2 (FAM)	CCAAGAGAAAGATCTAG
4:48575349_A/G	Forward	ACCAAAACCTGAATAATGCAGAGGAA
	Reverse	CGCCTCTGAATCTCGTGAGA
	Probe 1 (VIC)	TCCCAGAGAGTCTCG
	Probe 2 (FAM)	CCAGAGGGTCTCG
7:46758784_C/A	Forward	CTGCTTTGATCCTATCTGCTCCAA
	Reverse	CCCATGTTTTTGATGTTTGATCCAT
	Probe 1 (VIC)	TGTGTTTGGTTTGTGGTTGT
	Probe 2 (FAM)	TGTGTTTGGTTTTTGGTTGT

<sup>1</sup> Assay Name 依「染色體:位置\_參考等位基因/變異等位基因」格式命名。<sup>1</sup> Assay Name follows the format "chromosome:position\_reference/alternative allele".



圖六、本研究開發之 SNP 套組於驗證族群之分析結果：(a) 2:47705934\_A/G, (b) 5:19077016\_T/G, (c) 6:22509740\_T/A, (d) 4:64210791\_G/A, (e) 7:31829522\_T/C, (f) 8:34814345\_G/A, (g) 4:48575349\_A/G, (h) 7:46758784\_C/A。

Fig. 6. Allelic discrimination plots of TaqMan SNP genotyping assay developed in this study. (a) 2:47705934\_A/G, (b) 5:19077016\_T/G, (c) 6:22509740\_T/A, (d) 4:64210791\_G/A, (e) 7:31829522\_T/C, (f) 8:34814345\_G/A, (g) 4:48575349\_A/G, (h) 7:46758784\_C/A.

## 結 論

本研究利用 ddRADseq 技術進行甘藍的基因體分析，建立 SNP 分析流程，為甘藍育種提供重要的分子標誌。針對耐熱、抗黃葉病及園藝性狀優良的甘藍自交系、抗病育種材料及雜交一代品種等 112 份育種材料進行 ddRADseq，辨識出 745,196 個變異位點，其中包括 648,810 個 SNP 位點。進一步的遺傳分析呈現育種材料間的遺傳異質性與歧異性資訊，可作為自交世代推進與雜交組合選定之參考。依據篩選條件與基因型評估結果，開發了一套包含 8 個 SNP 的分子標誌分析套組，並驗證其在品系鑑別上的穩定性和可靠性。本研究為甘藍的育種提供了豐富的分子資源和分析工具，有望加速甘藍新品種的培育過程，提高其耐熱性、抗病性等重要性狀。研究使用之限制酶組合與操作分析流程亦可為其他蕓苔屬作物研究之參考。

## 參考文獻

1. 行政院農業委員會農糧署。2021。農業統計年報。  
<https://agrstat.coa.gov.tw/sdweb/public/trade/tradereport.aspx>. Accessed March 30, 2021.
2. 林延諭、吳靜霞、蕭政弘。2021。以基因體重定序開發甘藍雜交種子純度檢定用之 SNP 分子標誌。臺中區農業改良場研究彙報，150，1-12。
3. 蕭政弘。2012。臺灣中部地區外銷作物產業專集-甘藍。臺中區農業改良場特刊，122，77-75。
4. 蕭政弘。2013。甘藍新品種臺中 2 號育成。臺中區農業改良場 102 年度科技計畫研究成果發表會論文集，123，67-72。
5. Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., Johnson, E.A. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One*. 3(10), e3376.
6. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., Li, H. 2021. Twelve years of SAMtools and BCFtools. *Gigascience*. 10(2).
7. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., Del Angel, G., Rivas, M.A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K. Gabriel, S. B., Altshuler, D., Daly, M. J. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43(5), 491-498.
8. Fellers, J.P. 2008. Genome filtering using methylation-sensitive restriction enzymes with six base pair recognition sites. *The Plant Genome*. 1(2), 146-152.
9. Laila, R., Park, J.I., Robin, A.H.K., Natarajan, S., Vijayakumar, H., Shirasawa, K., Isobe, S., Kim, H. T., Nou, I. S. 2019. Mapping of a novel clubroot resistance QTL using ddRAD-seq in Chinese

- cabbage (*Brassica rapa L.*). BMC Plant Biology. 19, 1–9.
10. Langmead, B., Salzberg, S.L. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods. 9(4), 357-359.
  11. Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., Hoekstra, H.E. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One. 7(5), e37135.
  12. Poland, J.A., Brown, P.J., Sorrells, M.E., Jannink, J. L. 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One. 7(2), e32253.
  13. Schubert, M., Lindgreen, S., Orlando, L. 2016. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res. Notes. 9, 1-7.
  14. Sudan, J., Singh, R., Sharma, S., Salgotra, R.K., Sharma, V., Singh, G., Sharma, I., Gupta, S.K., Zargar, S.M. 2019. ddRAD Sequencing-Based Identification of Inter-Genepool SNPs and Association Analysis in *Brassica juncea*. BMC Plant Biology. 19(1), 594.

## Developing an SNP Assay Set for *Brassica oleracea* Using ddRADseq<sup>1</sup>

Yen-Yu Lin<sup>2\*</sup>, Ceng-Hung Hsiao<sup>3</sup>, Qun-Shan Wang<sup>4</sup> and Hsin-Yi Chang<sup>4</sup>

### ABSTRACT

This study focuses on *Brassica oleracea*, a major vegetable in Taiwan, using the ddRADseq technique, a next-generation sequencing method, for genomic analysis. Through the sequencing and analysis of 112 breeding materials, a total of 745,196 variant loci were successfully identified, including 648,810 SNP loci. After rigorous filtering, 3,565 high-quality SNP loci were retained. By utilizing the genotypic values of these SNPs, genetic heterogeneity and diversity data for *Brassica oleracea* breeding materials were constructed, providing rich reference information for material management and breeding. Eight sets of synthetic analysis kits were selected from these SNPs, capable of distinguishing 109 different varieties, and their utility was confirmed in 24 validation materials. The analysis kits demonstrated high stability and reliability within breeding materials, serving as a powerful tool for breeding efforts. This study not only established a molecular marker kit to assist *Brassica oleracea* breeding but also introduced an economical and rapid ddRADseq experiment and analysis pipeline for obtaining genotypic values of a large number of samples. This pipeline can serve as a beneficial tool for molecular-assisted breeding and gene localization in *Brassica oleracea*.

**Key words:** cabbage, SNP, ddRAD

---

<sup>1</sup>Contribution No.1087 from Taichung DARES, MOA.

<sup>2</sup>Former Assistant Researcher, Taichung DARES, MOA. Currently Assistant Researcher, Taiwan Agricultural Research Institute, MOA.

<sup>3</sup> Researcher and Deputy Director of Taichung DARES, MOA.

<sup>4</sup> Research Assistant, Department of Agronomy, National Taiwan University.

\*Corresponding author email: ylin@tari.gov.tw