

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¹

林延諭²、吳靜霞³、蕭政弘²

摘 要

單一核苷酸多型性(single nucleotide polymorphism, SNP)分子標誌因於基因組中數量繁多，且具有高再現性及可高度自動化分析的特性，為育種研究中重要的分子標誌。為建立重要經濟蔬菜作物甘藍(*Brassica oleracea* L. var. *capitata*)之 SNP 標誌開發流程，本研究以臺中區農業改良場於 102 年育成之耐熱性與結球性極佳的一代雜交品種甘藍‘台中 2 號’及其兩自交系親本為材料，進行基因體重定序，共計獲得 6,449,309 個 SNP 位點與 823,898 個 InDel。由 SNP 位點設計合成 28 組 Allele-Specific PCR 引子組、23 組切割擴增多型性序列(CAPS)標誌與 11 組 TaqMan SNP 分析套組，分別獲得 8、9 及 9 組可正確辨別甘藍‘台中 2 號’及其親本之基因型，顯示本試驗流程可快速開發雜交種子純度檢測標誌。豐富的序列資料與多樣的標誌類型，未來則可進一步應用於連鎖圖譜建立、重要性狀定位及分子標誌輔助選種。

關鍵詞：甘藍、單一核苷酸多型性

前 言

十字花科雲苔屬之甘藍物種(*Brassica oleracea*)中包含俗稱高麗菜之甘藍(var. *capitata*)、青花菜(var. *italica*)、花椰菜(var. *botrytis*)、芥藍(var. *alboglabra*)、羽衣甘藍(var. *acephala*)、球莖甘藍(var. *caulorapa*)及孢子甘藍(var. *gemmifera*)等 7 個重要亞種⁽¹³⁾，其中甘藍為臺灣最大宗的葉菜類蔬菜，年栽培面積約 8,000 公頃⁽²⁾。在臺灣甘藍可周年生產，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⁽³⁾。為此，臺中區農業改良場(以下簡稱本場)引入新的耐熱種質資源，於 102 年育成耐熱性與結球性極佳的一代雜交品種‘台中 2 號’⁽⁴⁾。

甘藍為典型異交作物，生產雜交種子時，利用孢子型的自交不親和及細胞質雄不稔系統，做為維持種子純度的主要策略，然而這兩種方法皆容易受到環境因素影響，使得混入雜交失敗的種子成為必然⁽¹³⁾。故如何有效率地控制雜交種子純度，為種苗產業的發展關鍵⁽¹⁶⁾。傳統上，採生長特性

¹ 臺中區農業改良場第 1001 號研究報告。

² 臺中區農業改良場助理研究員、研究員兼秘書。

³ 農糧署北區分署新竹辦事處課員。

檢定(grow-out-trials, GOTs)評估雜交種子的純度⁽⁵⁾，但因涉及不同生育階段的多個性狀調查，所需時間較長，對於投入的人力及檢定空間要求較高，同時也容易受到環境因素影響檢測結果⁽¹⁵⁾。

因具有較高的準確度及於時間、成本上的優勢，近 20 年來，各種分子標誌如限制片段長度多型性(restriction fragment length polymorphism, RFLP)、隨機增幅多型性 DNA(random-amplified polymorphic DNA, RAPD)、增幅片段長度多型性(amplified fragment length polymorphism, AFLP)簡單重複序列(simple sequence repeat, SSR)、相關序列增幅多型性(sequence-related amplified polymorphism, SRAP)及 SNP 等，在如水稻⁽⁹⁾、甘藍⁽¹⁴⁾、番茄⁽⁶⁾及甜瓜⁽¹²⁾等作物之品種鑑別及種子純度檢測上廣泛的使用。其中，不需要參考序列之 SSR 標誌因具有高再現性及高多型性等特性，使其使用較為普及⁽⁵⁾。近年，隨次世代定序的發展，同樣具有高再現性 SNP 標誌逐漸展現其發展優勢⁽⁸⁾，作為最常見且數量最豐富的遺傳變異類型，使其在相同物種的不同個體中，具有較 SSR 豐富之多型性⁽¹⁰⁾。因此，本研究以甘藍‘台中 2 號’及其親本為材料，進行基因體重定序，以快速辨識在親本間具有多型性之 SNP 位點，並開發 AS-PCR 引子組、TaqMan SNP 螢光標定引子組及 CAPS 標誌，開發之分子標誌及序列資料，即可直接應用於種子純度鑑定，未來也能應用於連鎖圖譜建立及分子標誌輔助選種。

材料與方法

一、植物材料

本研究以甘藍‘台中 2 號’與其親本自交系 T2 母本及 T3 父本為定序材料。另以 T2 及 T3 單株分別進行自交及雜交，收取種子後播種於穴苗盤，採其嫩葉 0.05 g 萃取 DNA 作為驗證材料。

二、植物核酸萃取與全基因組重定序

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 μ 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，接著以分光光度計測定 DNA 濃度，儲存於-20 $^{\circ}$ C。全基因組重定序委託有勁公司(有勁生物科技股份有限公司, 臺灣)以 Illumina HiSeq4000 進行 150 bp 雙邊定序。

三、SNP 與 InDel 探勘

定序資料以 atropos(v1.1.18)移除轉接子序列，再以 Seqtk 之 trimfq 依誤差閾值 0.05、最高不確定鹼基(-N)為 2 及最小序列長度(-L)為 35 等條件，進行品質修剪。通過品質修剪之序列先後以 BWA 根據 Ensembl 資料庫中 *Brassica oleracea* 之參考基因組序列(release-42)進行比對(reads mapping)，以

GATK 進行變異點偵測(variant calling)，產生之 VCF(Variant Call Format)檔案再以 Perl script 進行資料解析，並於R統計軟體中分析及篩選。SNP 篩選條件包含變異點層次(site-level)及樣品層次(sample-level)篩選，其篩選條件包含 QUAL(Phred-scaled probability)分數(1,000 以上)、甘藍‘台中 2 號’及其親本之基因型(GT)、定序深度(DP)及基因型品質(GQ)等。並以本次定序發現之變異點為基礎資料，標注篩選保留之 SNP 位點前後 50 bp 內其他變異點之數量，作為後續挑選依據，並將其 SNP 位點前後 600 bp 序列，以 NCBI BLAST 搜尋參考序列中相似序列之數量，作為引子專一性篩選條件。

四、分子標誌開發與樣品檢測

(一)Allele-Specific PCR(AS-PCR)引子設計：每一組 AS PCR marker(for SNP)包含 4 條引子，以 3'端有不同 SNP 核苷酸的 allele specific primer(AS primer)與對應的反向引子為基礎，並於 AS primer 5'端外側加上一條 AS 同向引子避免錯誤增幅並作為 PCR 正對照(flanking primers)，單次反應包含 3 條引子，增幅出 1 條 SNP 專一性條帶與 1 條 SNP flanking 條帶；另一對偶基因型則由另一條 AS primer 與相同之 flanking primers 組成。本研究中以 BatchPrimer3(<http://batchprimer3.bioinformatics.ucdavis.edu/>)設計 Allele-specific primers 與 allele-flanking primers，設定 primer size 15-30 mer、primer Tm 57-63°C、flanking product size 300 bp 及 3'端採 mismatch 設計，並委託源資公司(源資國際生物科技股份有限公司，臺灣)合成。

(二)CAPS 標誌設計：以 Perl script 分析具有 *EcoRI* 切位，且在親本間具有多型性之位點，並以 BatchPrimer3 設計 SNP flanking primers，預期產物大小設為 1,500 bp，進一步挑選 *EcoRI* 酶切後，親本間產物大小差異為 400 bp 以上之引子組，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)合成。

(三)TaqMan 分析套組建立：經篩選之 SNP 位點前後 2,500 bp 序列於 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。

(四)基因型分析：

1. Allele-Specific 引子組及 CAPS 標誌之聚合酶連鎖反應(polymerase chain reaction, PCR)樣品總體積為 15 μ L，內含 DNA 模版、5X Taq Master Mix(Fast-Run Taq Master Mix Kit，波仕特公司)、10 μ M 引子組及無菌蒸餾水。樣品混合後於 GeneAmp PCR System 9700 (Applied Biosystem, USA)進行聚合酶連鎖反應，先以 95°C 反應 2 min，接著依序以 95°C 反應 1 min、52°C 反應 40 sec 及 72°C 反應 40 sec，上述三步驟共重覆 35 個循環，最後以 72°C 反應 7 min。經聚合酶連鎖反應所增幅之產物，加入 6x loading dye，以 2%-3.5% 的 SFR agarose 在 TBE buffer 中，於 100 V 電壓進行電泳分離約 60 min，並以藍光檢視、照相及貯存影像。

2. TaqMan 分析套組之即時聚合酶連鎖反應(real-time PCR)，反應體積為 10 μ l，包含 2.5 μ l ddH₂O、5 μ l TaqMan GTXpress Master Mix(Thermo Fisher Scientific, USA)、0.5 μ l TaqMan SNP Genotyping Assay 20X(Thermo Fisher Scientific, USA)及 2 μ l template DNA，樣品混合後以即時聚合酶連鎖反應儀(QuantStudio 3 Real-Time PCR Systems, Thermo Fisher)進行反應，反應先以 60°C 30 秒進行螢光背景值預讀，接著以 95°C 20 秒活化聚合酶、循環階段為 95°C 1 秒及 60°C 20 秒共 40 個循環，最終以 60°C 30 秒進行反應結束，於每次循環及反應結束前偵測螢光值。

結果與討論

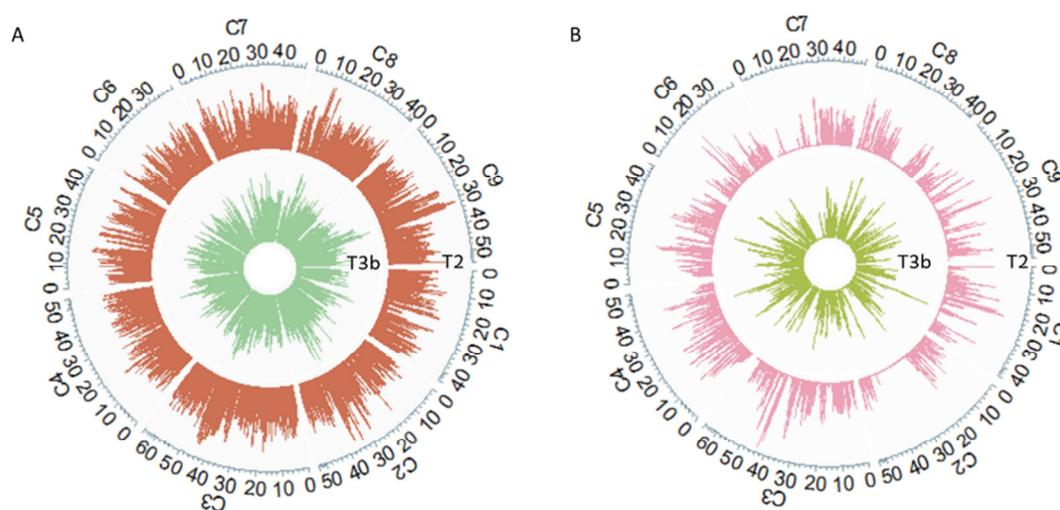
一、序列處理與 SNP 辨識

甘藍‘台中 2 號’及其母本 T2 與父本 T3 經 Illumina HiSeq 4000 150PE 定序後，分別獲得約 65.9 M、75.8 M 及 68.1 M 對序列，經 atropos v1.1.18 及 Seqtk v1.2-r94 軟體進行品質修剪後，以 BWA 對 *Brassica oleracea* 參考基因組序列進行比對，平均可成功比對之序列約占 97.8%(表 1)，在相近的定序資源下，高於 Shea 等學者(2018)於結球白菜(*Brassica rapa* var. *pekinensis*)之比對結果(61.1-63.4%)，與 Lee 等學者(2015)同為甘藍但定序量較低之比對結果相近(77.6-82.1%)。以 GATK 進行變異點辨識，共有 7.5 M 個變異點，其中 SNP 位點占 86.0%有 6,449,309 筆，InDel 占 11.0%有 823,898 筆，其他類型占 3.3%有 226,887 筆。為提高資料正確性，同時保留足夠數量資料供後續篩選，保留 QUAL 分數大於 1,000 之 3,763,587 筆 SNP，各染色體上 SNP 數量以第 3 條染色體 600,977 筆為最高，以第 6 條染色體的 313,262 筆為最低，與 Lee 等學者(2015)結果相同。兩親本在各染色體上之 SNP 數量相近，然其位點分布略有不同，其中 T2 上總計有 2,550,666 筆，在 T3 中共有 2,272,075 筆(圖 1)。基因型(GT)篩選部分，鹼基與參考序列不同(alternate)，且另一親與參考序列相同(reference)之 SNP 位點，T2 共有 586,387 筆，T3 有 458,879 筆，兩者合併，同時甘藍‘台中 2 號’基因型為異質結合者，即為具有多型性之 SNP 位點，總計有 994,433 筆。因 SNP 數量充足，進一步根據定序深度(DP)及基因型品質(GQ)篩選，在 SNP 位點上之 T2 之平均定序深度為 30.5，T3 為 27.9(圖 3)。為減少定序或序列比對錯誤產生之 SNP，因此將樣品 DP 小於 20 之 SNP 剔除，保留其中 209,377 筆 SNP。最後擷取 CHROM 及 POS 標籤，將 SNP 位點前後 50 bp 內有其他變異點者刪除，以提高引子設計之成功率，總計獲得 2,510 筆之 SNP，可供後續試驗分析使用。

表一、甘藍品種台中 2 號及其兩雜交親本之全基因體定序結果

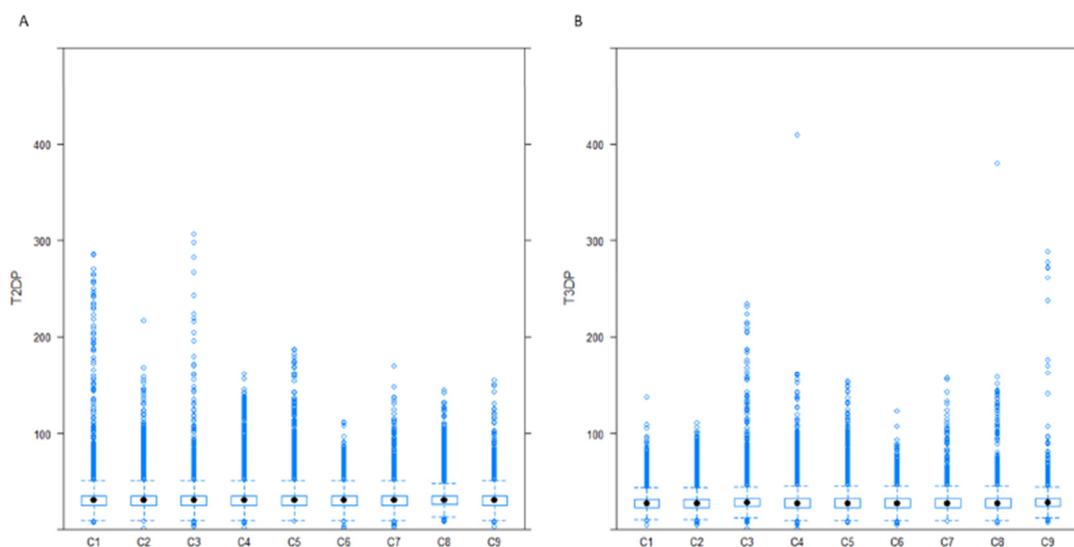
Table 1. Sequencing and mapping summary of Taichung No. 2 and its parental lines

Sample	F ₁ -Taichung No. 2	♀-T2	♂-T3
Raw total sequences:	131,757,962	151,462,288	135,929,524
Reads mapped:	128,936,417	148,187,999	132,965,713
Reads unmapped:	2,821,545	3,274,289	2,963,811
Mapping rate:	97.86%	97.84%	97.82%
Reads paired:	131,757,962	151,462,288	135,929,524
Reads MQ0:	17,429,476	19,416,070	17,664,568
Average length:	149	149	149
Maximum length:	150	150	150
Average quality:	37.9	37.5	37.7



圖一、甘藍品種台中 2 號兩親本間之 SNP 密度圖。(A)‘台中 2 號’父本 T3 及母本 T2 SNP 密度圖，密度圖以 100 kb 為窗口沿染色體滑動繪製。(B)‘台中 2 號’父本(T3)及母本(T2)獨特之 SNP 密度圖，密度圖以 100 kb 為窗口沿染色體滑動繪製。兩圖外圈為相應之染色體位置，單位為 Mb。

Fig. 1. SNP density plots of the parental lines of Taichung No. 2.(A)The SNP density of T3 and T2, in 100 kb sliding windows across the genome.(B)The density of unique SNPs found in T3 and T2, in 100 kb sliding windows across the genome. In both plots, the outer circle shows the chromosomes as blue bars, with length markers reported in Mb.

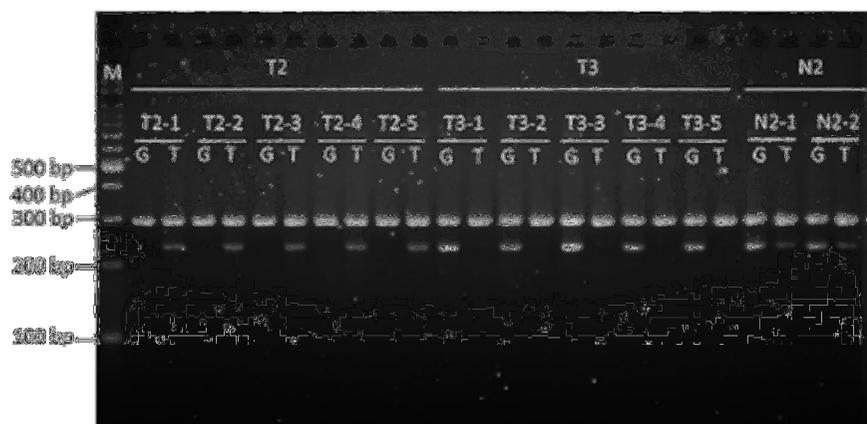


圖二、在甘藍‘台中 2 號’親本間具特異性之 SNP 位點定序深度。(A)甘藍‘台中 2 號’母本 T2、(B)甘藍‘台中 2 號’父本 T3。

Fig. 2. Box plots of SNPs DP(depth of coverage).(A)DP of unique SNPs found in T2.(B)DP of unique SNPs found in T3.

二、SNP 標誌與 CAPS 標誌之開發

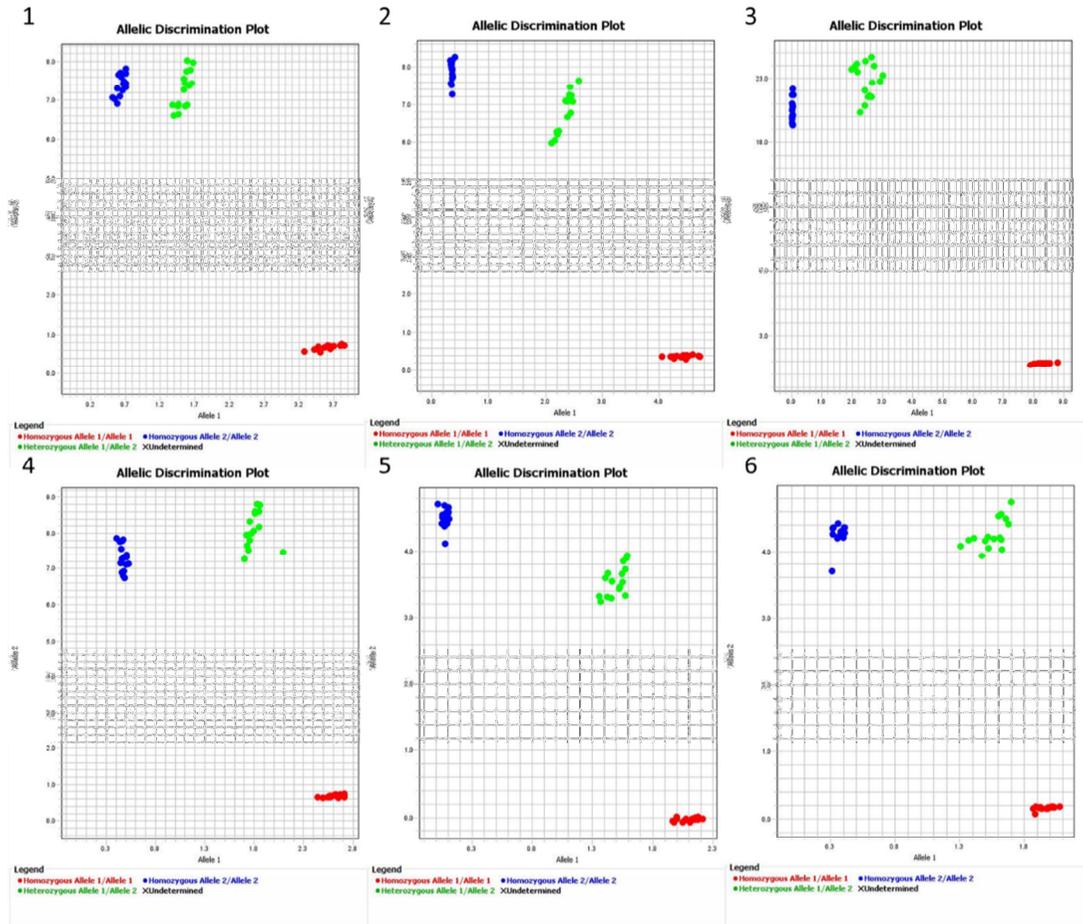
本試驗共合成 28 組 AS-PCR 引子組，其中有 8 組擴增結果符合預期(圖 3)，設計成功率為 28.6%，4 組引子因 AS 條帶與 flanking 條帶大小接近，不易判讀，其餘引子則有 AS 條帶訊號微弱或無法判讀之現象，然 flanking 條帶則皆可順利擴增，推測可能因 3'mismatch 設計使 AS 引子競爭優勢下降導致，藉由降低黏合溫度則可獲得較佳之 PCR 結果。在 CAPS 標誌開發部分，由通過篩選條件的 SNP 前後序列比對獲得 197 筆可被 *EcoRI* 酵素酶切之位點，以目標產物大小為 1,500 bp 設計引子，同時為提高電泳結果判讀效率，由酶切前後產物片段差異大於 400 bp 之引子組共 92 筆中，挑選 23 組進行合成，其中有 9 組擴增產物及 *EcoRI* 酶切片段大小符合預期，可用於雜交成功率檢定，標誌開發成功率為 39.1%。



圖三、AS-PCR 引子組 TDC801，檢測‘台中 2 號’(N2)、母本(T2)及父本(T3)樣品之結果，其中 T2 顯示為同質結合 T/T 基因型、T3 為同質結合之 G/G 基因型，N2 為異質結合體 G/T。
 Fig. 3. Analysis of T2, T3 and Taichung No.2 individuals by AS-PCR marker TDC801. The result showed all T2 individuals are homozygous(TT), T3 are homozygous(GG), and N2 are heterozygous(GT).

在 TaqMan 螢光標定引子組部分，以本場甘藍‘台中 2 號’及其父母本，共 48 個單株為驗證材料，驗證 11 組分析套組，結果有 9 組結果符合預期(圖 4)，而 Bo3g065600p3 及 Bo3g065600p7 雖可區分兩親本之基因型，但異質結合體與其中一個親本之基因型混雜，無法用於雜交成功率檢定，設計成功率為 81.8%。

如何有效率的管控雜交種子純度，為雜交種子生產過程中的關鍵環節⁽¹⁶⁾。在確認雜交種子純度時，除生長特性檢定外，以 DNA 分子標誌進行基因型鑑定的方法已被廣泛的接受。在臺灣多數的種苗業者仍以生長特性檢定控管雜交種子純度，部分使用分子標誌檢定之種苗公司仍以 SSR、SCAR 與 InDel 標誌為主要工具⁽¹⁾，然而這些標誌系統仍需以電泳方式進行結果判讀，不易導入自動化分析系統。另外，若使用之標誌非開發自目標品種(系)，仍需花費大量時間與資源進行引子擴增與多型性等驗證，如 Ye 等學者(2013)針對甘藍雜交品種‘Sugan 21’篩選 325 組引子，最終僅 7 組引子可用於雜交成功率檢定。本研究以目標品種甘藍‘台中 2 號’及其親本進行重定序，以 81.8% 的成功率開發可高度自動化分析的 TaqMan 分析套組，可直接投入於雜交種子純度檢定；充足的候選 SNP 位點與 AS-PCR 及 CAPS 標誌的成功開發，顯示本研究結果也可滿足有成本考量或建立連鎖圖譜、重要性狀定位及分子標誌輔助選種等不同之試驗需求。



圖四、本研究開發之 SNP 分析套組於 48 個甘藍‘臺中 2 號’及其親本之分析結果。

1: Bo1g025720p1、2: Bo1g025720p3、3: Bo3g065600p4、4: TDC501、5: TDC607、6: TDC701。

Fig. 4. Allelic discrimination plots of TaqMan SNP genotyping assay developed in this study

1: Bo1g025720p1、2: Bo1g025720p3、3: Bo3g065600p4、4: TDC501、5: TDC607、6: TDC701。

結 論

甘藍‘台中 2 號’及其親本經全基因組定序後，共計獲得 6,449,309 個 SNP 變異點，以 QUAL、基因型(GT)、定序深度(DP)、基因型品質(GQ)及鄰近變異點等條件篩選後，共有 2,510 筆 SNP 可供後續設計為 AS-PCR、TaqMan 分析套組或 CAPS 標誌使用。本試驗共計合成 11 組 TaqMan SNP 分析套組，以 48 個樣品進行檢測，結果顯示有 9 組可順利擴增，並正確鑑別基因型，成功率為 81.8%，開發之分析套組可直接應用於雜交一代種子之純度檢定，並大幅縮短分析時間。本研究建立之序列資料與分析流程，未來可用於其他自交系之 SNP 開發及供甘藍基因體研究使用。

參考文獻

1. 王聖善、朱詠筑、蔡佩樺、楊藹華 2019 利用 SNP 分子標誌建立胡瓜雜交種子純度檢測技術 臺南區農業改良場研究彙報 74: 27-35。
2. 行政院農業委員會統計資料庫 <https://agrstat.coa.gov.tw/sdweb/public/trade/tradereport.aspx> (下載日期：110.3.30)。
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. Lee, J., N.K. Izzah, M. Jayakodi, S. Perumal, H.J. Joh, H.J. Lee, S.-C. Lee, J.Y. Park, K.-W. Yang, I.-S. Nou, J. Seo, J. Yoo, Y. Suh, J. H. Lee, G. J. Choi, Y. Yu, H. Kim and T.-J. Yang. 2015. Genome-wide SNP identification and QTL mapping for black rot resistance in cabbage. *BMC Plant Biol.* 15: 32.
6. Liu Liwang, X. Hou, Y. Gong, Y. Zhang, K. Wang. and J. Zheng. 2004. Application of molecular marker in variety identification and purity testing in vegetable crops. *Mol. Plant Breed.* 2: 563-568.
7. Liu, L.-W., Y. Wang, Y.-Q. Gong, T.-M. Zhao, G. Liu, X.-Y. Li and F.-M. Yu. 2007. Assessment of genetic purity of tomato(*Lycopersicon esculentum* L.)hybrid using molecular markers. *Sci. Hortic.* 115: 7-12.
8. Mammadov, J., R. Aggarwal, R. Buyyarapu and S. Kumpatla. 2012. SNP markers and their impact on plant breeding. *Int. J. Plant Genomics* 2012: 1-11.
9. Moorthy, K.K., P. Babu, M. Sreedhar, V.S.A.K. Sama, P.N. Kumar, S.M. Balachandran and R.M. Sundaram. 2011. Identification of informative EST-SSR markers capable of distinguishing popular Indian rice varieties and their utilization in seed genetic purity assessments. *Seed Sci. Technol.* 39: 282-292.
10. Rafalski, J.A. 2002. Novel genetic mapping tools in plants: SNPs and LD-based approaches. *Plant Sci.* 162: 329-333.
11. Shea, D.J., M. Shimizu, E. Itabashi, N. Miyaji, J. Miyazaki, K. Osabe, M. Kaji, K. Okazaki and R. Fujimoto. 2018. Genome re-sequencing, SNP analysis, and genetic mapping of the parental lines of a commercial F1 hybrid cultivar of Chinese cabbage. *Breed. Sci.* 68: 375-380.
12. Silberstein, L., I. Kovalski, R. Huang, K. Anagnostou, M.M. Kyle Jahn and R. Perl-Treves. 1999. Molecular variation in melon(*Cucumis melo* L.)as revealed by RFLP and RAPD markers. *Sci. Hortic.* 79: 101-111.
13. Sorajjapinun, W. and P. Srinives. 2011. Chasmogamous mutant, a novel character enabling commercial hybrid seed production in mungbean. *Euphytica* 181: 217-222.

14. Yan, C., Y. Huang, Z. Liu, F. Guo, Z. Jiao, W. Yang, F. Zhu and Z. Qiu. 2020. Rapid identification of yellow-flowered gene *Bofc* in cauliflower (*Brassica oleracea* var. *botrytis*) by bulked segregant analysis and whole-genome resequencing. *Euphytica* 216: 26.
15. Ye, S., Y. Wang, D. Huang, J. Li, Y. Gong, L. Xu and L. Liu. 2013. Genetic purity testing of F1 hybrid seed with molecular markers in cabbage (*Brassica oleracea* var. *capitata*). *Sci. Hortic.* 155: 92-96.
16. Zhiyuan, F., X. Wang, Q. Dongyu and L. Guangshu. 2000. Hybrid seed production in cabbage. *J. New Seeds* 1: 109-129.

SNP Identification by Genome Re-sequencing for Genetic Purity Testing of F₁ Hybrid Seed in Cabbage¹

Yen-Yu Lin ², Ching-Hsia Wu ³ and Ceng-Hung Hsiao ²

ABSTRACT

Single nucleotide polymorphism(SNP)marker is an important type of DNA markers because it is highly abundant, highly reproducible and amenable to automation. To establish the SNPs development process, we performed genome re-sequencing and analysis of a heat-resistant F₁ hybrid cabbage(*Brassica oleracea* L. var. *capitata*)cultivar ‘Taichung No. 2’ and its two parental lines. A total of 6,449,309 SNPs and 823,898 InDel were identified. Markers of different systems were designed from the parent-unique SNPs and were validated using 48 individuals. As a result, 8 out of 28 allele-specific PCR for SNP, 9 out of 23 cleaved amplified polymorphic sequence(CAPS)markers and 9 out of 11 TaqMan SNP genotyping assays could be used in quality control of the cabbage hybrid seed. The results and data are useful for future studies on cabbage breeding.

Key words: cabbage, SNP

¹ Contribution No.1001 from Taichung DARES, COA.

² Assistant Researcher, Researcher and Senior Secretary of Taichung DARES, COA.

³ Officer of Northern Region Branch, Agriculture and Food Agency.