

# 利用小量次世代基因體定序開發長壽花 及聖誕紅微衛星標誌之研究<sup>1</sup>

葉志新<sup>2</sup>、廖芳心<sup>2</sup>

## 摘 要

微衛星分子標誌廣泛使用在 DNA 指紋分析、品種鑑定、遺傳歧異度分析、親緣關係及遺傳圖譜之研究，但是在新的物種開發 SSR 分子標誌是昂貴且費時的工作。現在，次世代定序技術提供了一個新的選擇，對沒有序列資訊的作物也可簡易地開發 SSR 分子標誌。在本試驗中，以 Roche/454 定序為基礎，開發 SSR 分子標誌，將長壽花及聖誕紅 DNA 樣品分別接 MID 後混合，完成 1/4 個 PicoTiterPlate 之定序，共獲得 60.9 與 53.3 Mbp 鹼基對（180,648 與 135,085 條序列），含 SSR 基因座並於兩端可設計引子者有 732 條與 514 條。分別以 6 個長壽花及聖誕紅品種測試，結果有 33 及 28 組引子具有多型性，符合作為 SSR 分子標誌者，佔所設計引子 50% 以上。相較於傳統選殖及定序方式，Roche/454 是一個便宜、快速而且可以大量取得微衛星分子標誌的技術。

關鍵詞：簡單重複序列、次世代定序、Roche/454 焦磷酸定序、長壽花、聖誕紅

## 前 言

微衛星序列（microsatellites）又稱為簡單重複序列（simple sequence repeat, SSR）或短縱列重複序列（short tandem repeat, STR），是以 1-6 個鹼基對為重複單元（motif），頭尾串連排列的小片段 DNA。SSR 序列存在於真核生物及原核生物中（Field and Wills, 1996；Gur-Arie *et al.*, 2000；Katti *et al.*, 2001；Weber, 1990），不論在基因編碼區或非

---

<sup>1</sup> 行政院農業委員會桃園區農業改良場研究報告第 432 號。

<sup>2</sup> 桃園區農業改良場助理研究員（通訊作者，zeamays@tydais.gov.tw）及研究員。

編碼區含量都相當豐富，約 2-30 kb 會出現一個 SSR (Kelkar *et al.*, 2010; Webster *et al.*, 2002)，而基因體 SSR (genomic SSR) 相較於基因 SSR (gene SSR) 有較高的多型性 (Blair *et al.*, 2003; Ellison and Shaw, 2010)。SSR 有高度多型性原因目前未明，可能是 DNA 複製時特定 SSR 基因座的重複序列產生不均等的重組 (unequal crossing-over)，或是 DNA 的滑動配對 (slipped-strand mispairing, SSM) 造成重複序列數改變，因而在物種間或個體間產生多型性 (Levinson and Gutman, 1987; Li *et al.*, 2002; Schlotterer, 2000; Schlotterer and Tautz, 1992)。SSR 分子標誌分析原理，是在重複序列的兩側保守性序列設計專一性的引子對，以基因組 DNA 作為模板，進行聚合酵素連鎖反應及電泳分析。因其具有多對偶基因 (multiallelic)、共顯性遺傳 (co-dominance)、再現性高、廣泛分布於基因組中及分析容易等優點 (Aggarwal *et al.* 2007; Mittal and Dubey, 2009; Saha *et al.*, 2006)，並符合孟德爾遺傳定律，近年來被廣泛應用於 DNA 指紋分析、親本與品種鑑定、遺傳圖譜建構、族群遺傳及分子標誌輔助育種等研究 (Guichoux, 2011; Knapik *et al.*, 1998; Mba *et al.*, 2001; Schuler *et al.*, 1996)。

傳統開發 SSR 分子標誌的策略，包含由小片段插入基因組文庫 (small-insert genomic library) 選殖及定序、建構豐富性微衛星序列基因組文庫 (SSR-enriched genomic library) 選殖及定序，或由基因序列資料庫中搜尋，另外由於 SSR 序列在物種演化中具有保守性，因此可利用序列比對方式，由已完成基因體定序之親緣相近物種開發兩者通用之 SSR 標誌 (Zane *et al.*, 2002)。次世代定序 (next generation sequencing, NGS) 又稱為高通量定序 (high throughput sequencing)，是以平行定序的方式，一次反應同步完成百萬條以上之定序，目前包括 Roche/454、Solexa/Illumina 及 AB SOLiD 等技術。其中 Solexa/Illumina 及 AB SOLiD 每次反應可完成  $160-500 \times 10^6$  條序列定序，每條序列長度為 50-100 bp；而 Roche/454 每次反應完成  $1 \times 10^6$  條序列，但長度可達 400-500 bp (Deschamps and Campbell, 2010)。其中由 Roche/454 序列直接搜尋 SSR，不需經過選殖及傳統定序程序，且單一序列長度已足夠於 SSR 重複單元兩側設計引子，只需進行 1/8 或 1/16 個 PicoTiterPlate 之反應即可獲得足夠之序列，並偵測出數百至數千個以上含 SSR 之序列，為一個發展 SSR 分子標誌快速又經濟的選擇 (Csencsics, 2010; Takayama, 2011)。

長壽花 (*Kalanchoe blossfeldiana* Poelln.) 及聖誕紅 (*Euphorbia pulcherrima* Willd ex Koltz) 是國內外重要盆花作物，國內年產量約 50 萬盆及 110-150 萬盆，目前已有

68 個聖誕紅及 24 個長壽花品種申請植物品種權（農糧署全球資訊網，2012）。而開發 SSR 分子標誌必需要有充足的序列資訊，新的物種在第一次發展 SSR 分子標誌時，通常需要投入大量經費來定序。目前長壽花及聖誕紅在基因資料庫中所公開的序列不多，因此，本研究以 Roche/454 高通量定序，開發長壽花及聖誕紅之 SSR 分子標誌，評估其作為品種鑑定之輔助工具，及應用於遺傳圖譜建構及族群遺傳等後續研究之可行性。

## 材料方法

### 一、DNA 萃取與 Roche/454 定序

分別取長壽花品種‘Calandiva Cher’及聖誕紅品種‘TYNY No.1’葉片 0.1 g，以 Qiagen 之 Mini Kit 植物基因組萃取套組，萃取基因組 DNA 10 µg，委送榮陽基因體研究中心（VYM Genome Research Center）進行 Roche/454 焦磷酸定序（Roche/454 pyrosequencing）。定序分析使用多樣品混合方式（multiplexing option），分別接上 11 個鹼基之 MID 標籤（Multiplex Identifier adaptors），長壽花之 MID 為 5'-ACACTACTCGT-3'，聖誕紅之 MID 為 5'-ACACGACGACT-3' 以 Roche/454 Genome Sequencer 之 FLX Titanium 套組，進行 1/4 的 70×75 PicoTiterPlate（PTP）反應（Roche Applied Science, Penzburg, Germany）。

### 二、序列分析及 SSR 引子設計

將 Roche/454 產出之短片段序列以 CLC Genomics Workbench 4 軟體（CLC bio-AsiaPac）分析及組裝，並以 SSRLocator 軟體（Maia *et al.*, 2008）篩選含 SSR 之序列，篩選原則以重複單元為單核苷酸（mono-nucleotide）者重複 12 次、雙核苷酸（di-nucleotide）者重複 8 次以上、三核苷酸（tri-nucleotide）者重複 5 次以上、四、五及六核苷酸（tetra, penta and hexa -nucleotide）者重複 4 次以上，並由 SSRLocator 軟體內建之 Primer3 模組設計引子。以 M13-tails 方式在正反兩條引子之 5'端分別加上 5'-ACGACGTTGTAAA-3'與 5'-CATTAAGTTCCCATTA-3'序列；另於 PCR 時需加入 tagF 與 tagR 兩股短引子，其序列分別為 5'-ACGACGTTGTAAA-3'與 5'-CATTAAGTTCCCATTA-3'，並將螢光（VIC, FAM, NED 及 PET）標定在 tagF 引子 5'端上（Hayden *et al.*, 2008; Schuelke, 2000）。

### 三、PCR 反應及資料分析

分別以長壽花及聖誕紅各 6 個品種進行多型性測試，長壽花品種為‘Calandiva Hayworth’、‘Taoyuan No.1’、‘NCHU No.1-Beacon’、‘Lea’、‘Taoyuan No.3’及 *K. garambiensis*；聖誕紅品種為‘Peterstar’、‘Red Splendor’、‘Sonora white’、‘Red velvet’、‘Red elf’及‘Dulce Rosa’。PCR 反應總體積為 10  $\mu$ L，內含 1X Bioline<sup>®</sup> ImmoBuffer (16 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>2</sub>, 0.01 % Tween-20, 100 mM Tris-HCl, pH 8.3)、20 mM MgCl<sub>2</sub>、2 mM dNTPs、0.8  $\mu$ M Universal 引子、0.4  $\mu$ M specific 引子、20 ng genomic DNA 及 0.25U Bioline<sup>®</sup> IMMOLASE<sup>™</sup> DNA Polymerase。PCR 反應條件為 95 $^{\circ}$ C 10 分鐘；92 $^{\circ}$ C 30 秒，63 $^{\circ}$ C 90 秒，72 $^{\circ}$ C 60 秒，循環 20 次；92 $^{\circ}$ C 15 秒，54 $^{\circ}$ C 30 秒，72 $^{\circ}$ C 60 秒，循環 40 次；72 $^{\circ}$ C 30 分鐘完成。反應之產物以 VIC:NED:6-FAM:PET=1:2:3:3 的比例混合，之後將綜合產物稀釋 10 倍（產物 18  $\mu$ l+水 162  $\mu$ l），委由基龍米克斯股份有限公司進行 ABI 3730 毛細管電泳分析，資料以 Gene Mapper<sup>®</sup> 4.0 (Applied Biosystem) 軟體分析。

## 結果與討論

經由 Roche/454 定序結果如表 1，長壽花獲得 180,648 條序列，總計 60,986,646 bp，平均長度為 337 bp，聖誕紅獲得 135,085 條序列，總計 53,321,795 bp，平均長度為 394 bp，Q40Plus Bases 分別為 86.4%及 83.9%。經由 CLC Genomics Workbench 4 軟體將 MID 及訊號品質低之序列去除後，進行序列比對及組裝，長壽花得到 78,977 條單一序列及 16,901 條組合序列，平均長度為 294 bp 及 420 bp，聖誕紅則有 57,700 條單一序列及 11,096 條組合序列，平均長度為 336 bp 及 435 bp。長壽花染色體數為  $2n=34$  及自然多倍體  $2n=68$ ，栽培種中普遍有多倍體化之情形 (Aida and Shibata, 2002)，基因體大小約為  $3\times 10^8$  bp 至  $1.4\times 10^9$  bp (Izumikawa *et al.*, 2008)，聖誕紅染色體數為  $2n=28$ ，基因體大小為 1.30 pg/1C 至 1.65 pg/1C (Galbraith *et al.*, 1983, The Royal Botanic Gardens, Kew, 2012)，約為  $1.27\times 10^9$  bp 至  $1.56\times 10^9$  bp (Doležel and Bartoš, 2005)。而在本試驗中定序所得之鹼基數僅佔長壽花與聖誕紅基因組中 1/6 至 1/30 的覆蓋率 (coverage)，將序列組裝成單一序列的意義不大，因此，在序列組裝時以高相似係數 (similarity coefficient) 設定，主要目的是排除重複被定序之序列。

表 1. 長壽花及聖誕紅 Roche/454 定序及組裝結果

Table 1. Samples and sequence data generated from the Roche/454 genome sequencer and assembling in *K. blossfeldiana* and *E. pulcherrima*.

	總定序條數	總鹼基數	單一序列及 組合序列數	平均單一序列及 組合序列長度
	No. of Sequence reads	Total bases	No. of singletons / contigs	Average length singletons / contigs
	no.	bp	no.	bp
長壽花 <i>K. blossfeldiana</i>	180,648	60,986,646	78,977/16,901	294/420
聖誕紅 <i>E. pulcherrima</i>	135,085	53,321,795	57,700/11,096	336/435

SSR 基因座的重複數以 5-40 次為主 (Selkoe and Toonen, 2006)，但有一些研究認為 4-5 次重複即為 SSR 基因座 (Brzyski, 2010; Culley, 2005; Pazouki *et al.*, 2010)，因此篩選 SSR 原則依據不同作者而異，但以重複單元重複 3-5 次以上者為多 (Abdelkrim *et al.*, 2009; Castoe *et al.*, 2010; Takayama *et al.*, 2011)。在本試驗中以單核苷酸重複 12 次，雙核苷酸重複 8 次以上，三核苷酸重複 5 次以上，四、五、六核苷酸重複 4 次以上為篩選原則，以期獲得較多對偶基因數之機會。SSR 基因座分析結果如表 2，長壽花含 SSR 基因座序列為 1,189 條 (佔全部序列之 1.24%)，可設計引子者 732 條；聖誕紅為 884 條 (佔全部序列之 1.28%)，可設計引子者 514 條。分析 SSR 重複單元之類型及數量結果如表 3，長壽花及聖誕紅含雙核苷酸 SSR 分別有 343 及 211 個，其中以 (AT)<sub>n</sub> 數量最多，三核苷酸 SSR 分別有 469 及 475 個，其中以 (ATT)<sub>n</sub> 及 (ACT)<sub>n</sub> 數量最多，四核苷酸 SSR 分別有 203 及 98 個，五核苷酸 SSR 分別有 102 及 84 個，六核苷酸 SSR 分別有 80 及 38 個。目前的研究發現在多種植物作物基因組中最常出現的雙核苷酸 SSR 以 (AG)<sub>n</sub> 及 (CT)<sub>n</sub> 為重複的單位，而三核苷酸 SSR 常見的為 (AAT)<sub>n</sub> 及 (ATT)<sub>n</sub> (Wang *et al.*, 1994)，與本試驗中兩個物種略有不同。分析長壽花及聖誕紅 SSR 重複次數如圖 1，重複 4-10 次者所占比例最高分別有 1066 及 750 個，重複 11-20 次者分別有 137 及 167 個，重複 21-30 次者分別有 12 及 14 個，重複 31-40 次者分別只有 1 及 2 個，而重複 41 次以上者長壽花有 11 個，而聖誕紅則有 2 個。

圖 2 為長壽花 Kb-AGCT-02 基因座之毛細管電泳圖，在 6 個測試品種中有 4 個對偶基因，一個好的 SSR 分子標誌除了對偶基因數量多，還要有加 A 效應（addition of adenine）完全及適度的跡痕條帶（stutter bands）等條件。跡痕條帶是協助判別是否為 SSR 分子標誌的重要訊號（Schlötterer, 2000；Schlötterer and Tautz, 1992），但要避免超過主要條帶之 40% 以免影響判讀。在本試驗中 SSR 分子標誌擴增條帶表現之範例如圖 3 所示，圖 3-A、圖 3-B 為表現良好的標誌，具有適度的跡痕條帶及加 A 效應完全，其中圖 3-A 為同質結合，圖 3-B 為異質結合，箭頭所指為跡痕條帶之波峰，星號所指為加 A 效應不完全之波峰。圖 3-C、圖 3-D 為跡痕條帶訊號過強，並於主要條帶兩側均出現，如與鄰近異質結合條帶重疊，將會造成判讀困難。圖 3-E 為多對偶基因（multiple allele）。圖 3-F 箭頭所指為雜訊，過多之雜訊將會影響判讀。圖 3-G 為加 A 效應不完全，兩個波峰高度相近，並相差 1 個鹼基。圖 3-H 為擴增條帶訊號微弱，難與雜訊區分，無法判讀。

表 2. 長壽花及聖誕紅 SSR 基因座分析結果

Table 2. Number of microsatellites loci identified in *K. blossfeldiana* and *E. pulcherrima*.

	含 SSR 基因座 序列數	可以設計引子 序列數	可擴增 序列數	具多型性 序列數
	No. of sequences containing SSR Loci	Sequences with amplifiable SSR Loci	No. of Amplified result	Polymorphism result
	no.	no.	no.	no.
長壽花 <i>K. blossfeldiana</i>	1,189	732	44/64	33
聖誕紅 <i>E. pulcherrima</i>	884	514	40/55	28

表 3. 長壽花及聖誕紅 SSR 重複單元之類型及數量

 Table 3. SSR type and number of repeat motifs in *K. blossfeldiana* and *E. pulcherrima*.

重複類型 Repeat type	重複單元 Repeat motif	<i>K. blossfeldiana</i>		<i>E. pulcherrima</i>	
		no.	total	no.	total
單核苷酸 mono-nucleotide	A (T)	28	29	21	29
	G (C)	1		8	
雙核苷酸 di-nucleotide	AT	264	343	133	211
	GC	0		1	
	AG/TC	48		45	
	AC/GT	31		32	
	AAC/TTG	18		56	
三核苷酸 tri-nucleotide	AAG/TTC	76	469	77	475
	AAT/TTA	99		136	
	ACC/TGG	35		23	
	ACG/AGC/CGT/CTG	69		16	
	ACT/ATC /AGT/ATG	96		130	
	AGG/TCC	58		32	
	CCG/GGC	18		5	
四核苷酸 tetra-nucleotide	AAAC/AAAG/AAAT/AACC/ AACG/AACT/AAGG/AAGT/ AATT/ACAG/ACAT/ACCC/ ACCG/ACCT/ACGC/ACGG/ ACGT/ACTC/ACTG/AGAT/ AGCG/AGCT/AGGG/CCCG		203		98
	五核苷酸 penta-nucleotide	AATTA/ACCGG/...		102	84
		六核苷酸 hexa-nucleotide	ACACAA/TGTTTG/...		80

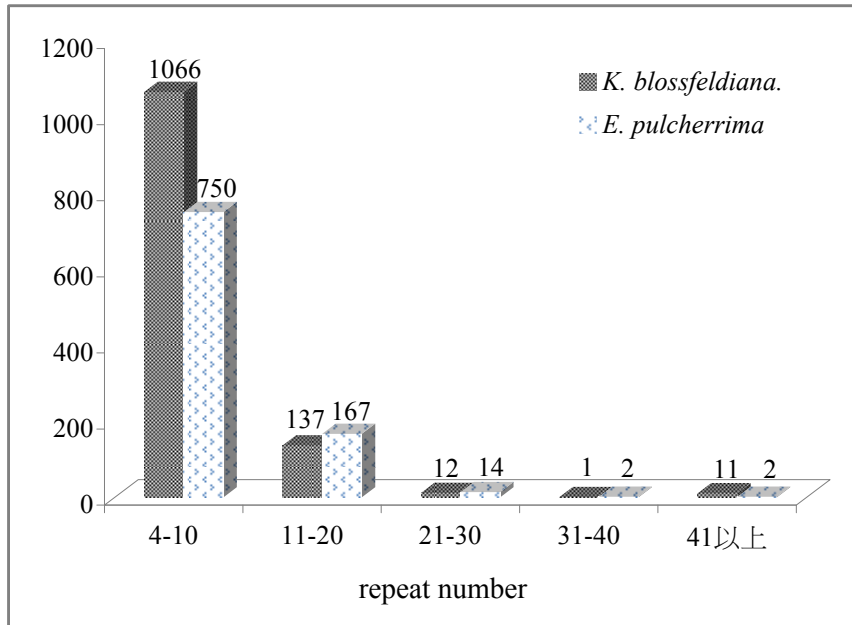


圖 1. 長壽花與聖誕紅 SSR 重複次數之分佈

Fig 1. Distribution of the number of SSR repeats in *K. blossfeldiana* and *E. pulcherrima*.

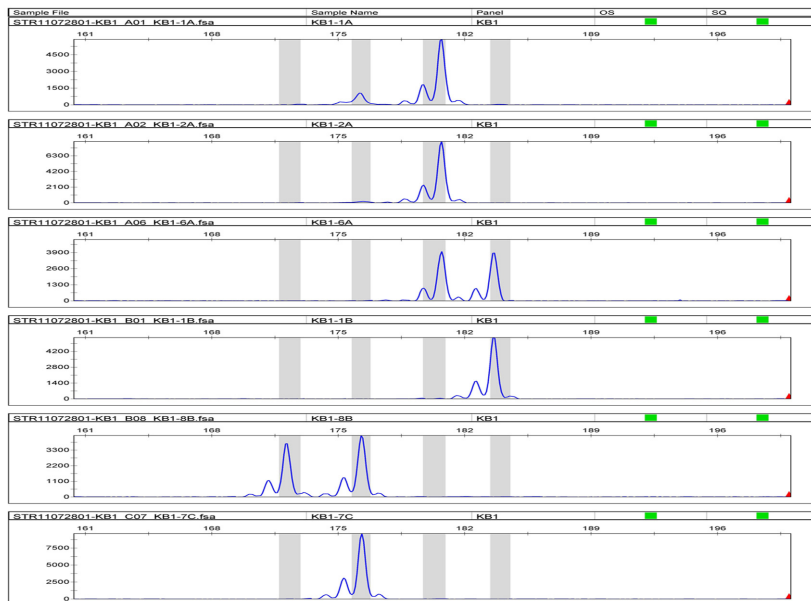


圖 2. 長壽花 Kb-AGCT-02 基因座之毛細管電泳圖。

Fig 2. Capillary electrophoresis result of *K. blossfeldiana* amplified by Kb-AGCT-02 marker.



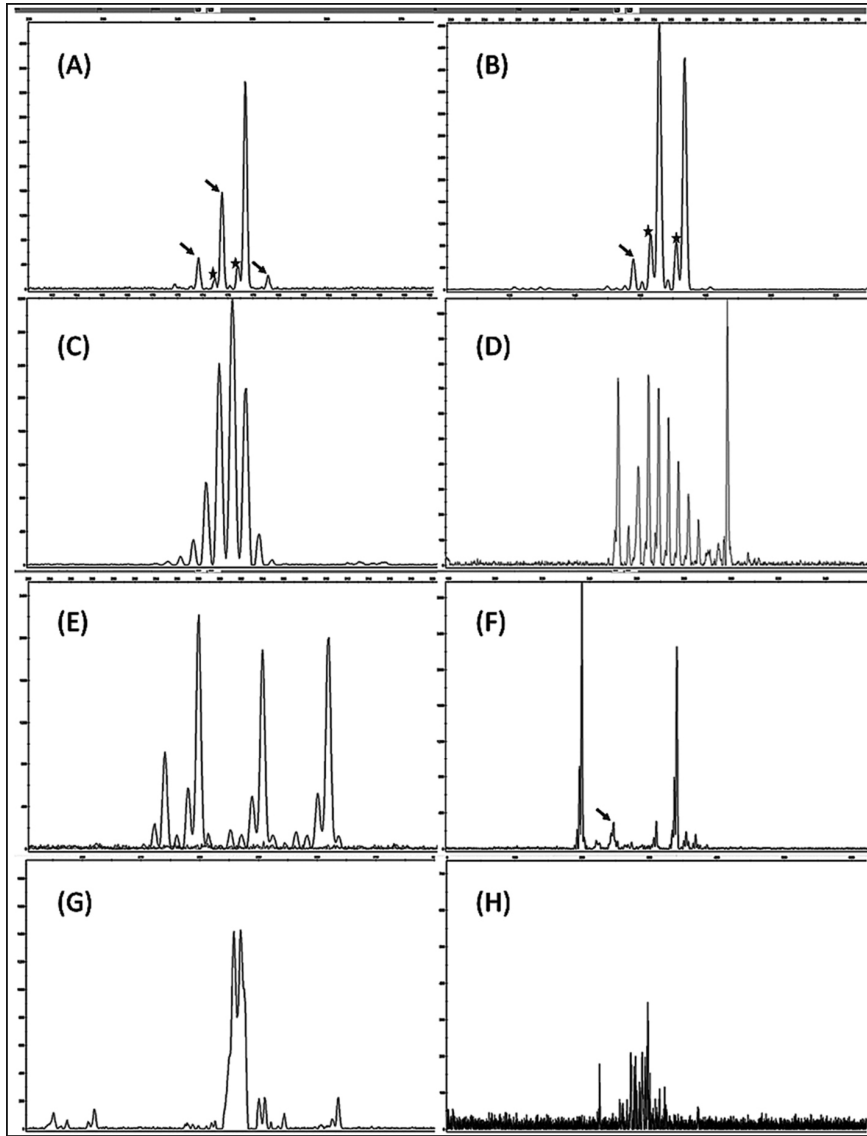


圖 3. SSR 分子標誌擴增條帶表現之範例，(A)同質結合；(B)異質結合，箭頭所指為跡痕條帶之波峰，星號所指為加 A 效應不完全所之波峰；(C)跡痕條帶訊號過強；(D)異質結合且跡痕條帶訊號過強；(E)多對偶基因(multiple allele)；(F)箭頭所指為雜訊；(G)加 A 效應不完全；(H)擴增條帶訊號微弱，難與雜訊區分，無法判讀。

Fig 3. Examples of PCR products amplified by SSR primer pairs in this study. (A)homozygote; (B)heterozygote, the arrow is stutter bands, asterisk is addition of adenine incomplete; (C)stutter bands excessively over; (D)heterozygote and stutter bands excessively over; (E)multiple allele; (F)the arrow is noise; (G)addition of adenine incomplete; (H) noise.

長壽花及聖誕紅分別設計 64 及 55 組引子測試，結果如表 2 所示，長壽花可以擴增預期條帶有 44 組，其中 33 組具有多型性，聖誕紅可以擴增預期條帶有 40 組，其中 28 組具有多型性，符合作為的 SSR 分子標誌者，佔所設計引子 50% 以上。Abdelkrim 等 (2009) 以 1/16 的 Roche/454 進行藍鴨 (*Hymenolaimus malacorhynchos*) 定序，得到 17,215 條序列，其中含 SSR 基因座序列為 231 條，最後獲得 24 個 SSR 分子標誌，而定序經費只需要 1500 美金。Castoe 等 (2010) 使用 Roche/454 FLX 來進行銅頭蛇 (*Agkistrodon contortrix*) 定序，得到 128,773 條序列，含 SSR 基因座序列者有 14,612 條 (11.3%)，其中有潛力設計引子的有 4,564 條序列，由抽 DNA 到完成定序只花了 9 天的時間。另外 Santana 等 (2009) 利用 ISSR-PCR 及 FIASCO (Fast Isolation by AFLP of Sequences COntaining repeats) 方式篩選包含 *Fusarium circinatum*, *Sirex noctilio* 和 *Deladenus siricidicol* 三個物種之 SSR 豐富性基因組文庫，再以 Roche/454 定序，最後獲得 1040-1840 條序列，可設計引子有 159-336 個，其中以 FIASCO 方式有較大的比例可以設計引子，而 ISSR-PCR 則因 SSR 序列太靠近兩側，不易設計引子。

傳統開發 SSR 分子標誌，需要建構成小片段插入基因組基因庫，再以含有微衛星序列的放射性探針篩選，傳統法 (Sanger method) 定序，設計引子，PCR 條件最佳化，最後測試是否具多型性；而以 NGS 法只需要抽取基因組 DNA，再經 NGS 定序，搜尋含 SSR 序列，同樣再設計引子，測試 PCR 條件及多型性，由於 NGS 定序數量非常大，因此，僅需要最小量的反應數，就足夠開發 SSR 分子標誌 (Santana, 2009)。藉由 NGS 技術可減少在選殖所耗費的人力及經費，由於單一序列的定序成本也大幅降低，且 Roche/454 定序之長度已接近傳統定序法，並可在一次試驗中同時進行不同物種的定序，已是發展 SSR 分子標誌重要方法。目前已有許多報告發表，如在微生物與腺蟲 (Santana, 2009)、鱗翅類 (*Euphydryas aurinia*) (Sinama *et al.*, 2011)、甲蟲 (*Ips typographus*) (Stoeckle and Kuehn, 2011)、蛇類 (*Vipera ursinii*) (Metzger *et al.*, 2011)、蜥蜴 (*Amphibolurus norrisi*, *Ctenotus atlas* and *Nephrurus stellatus*) (Smith *et al.*, 2011)、蝙蝠 (*Artibeus lituratus*) (McCulloch and Stevens, 2011) 及鯊魚 (*Galeorhinus galeus*) (Chabot and Nigenda, 2011) 等；在植物有香蒲 (*Typha minima*) (Csencsics *et al.*, 2010)、大美木豆 (*Pericopsis elata*) (Micheneau *et al.*, 2001)、金合歡 (*Acacia harpophylla*) (Lepais and Bacles, 2011)、桃花心木 (*Khaya senegalensis*) (Sexton *et al.*, 2010) 等，分別以 1/16-1/4 的 Roche/454 反應，即可開發出研究所需之 SSR 分子標誌。在本研究中，長壽花及聖誕紅獲得具開發 SSR 之潛

力者分別有 514 及 732 條序列，經過測試，已經篩選出 33 及 28 個具有多型性之 SSR 分子標誌，可作為品種鑑定及親緣關係之研究，顯示以 Roche/454 定序開發長壽花及聖誕紅 SSR 分子標誌，的確是一個快速、簡單且效率高的方式。

## 誌 謝

本研究試驗期間承蒙台灣大學農藝學系胡凱康老師、陳凱儀老師及林延諭同學提供實驗方法及資料分析指導，毛清原助理協助實驗室工作，以及榮陽基因體研究中心定序分析，謹致謝忱。

## 參考文獻

- 農糧署全球資訊網。2012。植物品種權公告查詢系統(<http://newplant.coa.gov.tw/>)。
- Abdelkrim, J., B. C. Robertson, J.-A.L. Stanton, and N.J. Gemmell. 2009. Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. *Bio.Techniques* 46:185-192.
- Aggarwal, R.K., P.S. Hendre, R.K. Varshney, P.R. Bhat, V. Krishnakumar and L. Singh. 2007. Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. *Theor. Appl. Genet.* 114(2):359-372.
- Aida, R. and M. Shibata. 2002. High frequency of polyploidization in regenerated plants of *Kalanchoe blossfeldiana* 'Tetra Vulcan'. *Plant Biotech* 19:329-334.
- Blair, MW, F. Pedraza, H.F. Buendia, E. Gaita'n-Solís, S.E. Beebe, P. Gepts and J. Tohme. 2003. Development of a genome-wide anchored microsatellite map for common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 107:1362-1374.
- Brzyski, J.R. 2010. Isolation and characterization of microsatellite markers in the rare clonal plant, *Spiraea virginiana* (Rosaceae). *Am. J. Bot.* 97:20-22.
- Castoe, T.A., A.W. Poole, Gu W., A.P.J. de Koning, J.M. Daza, E.N. Smith and D.D.

- Pollock. 2010. Rapid identification of thousands of copperhead snake (*Agkistrodon contortrix*) microsatellite loci from modest amounts of 454 shotgun genome sequence. *Mol. Ecol. Resour.* 10:341-347.
- Chabot, C.L., S. Nigenda. 2011. Characterization of 13 microsatellite loci for the tope shark, *Galeorhinus galeus*, discovered with next-generation sequencing and their utility for eastern Pacific smooth-hound sharks (*Mustelus*). *Conservation Genet. Resour.* 3:553-555.
- Csencsics, D., S. Brodbeck, R. Holderegger. 2010. Cost-effective, species-specific microsatellite development for the endangered dwarf Bulrush (*Typha minima*) using next-generation sequencing technology. *J. Hered.* 101:789-793.
- Culley, T. 2005. Characterization of newly developed microsatellite loci in the stemmed yellow violet, *Viola pubescens* (Violaceae). *Mol. Ecol. Notes* 5:882-884.
- Deschamps, S. and M.A. Campbell. 2010. Utilization of next-generation sequencing platforms in plant genomics and genetic variant discovery. *Mol. Breed.* 25:553-570.
- Doležel, J. and J. Bartoš. 2005. Plant DNA flow cytometry and estimation of nuclear genome size. *Ann. Bot.* 95:99-110.
- Ellison, C.K. and K.L. Shaw. 2010. Mining non-model genomic libraries for microsatellites: BAC versus EST libraries and the generation of allelic richness. *BMC Genomics* 11:428.
- Field, D. and C. Wills. 1996. Long, polymorphic microsatellites in simple organisms. *Proc. Biol. Sci.* 263:209-215.
- Galbraith, D.W., K.R. Harkins, J.M. Maddox, N.M. Ayres, D.P. Sharma and E. Firoozabady. 1983. Rapid flow cytometric analysis of the cell cycle in intact plant tissues. *Science* 220:1049-1051.
- Guichoux, E, Lagache L, Wagner S, Chaumeil P, Léger P, Lepais O, Lepoittevin C, Malausa T, Revardel E, Salin F, Petit RJ. 2011. Current trends in microsatellite genotyping. *Mol. Ecol. Resour.* 11:591-611.
- Gur-Arie, R., C.J. Cohen, Y. Eitan, L. Shelef, E.M. Hallerman and Y. Kashi. 2000. Simple sequence repeats in *Escherichia coli*: abundance, distribution, composition, and polymorphism. *Genome Res.* 10:62-71.

- Izumikawa, Y., S. Takei, I. Nakamura, M. Mii. 2008. Production and characterization of inter-sectional hybrids between *Kalanchoe spathulata* and *K. laxiflora* (= *Bryophyllum crenatum*) . *Euphytica* 163:123-130.
- Katti, M.V., P.K. Ranjekar, and V.S. Gupta. 2001. Differential distribution of simple sequence repeats in eukaryotic genomes sequences. *Mol. Biol. Evol.* 18:1161-1167.
- Kelkar, Y.D., N. Strubczewski, S.E. Hile, 2010. What is a microsatellite: a computational and experimental definition based upon repeat mutational behavior at A/T and GT/AC repeats. *Genome Biol. Evol.* 2:620-635.
- Knapik, E.W., A. Goodman, M. Ekker. 1998. A microsatellite genetic linkage map for zebrafish (*Danio rerio*) . *Nat. Genet.* 18:338-343.
- Lepais, O. and C.F.E. Bacles. 2011. Comparison of random and SSR-enriched shotgun pyrosequencing for microsatellite discovery and single multiplex PCR optimization in *Acacia harpophylla* F. Muell. Ex Benth. *Mol. Ecol. Resour.* 11:711-724.
- Levinson, G., G.A. Gutman. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* 4:203-221.
- Li, Y.C., A.B. Korol, T. Fahima, A. Beiles and E. Nevo. 2002. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Mol. Eco.* 11:2453-2465.
- Maia, L.C.da, D.A. Palmieri, V.Q. de Souza, M.M. Kopp, F.I. de Carvalho, A.C. de Oliveira. 2008. SSR Locator: Tool for simple sequence repeat discovery integrated with primer design and PCR simulation. *Int J Plant Genomics.* 2008:412696.
- Mba, R.E.C., P. Stephenson, K. Edwards, S. Melzer, J. Nkumbira, U. Gullberg, K. Apel · M. Gale, J. Tohme and M. Fregene. 2001. Simple sequence repeat (SSR) markers survey of the cassava (*Manihot esculenta* Crantz) genome: towards an SSR-based molecular genetic map of cassava. *Theor. Appl. Genet.* 102:21-31.
- McCulloch, E.S. and R.D. Stevens. 2011. Rapid development and screening of microsatellite loci for *Artibeus lituratus* and their utility for six related species within Phyllostomidae. *Mol. Ecol. Resour.* 11:903-913.
- Metzger, C., A.L. Ferchaud, C Geiser and S. Ursenbache. 2011. New polymorphic microsatellite markers of the endangered meadow viper (*Vipera ursinii*) identified by

- 454 high-throughput sequencing: when innovation meets conservation. *Conservation Genet. Resour.* 3:589-592.
- Micheneau, C., G. Dauby, N. Bourland, J.L. Doucet and O. J. Hardy. 2011. Development and characterization of microsatellite loci in *Pericopsis elata* (fabaceae) using a cost-efficient approach. *Am. J. Bot.* 98:e268-e270.
- Mittal, N. and A. Dubey. 2009. Microsatellite markers-A new practice of DNA based markers in molecular genetics. *Phcog. Rev.* 3:235-246.
- Pazouki, L., M. Mardi, P.S. Shanjani, M. Hagidimitriou, S.M. Pirseyedi, M.R. Naghavi, Avanzato D., E. Vendramin, S. Kafkas, B.Ghareyazie 2010. Genetic diversity and relationships among *Pistacia* species and cultivars. *Conserv. Genet.* 11:311-318.
- Saha, M.C., J.D. Cooper, M.A.R. Mian, K. Chekhovskiy, and G.D. May, 2006. Tall fescue genomic SSR markers: development and transferability across multiple grass species. *Theor. Appl. Genet.* 113:1449-1458.
- Santana, Q.C., M.P.A. Coetzee<sup>1</sup>, E.T. Steenkamp, O.X. Mlonyeni<sup>1</sup>, G.N.A. Hammond, M.J. Wingfield, and B.D. Wingfield. 2009. Microsatellite discovery by deep sequencing of enriched genomic libraries. *BioTechniques* 46:217-223.
- Schlötterer, C. 2000. Evolutionary dynamics of microsatellite DNA. *Chromosoma* 109: 365–371.
- Schlötterer, C. and D. Tautz. 1992. Slippage synthesis of simple sequence DNA. *Nucleic Acids Res.* 20:211-215.
- Schuelke, M. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nat. Biotechnol.* 18:233-234.
- Schuler, G.D., Boguski M.S., Stewart E.A. *et al.* 1996. A gene map of the human genome. *Science* 274:540-546.
- Selkoe, K.A. and R.J. Toonen. 2006. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecol. Let.* 9:615-629.
- Sexton, G.J., C.H. Frere, M.J. Dieters, I.D. Godwin, and P.J. Prentis. 2010. Development and characterization of microsatellite loci for *Khaya senegalensis* (meliaceae) . *Am. J. Bot.* 97:e111-e113.
- Sinama, M., V. Dubut, C. Costedoat, A. Gilles, M. Junker, T. Malausa, J.F. Martin, G.

- Nève, N. Pech, T. Schmitt, M. Zimmermann and E. Megléc. 2011. Challenges of microsatellite development in Lepidoptera: *Euphydryas aurinia* (Nymphalidae) as a case study. *Eur. J. Entomol.* 108:261-266.
- Smith, A.L., M.G. Gardner, C.M. Bull, D.A. Driscoll. 2011. Primers for novel microsatellite markers in “fire-specialist” lizards (*Amphibolurus norrisi*, *Ctenotus atlas* and *Nephrurus stellatus*) and their performance across multiple populations. *Conservation Genet. Resour.* 3:345-350.
- Stoeckle, B.C. and R. Kuehn. 2011. Identification of 18 polymorphic microsatellite loci in the spruce bark beetle *Ips typographus* (Coleoptera: Scolytidae) using high-throughput sequence data. *Eur. J. Entomol.* 108:169-171.
- Takayama, K; Lopez, P; König, C; Kohl, G; Novak, J; Stuessy, TF. 2011: A simple and cost-effective approach for microsatellite isolation in non-model plant species using small-scale 454 pyrosequencing. *Taxon* 60:1442-1449.
- The Royal Botanic Gardens, Kew. 2012. Plant DNA C-values Database. (<http://data.kew.org/cvalues/>)
- Wang Z., J.L. Weber, G. Zhong and S.D. Tanksley. 1994. Survey of plant short tandem DNA repeats *Theor. Appl. Genet.* 88:1-6.
- Weber, J.L. 1990. Informativeness of human (dC-dA) n. (dG-dT) n polymorphisms. *Genomics* 7:524-530.
- Webster, M.T., N.G.C. Smith, H. Ellegren. 2002. Microsatellite evolution inferred from human-chimpanzee genomic sequence alignments. *P. Natl. Acad. Sci. U.S.A.* 99: 8748-8753.
- Zane, L., L. Bargelloni and T. Patarnello. 2002. Strategies for microsatellite isolation: a review. *Mol. Ecol.* 11:1-16.

# Development of Microsatellite Markers in *Kalanchoe blossfeldiana* and *Euphorbia pulcherrima* using Small-scale Next Generation Genome Sequencing<sup>1</sup>

Chih-Hsin Yeh<sup>2</sup> and Fang-Hsin Liao<sup>2</sup>

## Abstract

Microsatellite markers are widely used in DNA fingerprint, variety identification, genetic relationship and genetic mapping. It is an expensive and time-consuming job to *de novo* isolate and to characterize SSR markers for unexamined species. Recently, high-throughput sequencing technologies provide new opportunities for locating microsatellites in not-previously-sequenced orphan crop. Here we report results of 454 pyrosequencing to identify a large number of microsatellites from *Kalanchoe blossfeldiana* and *Euphorbia pulcherrima*. Using a combination of a total genomic library and small-scale pyrosequencing (1/4 of a PicoTiterPlate) with multiplexing option, we determined 60.9 and 53.3 Mbp of sequence (180,648 and 135,085 reads) in the genomes of the two species and identified 732 and 514 microsatellite loci that had flanking regions suitable for primer design. We tested PCR amplification, reproducibility, and polymorphism of 28 for *E. pulcherrima* and 33 primer pairs for *K. blossfeldiana* using six individuals. As a result, 50% of markers were clearly genotyped, efficaciously detected polymorphisms. This method is low-cost, time-effective and large-scale in comparison with traditional approaches involving cloning and sequencing.

Key words: simple sequence repeat (SSR), Next Generation Sequencing (NGS), 454 pyrosequencing, *Kalanchoe blossfeldiana*, *Euphorbia pulcherrima*

---

<sup>1</sup>. Contribution No.432 from Taoyuan DARES, COA.

<sup>2</sup>. Assistant Researcher (Corresponding author, zeamays@tydais.gov.tw) and Researcher, respectively, Taoyuan DARES, COA.