

次世代定序在玉米基因定位與育種上的應用

鄭舒允

中央研究院植微所

摘 要

次世代定序 (Next Generation Sequencing) 能快速且大量的產生 DNA 序列，而 GBS (Genotyping By Sequencing) 即是以次世代定序為基礎的應用之一。GBS 可以快速的找到不同樣本間涵蓋全基因組的單一核苷酸多型性 (Single Nucleotide Polymorphism, SNP) 以及酵素截切差異位置藉此獲得大量的分子標誌。在基因定位的目的上，GBS 的應用可以快速的填補全基因組標誌數的不足，配合特殊族群的建立方式，能夠有效的限縮目標基因的可能位置，減少人力、財力以及時間的消耗。本文將介紹藉由 GBS 的方式進行全基因組 miRNA 表現數量性狀基因座關連分析 (miRNA eQTL Genome-wide Association Study) 以及異交不親和 *gal* 基因定位的應用。

關鍵詞：次世代定序、GBS、SNP、GWAS、異交不親和

前 言

利用次世代定序檢測基因型

次世代定序 (Next Generation Sequencing) 技術是以 Sanger sequencing 為基礎，目前以 Illumina 公司的解序技術為最大宗，主要步驟如下：1. 萃取 DNA; 2. 建立 Library; 3. 將建構好的 Library 解序; 4. 序列分析。游離核苷酸帶有不同的螢光，一旦合成股成功結合上新的核苷酸，就會釋放螢光由螢光感測器偵測，從而得知每一個位置的序列。由於螢光偵測以及電腦的應用，NGS 能夠快速且大量的產生短片段 (約 100-200 bp) 序列 (Fig. 1)⁷。

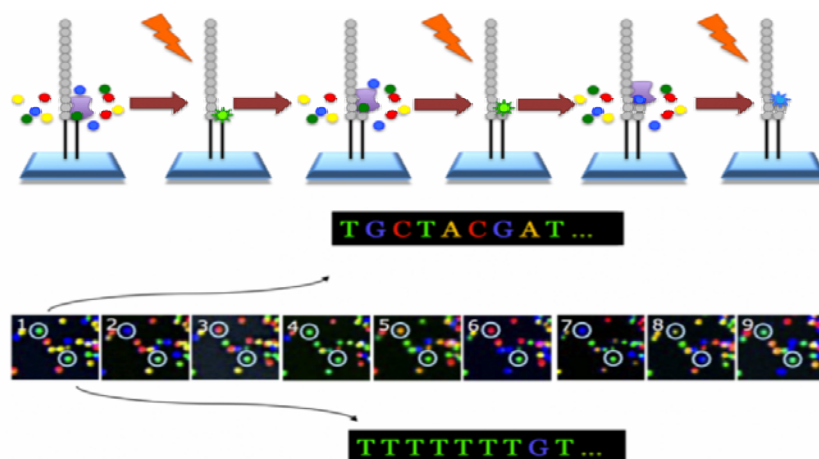


Fig. 1. 次世代定序原理。

由於技術的進步，次世代定序的價格從 2001 年逐年降低，如果將與人類基因組大小相近的物種 (3.2Gbp) 進行次世代定序，所需的價格只有 2001 年的十萬分之一 (Fig. 2)¹³。

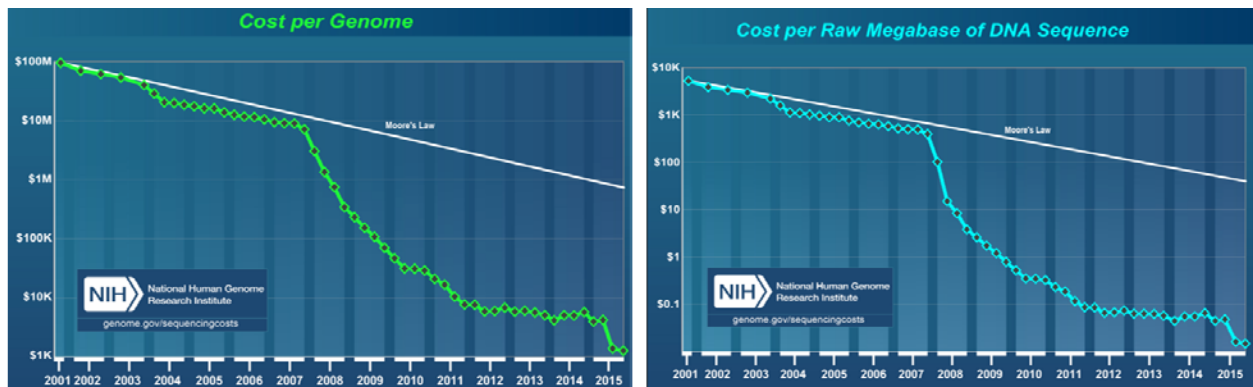


Fig. 2. 次世代定序以人類基因組為例的價格波動。

- (A) 進行完整解序所需價格
- (B) 原始數據中平均每百萬核苷酸所需價格

因為次世代定序的價格越來越合理，使得這個技術不僅能成為建立全基因組分子標誌的方式之一，另外包括快速的找出樣本間的多型性、樣本基因表現差異，甲基化修飾差異等都是應用範圍。目前在利用次世代定序進行基因型鑑定的方式有很多，包括 Multiplex shotgun genotyping (MSG)、Restriction association DNA sequencing (RAD-seq)、Genotyping-by-sequencing (GBS) 等，這些方法在建構 library 的過程中結合指引條碼(index barcode)，使得不同的樣本能夠混和在一起進行解序，進而降低每個樣品解序的成本 (Table 1)¹⁴。

Table 1. A technical comparison of current genotyping methods using next-generation sequencing of multiplex barcoded libraries

Method	Random shearing	Size selection	Fragment size	Enzymes [†]	Multiplexing level [‡]	Analysis tool(s)	Reference
Multiplex shotgun genotyping	No	Yes	Size selected	<i>MseI</i>	96 (up to 384)	Burrows-Wheeler alignment tool	Andolfatto et al., 2011
Restriction association DNA sequencing (RAD-seq)	Yes	Yes	Size selected	<i>SbfI</i> <i>EcoRI</i>	96	Custom Perl scripts	Baird et al., 2008
Double digest RAD-seq	No	Yes	Size selected	<i>EcoRI</i> and <i>MspI</i>	48 [§]	MUSCLE [¶]	Peterson et al., 2012
2b-restriction association DNA	No	No	33–36 bp	<i>BsaXI</i> [#]	NA ^{††}	Custom Perl scripts	Wang et al., 2012
Genotyping-by-sequencing	No	No	<350 bp	<i>ApeKI</i> ^{‡‡}	48 (up to 384)	TASSEL ^{§§}	Elshire et al., 2011
Genotyping-by-sequencing – two enzyme	No	No	<350 bp	<i>PstI</i> and <i>MspI</i>	48 (up to 384)	TASSEL	Poland et al., 2012a
Sequence-based genotyping	No	Yes	Size selected	<i>EcoRI</i> and <i>MseI</i> <i>PstI</i> and <i>TaqI</i>	32	Burrows-Wheeler alignment tool and unified genotyper	Truong et al., 2012
Restriction enzyme sequence comparative analysis	No	Yes	Size selected	<i>MseI</i> <i>NlaIII</i>	NA ^{¶¶}	Burrows-Wheeler alignment tool and Samtools	Monson-Miller et al., 2012

[†]All of these approaches can use different enzymes. Shown are the enzyme(s) used in the initial study.

[‡]All of these methods have the possibility to increase the number of multiplexed samples using additional unique barcodes. The multiplex level as reported in the reference paper. Given in parenthesis are subsequent increases.

[§]Combinatorial barcoding is possible, placing a barcode on each end of the DNA fragment. Using a set of 48 adapter P1 barcodes and × 12 polymerase chain reaction (PCR) 2 indices it is possible to uniquely label 576 individuals (48 [adapter P1 barcodes] × 12 [PCR2 indices]). This method would require paired-end sequencing.

[¶]MUSCLE, multiple sequence comparison by log-expectation.

[#]Uses type IIB restriction endonucleases.

^{††}NA, not applicable.

^{‡‡}Has been successfully applied to using *PstI* and *HindIII* (E. Buckler and R. Elshire, personal communication, 2012).

^{§§}TASSEL, trait analysis by association, evolution, and linkage.

^{¶¶}96-plexing reported but unpublished.

經由次世代定序進行基因鑑定獲得的資訊，可以更進一步的使用在許多不同的範疇，包括遺傳上的相關性圖譜建立 (Association mapping)、連鎖圖譜建立 (Linkage mapping) 以及育種上的基因組選擇 (Genomic selection) 等，讓基因組的資訊更加完備 (Fig. 3)¹⁴。

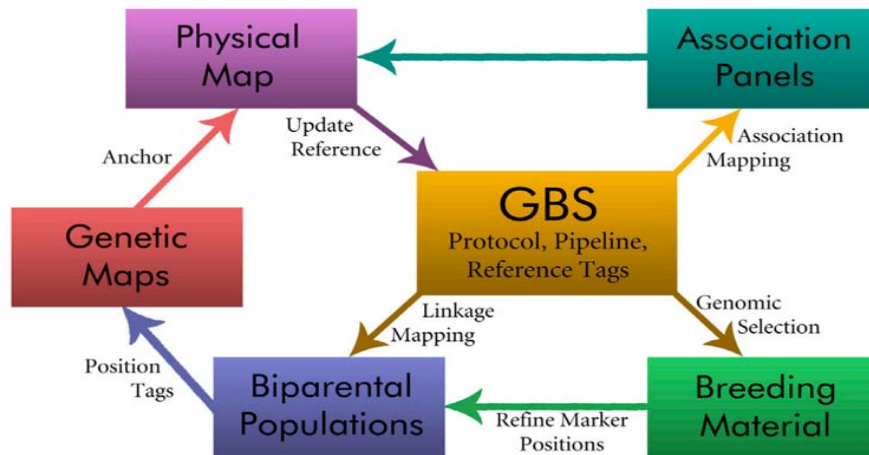


Fig. 3. 次世代定序進行基因型鑑定資訊應用。

次世代定序在基因定位的應用

數量性狀基因座 (Quantitative Trait Loci, QTL) 研究是遺傳研究的範疇，一般作物的重要農藝性狀皆屬於此⁹，其中數量基因調控性狀最顯著的例子為番茄果實大小的差異性⁶。目前已在不同的作物物種中找到許多調控不同性狀的 QTLs，包括了大麥中與種子休眠、條紋病 (stripe rust) 抗性相關的調控基因³；多年生黑麥草生長發育以及耐寒性相關的數量基因²³；豌豆基因組中定位得到影響種子產量、產量成分與生長特性的 QTLs²¹；玉米種子的耐淹水相關 QTLs¹⁶；水稻中則有找到控制淹水耐性的 QTLs²⁰、調控種子寬與重的 QTLs 基因-GW2¹⁸與 GW5 的選殖²²、小麥的產量調控 QTLs 選殖¹⁵等。

現今定位 QTLs 並選殖出 QTL 基因的方式以 positional cloning 與 association mapping 為主 (Fig. 4)¹⁹，Positional cloning 主要利用估計遺傳標誌或分子標誌與目標性狀間的相關性來決定，包含三個步驟：1.找到目標性狀有差異的親本；2.建構有多型性存在的遺傳圖譜；3.QTLs 相關性分析¹¹。GWAS 是從演化上的觀點尋找同物種不同品種、品系、地方種等個體的不同等位基因間連鎖不平衡 (Linkage disequilibrium, LD)與性狀間相關性的策略。LD 的解釋為「兩個以上基因座的等位基因出現非隨機連鎖」的可能，其高低程度可藉由預期的單套體 (haplotype) 頻度與觀測到的單套體頻度間差異求得⁵。GWAS 應用於植物 QTLs 找尋上擁有三項優勢：1.不需大量的試驗族群與時間；2.有更高的遺傳解析度；3.可以在同一基因座上找到演化成不同等位基因的功能性基因¹⁹。GWAS 的應用包括了高粱¹²、玉米¹⁰、乳牛¹⁷等，相當廣泛。

但上述的方式會受限於圖譜解析度不夠高，真正能夠從圖譜上定位到的位置可能距離實際影響基因仍有距離，降低在育種上的應用¹⁹。而次世代定序快速且大量找到涵蓋全基因組的標誌，就可以有效解決圖譜解析度不足的問題，讓找到的連鎖標誌相當靠近目標基因。

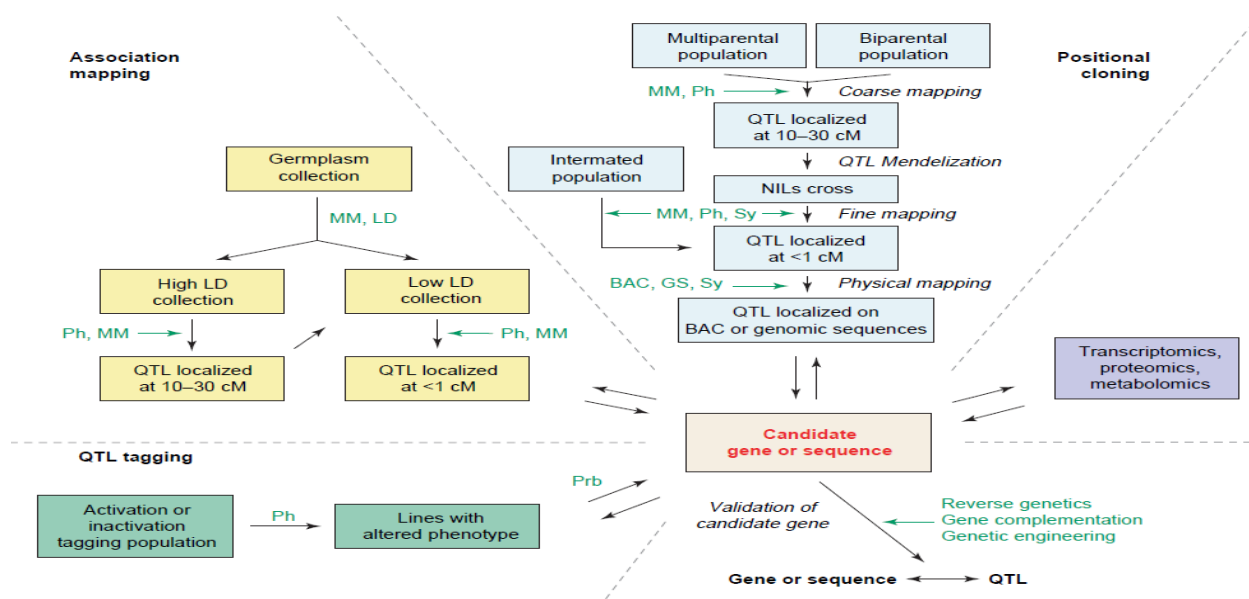


Fig. 4. 利用分子標誌方式進行數量性狀基因座定位的不同流程。

案例一：避免 GMO 污染的救世主-異交不親和 *gal* 基因

研究發現玉米異花授粉的情況在相隔 650 公尺仍能發生，其花粉更能夠飄到數公里之遠，因此如何防範不同種植區玉米品種間花粉互相污染是相當重要的。玉米的異交不親和基因 (cross-incompatibility gene) *gametophyte factor 1 (gal)* 即為解決方式之一，*gal* 基因的特性是母株的柱頭是帶有 *Ga1-S* 等位基因時，帶有 *gal* 等位基因的花粉無法有效的延長花粉管而授精⁸ (Table 2)

Table 2. 異交不親和等位基因間的關係

	<i>Ga1-S</i> (♂)	<i>Ga1-M</i> (♂)	<i>gal</i> (♂)
<i>Ga1-S</i> (♀)	✓	✓	X
<i>Ga1-M</i> (♀)	✓	✓	✓
<i>gal</i> (♀)	✓	✓	✓

[†]*Ga1-S*, gametophyte factor 1-strong allele; *Ga1-M*, gametophyte factor 1-male; *gal*, gametophyte factor-1; ✓, cross compatible; X, cross incompatible; ♂, male; ♀, female.

Bloom 等學者²以帶有 *gal/gal* 等位基因的 B73 自交系為母本與帶有 *Gal/Gal* 基因座的 Hp301 自交系為父本，建構 192 株自交五代的重組自交系 (recombinant inbred lines, RILs)；利用 1106 個 SNP 標誌，將 *gal* 基因座定位於第四條染色體短臂上 (Fig. 5)，又進一步利用 GBS 標誌將候選基因的範圍縮小到僅有 13 個基因的 2.6Mb 左右 (Fig. 6)。這些候選基因中，GRMZM2G135056 的功能可能與花粉交互作用有關，而 GRMZM2G039983 則是另一個與花粉管生長有關的候選基因 (Table 3)。

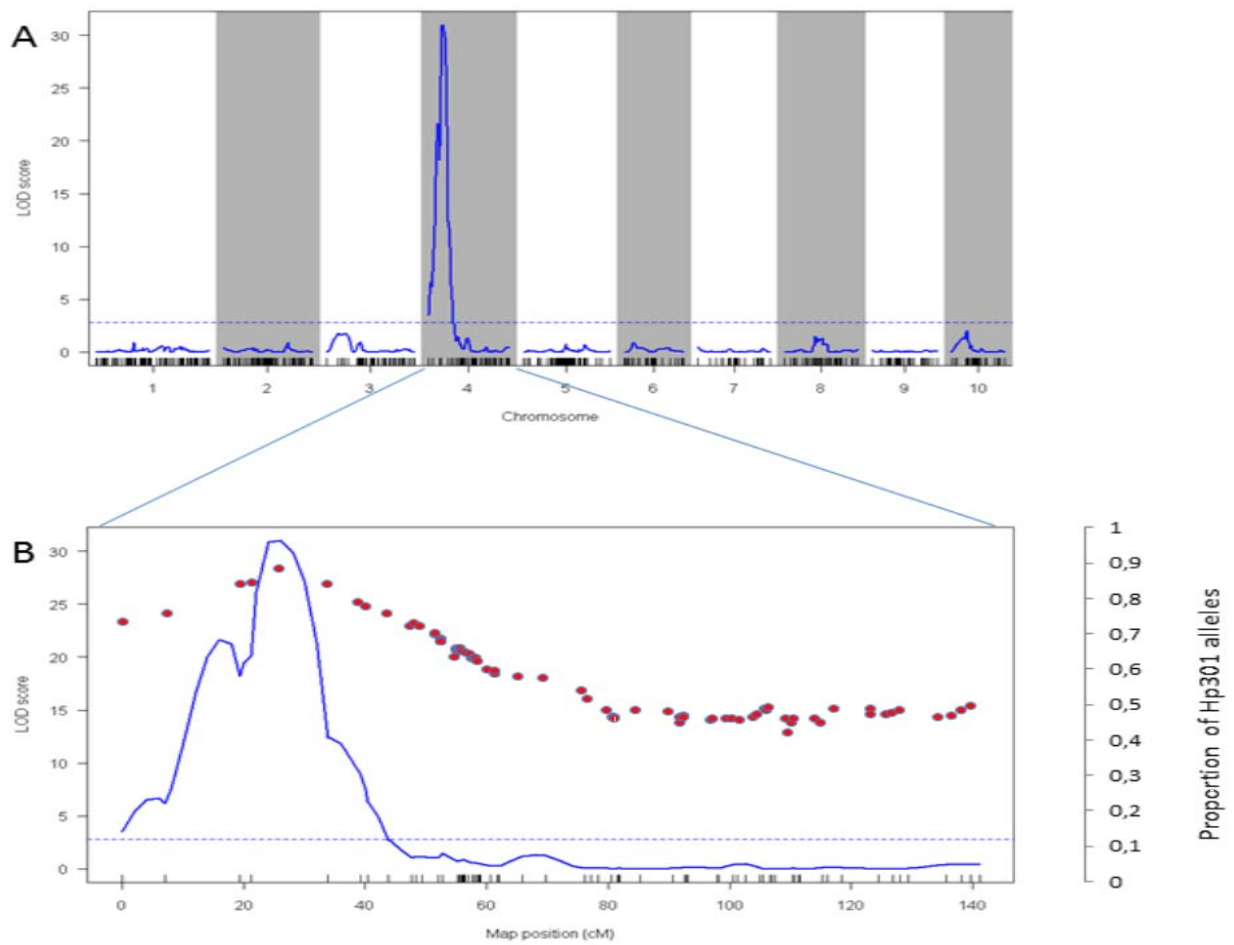


Fig. 5. B73xHp301 重組自交系的全基因組檢測。

(A) 全基因組檢測影響重組自交系接受 *gal/gal* 花粉的表現型結果

(B) 放大檢視定位到 *Gal* 基因的區間，紅色圓點代表 Hp301 等位基因的頻度

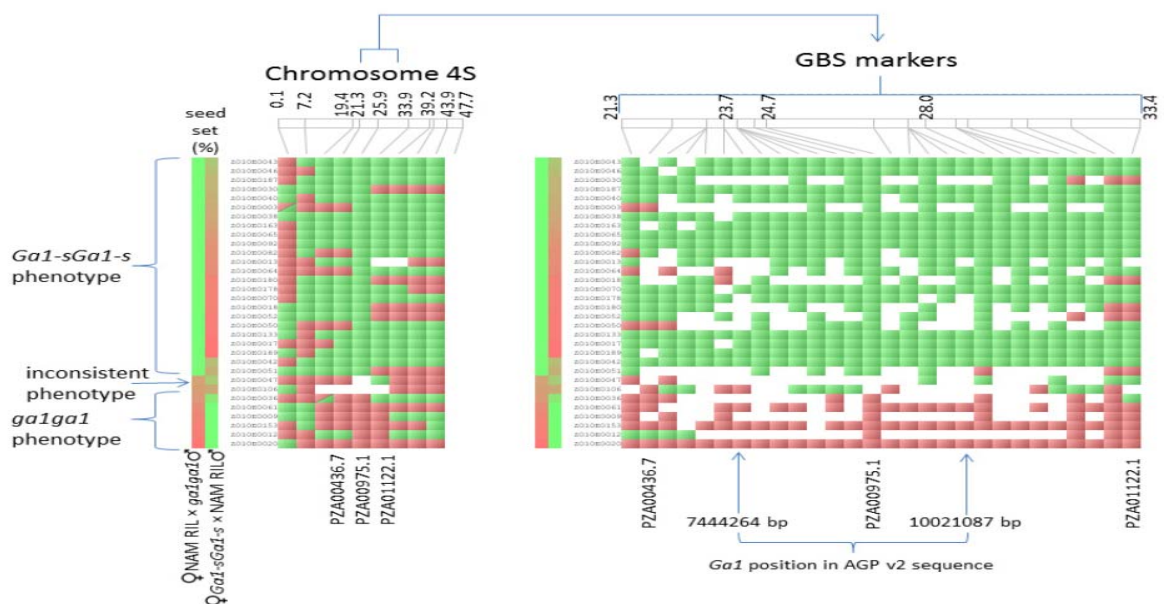


Fig. 6. 利用 GBS 的標誌去限縮並尋找更接近目標基因座的標誌。

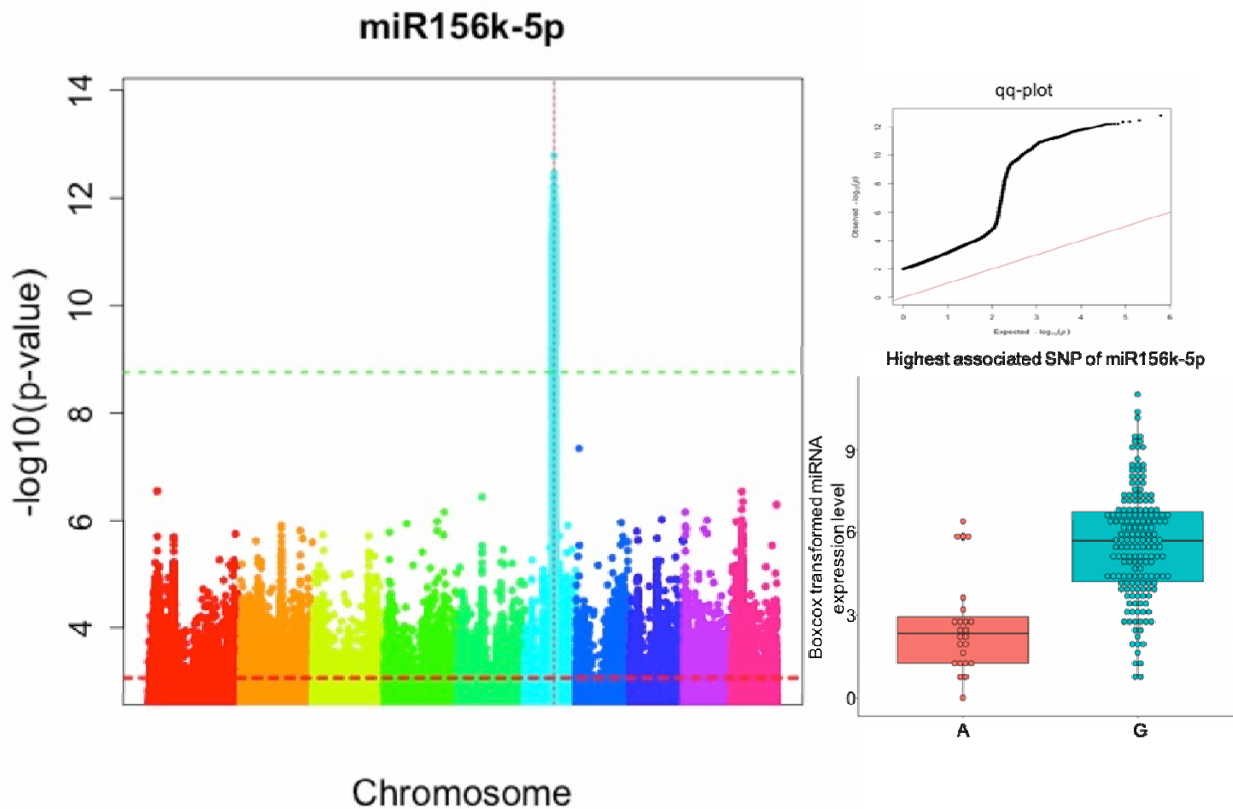
Table 3. *gal1* 基因座可能區間內的候選基因

Gene ID	AGP v2 sequence positions (bp)		Conserved domains
	Transcript start	Transcript end	
GRMZM2G012821	7616846	7618466	F-box domain cyclin-like
GRMZM2G424553	7653177	7691914	kinesin motor domain
GRMZM2G135056	7780877	7782970	nodulin-like
GRMZM2G181073	8078275	8079905	
GRMZM2G029496	8305887	8308705	
GRMZM5G835418	8899536	8900563	
AC196002.2_FG002	8901387	8901950	
AC201986.3_FG002	9183034	9183546	
GRMZM2G702344	9259652	9260731	
GRMZM5G817995	9325329	9325631	
GRMZM2G419836	9351020	9354236	thioredoxin-like fold
GRMZM2G027021	9485207	9494351	GTP-binding protein hflX
GRMZM2G039983	9589010	9592389	Xklp2 targeting protein, WDL1

來自藍河有機種子公司 (Blue River Organic Seed)¹ 的玉米商品「PuraMaize™」，即是利用這種異交不親和基因，讓商品本身的雌花僅能接受自己的商品的花粉，從而隔絕其他周遭玉米花粉的汙染，且特別強調可以在有機栽種上避免轉基因花粉的汙染。

案例二：miRNA 表現數量性狀基因座關連分析 (miRNA eQTL Genome-wide Association Study)

microRNA (miRNA) 是基因表現相當重要的調控因子之一，在動植物的不同生長發育時期都需要 miRNA 的參與，但是調控 miRNA 表現的基因座在玉米基因組裡仍未發現，Chen 等⁴，將玉米 282 個 NAM 自交系的發芽後根尖以及開花期的劍葉下一葉為目標，利用次世代定序的方式將成熟 miRNA 的表現量化 (smallRNA-seq) 獲得性狀，結合已知的基因型數據 Maize hapmap 3.21 進行全基因組關連性分析，找到調控 miRNA 表現的數量基因座 (Fig. 7)。



結 語

次世代定序能夠快速且大量的提供序列，結合不同分子標誌的找尋方式，能有效的找出涵蓋全基因組的差異性標誌，再加上不同指引條碼的組合，使得多個樣本一起解序成為可能，大大降低解序所需的時間與成本。Genotyping-by-Sequencing 即是其中一種方式，不僅能如本文介紹的應用在基因座的測定，還可尋找品種專一性的標誌來建立品種特有指紋、使用於演化的研究以瞭解變異。此外，這些大範圍高密度的標誌，更能應用在基因組選拔 (genome selection) 與基因組預測 (genome prediction) 等的未來育種策略上。

參考文獻

1. Blue River Organic Seed
<http://www.blueriverorgseed.com/products/corn/puramaize/145>
2. Bloom, J. C. and J. B. Holland. 2012. Genomic localization of the maize cross-incompatibility gene, Gametophyte factor 1 (ga1). *Maydica* 56-1782.
3. Castro, A. J., X. Chen, P. M. Hayes, S. J. Knapp, R.F. Line, T. Toojinda, and H. Vivar. 2002. Coincident QTL which determine seedling and adult plant resistance to stripe rust in barley. *Crop Sci.* 42: 1701-1708.
4. Chen, S. Y., M. H. Su, K. A. Kremling, N. K. Leapk, M. C. Romay, K. L. Swarts, Q. Sun, P. J. Bradbury, and E. S. Buckler. Identification of miRNA eQTLs in Maize Seedling Root and Mature Leaf Tissue by GWAS (on processing).
5. Flint Garcia, S. A., J. M. Thornsberry, and E. S. Buckler. 2003. Structure of linkage disequilibrium in plants. *Annu. Rev. Plant Biol.* 54: 357-374.
6. Frary, A., T. C. Nesbitt, A. Frary, S. Grandillo, E. V. D. Knaap, B. Cong, J. Liu, J. Meller, R. Elber, K. B. Alpert, and S. D. Tanksley. 2000. Fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. *Sci.* 289: 85-88.
7. Illumina website
<https://www.illumina.com/>
8. Liu, X., H. Sun, P. Wu, Y. Tian, D. Cui, C. Xu, S. Li, P. Li, H. Zhang, T. Chen, D. Li, X. Zhao, Y. Zhang, Y. Xue, and H. Chen. 2014. Fine mapping of the maize cross-incompatibility locus gametophytic factor 1 (ga1) using a homogeneous population. *Crop Science* 54: 1-9.
9. Kearsley, M. J. and A. G. L. Farquhar. 1998. QTL analysis in plants; where are we now? *Heredity* 80: 137-142.
10. Massman J. M., A., Gordillo, R. E. Lorenzana, and R. Bernardo. 2013. Genomewide predictions from maize single-cross data. *Theor. Appl. Genet.* 126: 13-22
11. Mauricio, R. 2001. Mapping quantitative trait loci in plants: uses and caveats for evolutionary biology. *Nat. Rev. Genet.* 2: 370-381.
12. Morris G. P., P. Ramu, S. P. Deshpande, C. T. Hash, T. Shah, H. D. Upadhyaya, O. Riera-Lizarazu, P. J. Brown, C. B. Acharya, S. E. Mitchell, J. Harriman, J. C. Glaubitz, E. S. Buckler, and S. Kresovich. 2013. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *PNAS* 110: 453-458
13. National human genome research institute website
<https://www.genome.gov/>

14. Poland, J. A. and T. W. Rife. 2012. Genotyping-by-Sequencing for plant breeding and genetics. *The Plant Genome*. 5: 92-102.
15. Quarrie, S. A., S. Pekic Quarrie, R. Radosevic, D. Rancic, A. Kaminska, J. D. Barnes, M. Leverington, C. Ceoloni, and D. Dodig. 2006. Dissecting a wheat QTL for yield present in a range of environments: from the QTL to candidate genes. *J. Exp. Bot.* 57: 2627-2637.
16. Qiu, F., Y. Zheng, Z. Zhang, and S. Xu., 2007. Mapping of QTL associated with waterlogging tolerance during the seedling stage in Maize. *Ann. Bot.* 99: 1067-1081.
17. Schaeffer L. R. 2006. Strategy for applying genome-wide selection in dairy cattle. *J. Anim. Breed. Genet.* 123: 218-223
18. Song, X. J., W. Huang, M. Shi, M. Z. Zhu, and H. X. Li. 2007. A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat. Genet.* 39: 623-630.
19. Salvi, S. and R. Tuberosa. 2005. To clone or not to clone plant QTLs: present and future challenges. *Trends Plant Sci.* 10: 297-304.
20. Toojinda, T., M. Siangliw, S. Tragoonrung, and A. Vanavichit. 2003. Molecular genetics of submergence tolerance in rice: QTL analysis of key traits. *Ann. Bot.* 91: 243-253.
21. Timmerman-Vaughan, G. M., A. Mills, C. Whitfield, T. Frew, R. Butler, S. Murray, M. Lakeman, J. McCallum, A. Russell, and D. Wilson. 2005. Linkage mapping of QTL for seed yield, yield components, and developmental traits in pea. *Crop Sci.* 45: 1336-1344.
22. Weng, J., S. Gu, X. Wan, H. Gao, T. Guo, N. Su, C. Lei, X. Zhang, Z. Cheng, X. Guo, J. Wang, L. Jiang, H. Zhai, and J. Wan., 2008. Isolation and initial characterization of GW5, a major QTL associated with rice grain width and weight. *Cell Res.*: 1-11.
23. Yamada, T., E. S. Jones, N. O. I. Cogan, A. C. Vecchies, T. Nomura, H. Hisano, Y. Shimamoto, K. F. Smith, M. D. Hayward, and J. W. Forster. 2004. QTL analysis of morphological, developmental, and winter hardiness-associated traits in perennial ryegrass. *Crop Sci.* 44: 925-935.