

수박 엘리트 계통의 GBS를 통한 마커이용 육종용 SNP 마커 개발

이준우 · 손병구 · 최영환 · 강점순 · 이용재 · 제병일 · 박영훈

Development of an SNP set for marker-assisted breeding based on the genotyping-by-sequencing of elite inbred lines in watermelon

Junewoo Lee · Beunggu Son · Youngwhan Choi · Jumsoon Kang · Youngjae Lee · Byoung Il Je · Younghoon Park

Received: 17 August 2018 / Revised: 17 August 2018 / Accepted: 17 August 2018

© Korean Society for Plant Biotechnology

Abstract This study was conducted to develop an SNP set that can be useful for marker-assisted breeding (MAB) in watermelon (*Citrullus lanatus* L.) using Genotyping-by-sequencing (GBS) analysis of 20 commercial elite watermelon inbreds. The result of GBS showed that 77% of approximately 1.1 billion raw reads were mapped on the watermelon genome with an average mapping region of about 4,000 Kb, which indicated genome coverage of 2.3%. After the filtering process, a total of 2,670 SNPs with an average depth of 31.57 and the PIC (Polymorphic Information Content) value of 0.1–0.38 for 20 elite inbreds were obtained. Among those SNPs, 55 SNPs (5 SNPs per chromosome that are equally distributed on each chromosome) were selected. For the understanding genetic relationship of 20 elite inbreds, PCA (Principal Component Analysis) was carried out with 55 SNPs, which resulted in the classification of inbreds into 4 groups based on PC1 (52%) and PC2 (11%), thus causing differentiation between the inbreds. A similar classification pattern for PCA was observed from hierarchical clustering analysis. The SNP set developed in this study has the potential for application to cultivar identification, F1 seed purity test, and marker-assisted backcross (MABC) not only for 20 elite inbreds but also for diverse resources for watermelon breeding.

Keywords Cultivar identification, GBS, MABC, SNP, watermelon

J. Lee · B. Son · Y. Choi · J. Kang · Y. Lee · B. Il Je · Y. Park (✉)
부산대학교 생명자원과학대학 원예생명과학과
(Department of Horticultural Bioscience, Pusan National University
(PNU), Miryang, 50464, Korea)
e-mail: ypark@pusan.ac.kr

서 언

수박(*Citrullus lanatus* L., $2n=2x=22$)은 남아프리카 원산의 박과 작물 중 하나로 건생식물 속 *Citrullus Schard*에 속한다. 수박은 오이(*Cucumis sativus* L.), 멜론(*Cucumis melo* L.), 호박(*Cucurbita moschata & maxima*)과 같이 경제적으로 중요한 위치를 차지하고 있는 과채류로서 전 세계 재배 면적의 16%를 차지하고 있다. 국내의 수박 생산액은 9,901억원(2015년 기준)으로 전체 농업 생산액의 2.1%를 차지하며, 종자 생산량은 2017년 1045 kg으로 전체 채소 종자 생산량의 1%를 차지했다(Korean seed Association 2017).

최근 수박 품종육성에 분자마커(Molecular marker) 기술이 적극적으로 활용되고 있다. 이 중 마커이용선발(Marker-assisted selection, MAS)은 특정 형질과 연관된 염색체상 위치나 유전자로부터 개발된 마커를 이용하여 개체선발 효율을 극대화한다(Bertrand and David 2008). 반면 분자마커이용 여교잡(Marker-assisted backcross, MABC)은 반복친으로부터 이입된 염색체 영역을 분자마커를 이용하여 판단하고 선발함으로써 반복친의 특성을 회복하는데 걸리는 여교잡 세대 수와 비용을 단축시키는 육종기술이다(Edwards and Batley 2010). 기존의 여교잡 과정은 많은 시간과 노동력이 소요되는데 반해, MABC를 이용한 품종육성은 2~3번의 여교잡 만으로 전통적 육종방법 이상의 결과를 얻을 수 있는 강점이 있다(Hasan et al. 2015). MABC는 벼와 밀과 같은 식량작물의 신 품종 육성 과정에 사용된 사례가 다수 보고되었지만(Gupta et al. 2010; Neeraja et al. 2007) 수박을 포함한 채소작물에 적용된 사례는 미흡한 실정이다. 분자마커는 재배종의 유전적 다양성이 협소한 작물의 품종식별과 F1 종자의 채종 시 순도 검정에도 효과적으로 이용되고 있다. 정확한 품종식별 기술은 신 품종등록 및 육종가의 지적재산권보호에 필수적이다. 채종된 F1 종자 내 화분유입, 자식주, 종자혼입 등에 의한 비

F1 종자의 비율을 분석하는 순도검정은 상업 육종에 있어 중요한 과정이다(Kiruthika and Padmanabha 2018). 특히 수박은 품종 간 상호 유연관계가 매우 유사한 것으로 알려져 있으며 (Che et al. 2003; Jarret et al. 1997; Kwon et al. 2015; Lee et al. 1996), 따라서 표현형질에 기반한 품종식별 및 F1 품종의 순도검정에는 많은 한계가 있다.

수박의 기존 품종식별 및 순도검정용 마커로 RAPD, SSR, AFLP 등이 활용되었다(Choi et al. 2012; Levi and Thomas 2007). 이들 마커 타입에 비해 최근 차세대 염기서열분석 (Next Generation Sequencing, NGS)에 기반한 전장유전체 분석(Whole genome sequencing, WGS)은 대량의 단일염기다형성(Single nucleotide polymorphism, SNP), 삽입/결실(Insertion/Deletion, Indel)과 같은 염기서열변이 정보를 확보할 수 있는 장점이 있는 반면, 비용적 부담이 큰 단점이 있다. 하지만 이러한 문제점은 최근 Genotyping-by-sequencing (GBS) 기술을 통해 SNP 수를 적당히 줄이고 더 많은 수의 샘플을 분석할 수 있게 됨으로써 극복 되고 있다. GBS 기술은 NGS 기반 방식 중 하나로, 유전체를 제한 효소로 절단하여 얻은 서열을 부분적으로 해독하는 특징을 지니고 있다. GBS는 유전자연관지도 작성, 유전체연관분석(Genome-wide association study, GWAS)을 통한 형질연관 유전자 탐색 및 MAS 및 MABC용 분자마커 선발, 그리고 유전적 다양성 분석을 통한 유전집단 구조 분석 및 품종식별용 마커 개발 등 다양한 분야에서 활발히 사용되고 있다(Elshire et al. 2011; Glaubitz et al. 2014; He et al. 2014).

본 연구는 국내 육종 회사에서 개발된 수박 우량 육성계통 (Elite line) 20종을 대상으로 GBS 기반 SNP를 대량 탐색하고, 이를 이용한 MABC, 품종식별, 그리고 F1 순도검정용 SNP 마커세트를 개발하고자 수행되었다.

재료 및 방법

식물 재료

국내 민간육종회사(SN)의 수박 육종프로그램으로 개발된 F1 품종의 부모계인 20개 계통의 잎 샘플을 분양 받아 실험에 사용하였다.

GBS library 제작 및 NGS

각 계통의 유묘로부터 얻어진 본엽 3~4매를 액체질소 처리 후 막자사발로 마쇄한 뒤, GeneAll®GeneEx™Plant kit (GeneAll Biotechnology Co., LTD, Seoul, Korea) 를 이용하여, 농도 100 µg/µl, 순도 260/280=1.8~2.0, 260/230=1.6 이상인 genomic DNA를 추출하여, 정량화 하였다. 추출된 DNA sample의 adapter ligation을 위해, 제한효소 ApeKI을 이용해 75°C에서

2시간 동안 digestion 하였다. 이후 96-well plate에 1.66x ligase buffer, ATP와 T4 ligase가 포함된 solution을 처리하여 adapter ligation을 수행하였고, DNA sample들을 pooling한 후 QIAquick PCR Purification kit로 clean up 하고, PCR 을 95°C에서 2 min; 98°C에서 30 s, 65°C에서 30 s, 72°C에서 30 s (18 cycles); 72°C에서 5 min의 조건으로 수행하였다. GBS library의 염기서열 분석은 Hiseq2000 (Illumina Inc, San Diego, CA, USA)을 이용하여 수행되었다.

SNP 탐색

Barcode sequence 정보를 이용하여 demultiplexing 후, adapter sequence 제거 및 sequence quality trimming을 하였다. Cutadapt (v.1.8.3) 프로그램을 이용하여 adapter trimming을 수행하였고, SolexaQA (v.1.13) package의 DynamicTrim과 LengthSort 프로그램을 이용하여 trimming 및 quality control을 다음의 조건에 따라 수행하였다; 1) Probability value ≥ 0.05 , 2) Phred score ≥ 20 , 3) minimum length of reads ≥ 25 bp. 전처리 과정을 거친 read들의 mapping은 BWA (0.6.1-r104) 프로그램을 이용하여 ICuGI에서 얻은 표준유전체 정보 (*Citrullus lanatus* cv. 97103)를 바탕으로 수행한 후 SNP 탐색을 위한 SAM format file을 제작하였다. SEEDERS in-house script를 이용하여 SNP validation을 수행한 후, SAMtools(v.0.1.16)을 이용하여 SNP filtering (min.depth ≥ 3 , MAF $>5\%$, missing data $<30\%$)을 한 후 SNP matrix를 작성하였다. 이후 SEEDERS in-house script를 활용하여, reference gene position (ICuGi)을 기반으로 SNP annotation을 수행하고, SNP filtering과 genotyping 조건(reference sequence와 동일한 homozygous - 'AA', 다른 homozygous-'BB', heterozygous - 'AB', missing - '-')에 부합하는 genotype 데이터를 완성하였다.

MABC용 마커 개발을 위한 SNP 선발

MABC용 마커 개발을 위해, GBS 분석결과로 얻은 SNP genotype data 내에서 다형성 정도를 나타내는 척도인 PIC (Polymorphic information content) value가 0.3 이상인 SNP들을 염색체 별로 5개씩 선발하였다. PIC value는 아래의 공식을 이용하여 산출되었다(Nagy et al. 2012).

$$PIC = 1 - \sum_{i=1}^l P_i^2 - \sum_{j=i+1}^l P_i^2 P_j^2$$

P_i 와 P_j 는 각각 전체 l 개수의 allele 중에서 i 번째 및 j 번째 allele의 빈도수를 나타낸다.

주성분분석(PCA) 과 hierarchical clustering

수박 20개 우량 육성 계통의 SNP 정보를 바탕으로 GAPIT (Genome Association and Integrated Tool) 패키지 (R program

v.3.4.2, R Foundation for Statistical Computing, Vienna, Austria) 를 이용하여 수박의 유전형 데이터를 표준화 하고 주성분 분석(PCA)을 실시하였다. 각 계통간 유전적 관계(genetic relationship)을 나타내는 Hierarchical clustering은 VanRaden Kinship matrix (VanRaden 2008)를 통해 분석 및 작성되었다.

결과 및 고찰

GBS 및 SNP 탐색

20개 수박 계통에 대한 GBS 결과, 약 1,100,000천개의 raw read를 얻었다(Fig. 1). 이 중 수박 유전체에 mapping 된 read는 77%로, 각 계통의 평균 mapping region은 약 4,000 Kb로 reference genome coverage의 비율은 2.3%, 각 SNP의 평균 depth는 31.57이었다. Trimming 된 데이터 read의 평균은 517 만개였고, 이들의 평균 길이는 81.91 bp였다(Table 1). SNP filtering을 통해 총 2670개의 SNP를 얻었다. 순화된 수박 180

여개의 수박에서 filtering 후 약 5300개의 SNP를 얻은 경우 (Nimmakayala et al. 2014)와 86개 품종에서 약 1만 4천개의 SNP를 얻은 것(Reddy et al. 2014)과 비교 했을 때, 상대적으로 적은 양인 것을 알 수 있었다. 이는 우리가 20개의 품종만으로 분석을 하였고, 사용된 수박들의 유전적 다양성이 매우 낮은 우량 육종계통(advanced breeding line)이었기 때문인 것으로 보인다. SNP 들은 전반적으로 염색체 모든 영역에서 탐색되었다. SNP의 빈도는 각 염색체마다 차이가 있었지만 교차가 빈번히 일어나는 경향이 있는 염색체 말단 영역에서 대체적으로 높았다. 이는 centromere영역보다 paracentromeric 영역에 높은 빈도로 SNP가 존재하는 결과와 유사하였다 (Park et al. 2018). 각 SNP들은 intron 21%, CDS 26%, intergenic 영역에 53% 비율로 분포되어 있는 것을 확인하였다. 같은 박과인 멜론의 경우 50%의 SNP들이 intergenic 영역에 분포하고, 24%가 intron, 26%가 CDS에 분포한 사례와 비교했을 시에 비슷한 경향을 나타내는 것을 알 수 있었다(Pavan et al. 2017).

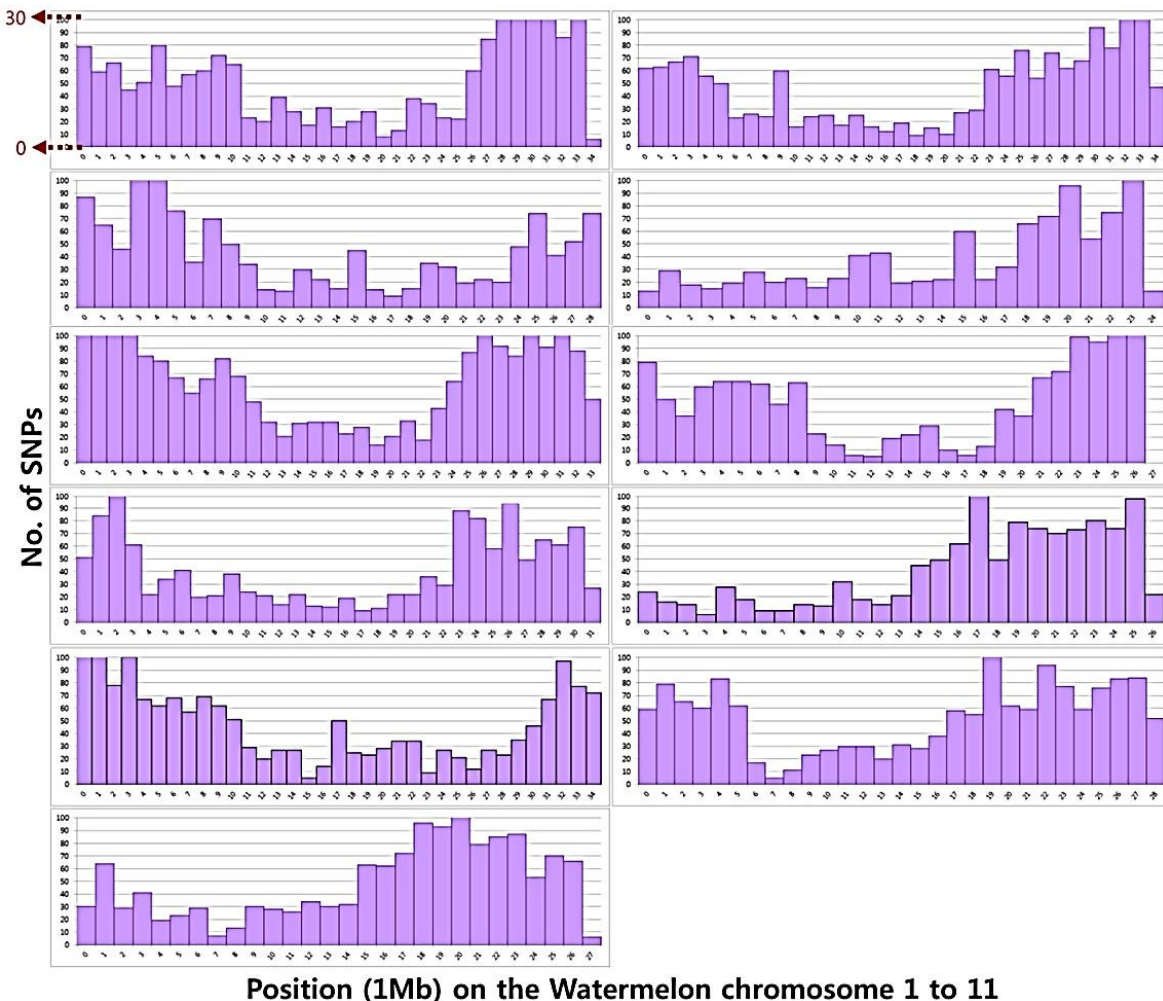


Fig. 1 Distribution and number of SNPs in 11 chromosomes revealed by the Genotype-by-Sequencing (GBS) of 20 elite inbred lines watermelon (*C. lanatus* L.)

Table 1 The statistics result of the Genotype-by-Sequencing (GBS) of 20 elite inbred lines in watermelon (*C. lanatus* L.)

Lines	The total length of trimmed reads	Avg. length of trimmed reads (bp)	Avg. depth of the mapped region
SN1	475,585,923	82.33	33.75
SN2	541,324,508	82.15	37.44
SN3	553,274,594	82.19	38.5
SN4	274,244,171	81.83	24.37
SN5	369,147,743	81.88	28.3
SN101	583,031,239	81.82	37.77
SN102	644,335,167	81.6	40.49
SN103	510,190,422	81.82	34.17
SN104	372,234,863	82.03	30.34
SN105	573,089,105	82.01	37.62
SN106	576,175,917	81.91	40.12
SN107	384,125,892	81.9	30.48
SN601	276,152,796	81.39	23.57
SN602	387,024,316	82.03	30.5
SN603	208,753,526	81.69	19.82
SN608	371,587,725	82.34	29.59
SN610	354,909,132	81.99	29.27
SN611	194,310,572	81.58	20.45
SN612	519,741,302	81.91	38.59
SN613	294,386,077	81.74	26.31

유용 SNP 선발

Filtering 된 2670개의 SNP 중에서 MABC용 분자 마커 개발을 위해 PIC value 가 0.3이상인 1445개의 SNP를 이용하여 분석에 사용하였다. 이들 SNP중에서 각 염색체 별로 물리적으로 비교적 균등한 거리로 분포된 5개 SNP 선발하였다(Fig. 2, Table 2). 총 2670개 SNP의 전체 PIC value의 범위는 0.1에서 0.38이었다. SSR, AFLP나 RAPD처럼 multi-allelic한 특성을 가진 마커들의 경우, PIC value 범위가 0.5에서 1.0인 반면, SNP는 bi-allelic한 특성을 가지고 있기 때문에 일반적으로 PIC value의 범위가 0에서 0.5로 비교적 낮다(Yang et al. 2016). 더욱이 본 연구에서 사용된 계통들이 상용 F1품종의 부모계라는 점에서 원예적 특성이 유사하고, 유전적 다양성이 협소하다는 점은 PIC value의 범위가 낮게 관찰된 원인이라고 판단된다. 소비 시장에 적합한 품종을 얻기 위해, 유전적 거리가 매우 가까운 양친을 사용하여 F1품종을 개발하는 경우가 다수이기에 표현형 기반의 전통적 방법이나 특정 분자마커만을 이용한 품종 지문화에는 어려움이 있었다(Kwon et al. 2015).

주성분 분석과 hierarchical clustering

선발한 55개의 SNP에 대한 MABC, 품종식별 및 순도검정에 활용 가능성을 검토하기 위해 이들 SNP를 이용하여 20개 계통의 유연관계를 PCA 및 Hierarchical clustering 분석을 수행하였다. PCA 결과, 계통간 유전적 변이의 약 52%를 설명하는 주성분 1과 11%를 설명하는 주성분 2를 통해(PC1, PC2) 20개 계통을 총 네 개의 그룹(major group A, B, C, D)으로 분류할 수 있었다(Fig. 3). 특히 주성분 1에 의한 분류결과는 2670개의 SNP를 이용한 PCA 수행 결과(Fig. 4)와 매우 유사하여 염색체 영역별로 선발된 55개의 SNP만으로도 20개 계통의 유전적 다양성을 대표할 수 있으리라 판단되었다. Hierarchical clustering 분석 결과에서는 20개 계통을 두 개의 major group (A, B)으로 분류할 수 있었으며, 각 major group은 다시 두 개의 subgroup (A-1, A-2, B-1, B-2)으로 분류되어 PCA 분석과 유사한 경향을 나타내었다(Fig. 5). 유연관계 분석결과, 사용된 55개 SNP 조합으로 모든 20개 계통을 분류할 수 있어 품종식별 및 F1 품종의 순도검정용 마커로서 활용 가능성을 확인하였다. 유연관계 및 분류 패턴으로 미루어 보아 20개 계통이 market segment에 따른 품종군(Jubilee type, Crimson type, 대과종, 소과종 등)에 따라 4개의 그룹으로 분류된 것으로 추정된다. 본 연구에서 개발된 55개 SNP set은 20개 계통 간 품종식별 및 순도검정용 마커로서 활용 가치가 매우 높으며 이들 우량 계통들을 소재로 한 MABC에도 직접 이용이 가능하리라 본다. 특히 이들 20개 계통이 유전적 다양성이 매우 협소

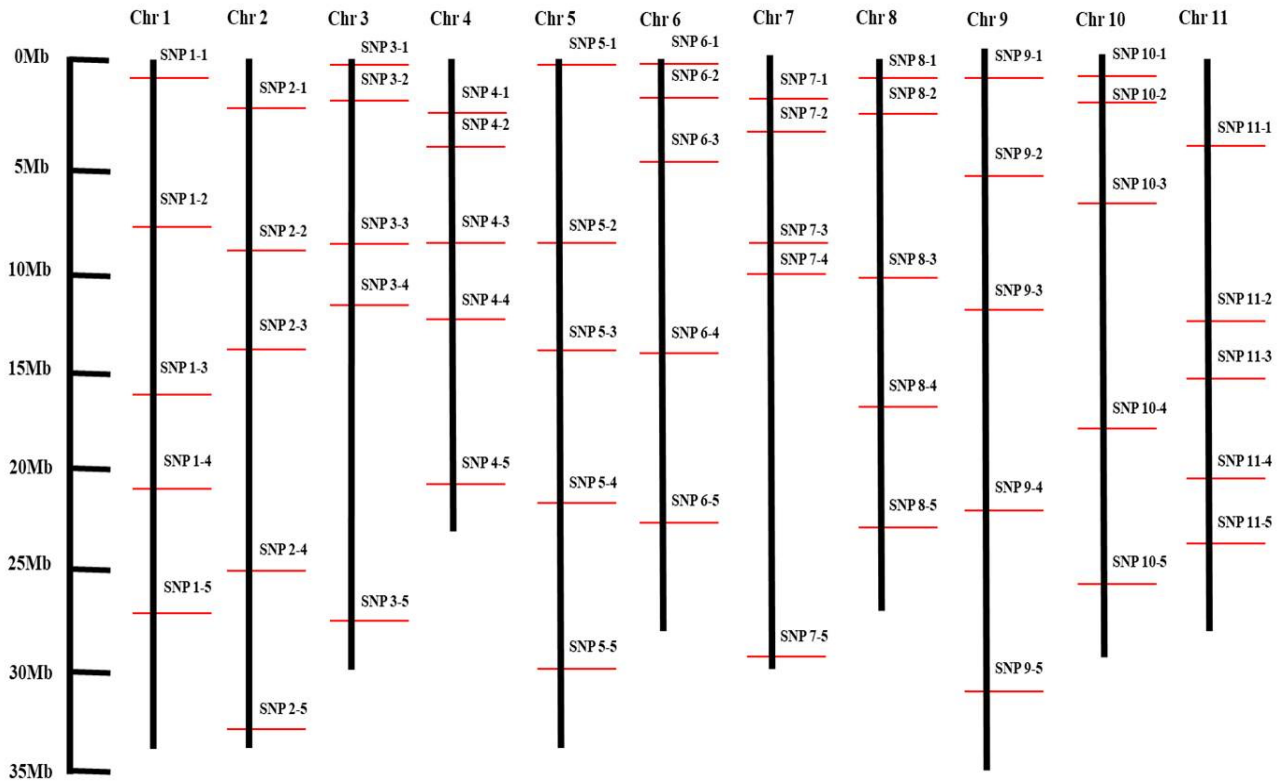


Fig. 2 Genomic location of 55 SNPs selected in this study

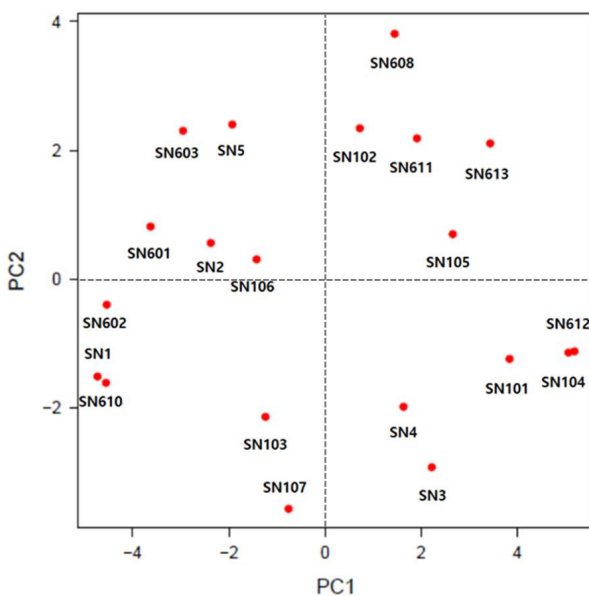


Fig. 3 Principal component analysis (PCA) of 20 elite watermelon inbred lines (*C. lanatus* L.) based on 55 SNPs

한 F1 품종의 부모계임에도 불구하고, 55개 SNP로 모든 계통을 식별할 수 있다는 점은 타회사의 육종소재나 품종에도 충분히 확대 적용이 가능할 수 있음을 시사한다. 단 fluidic chip이나 array와 같은 SNP 자동화 분석 플랫폼(Thomson 2014;

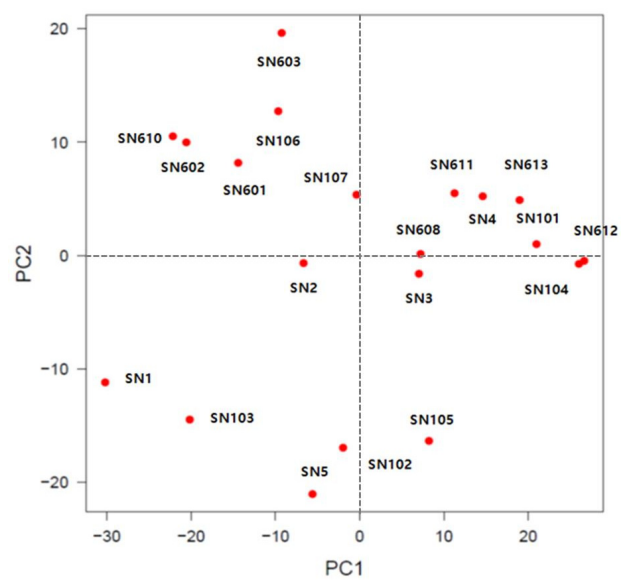


Fig. 4 Principal component analysis (PCA) of 20 elite inbred lines watermelon (*C. lanatus* L.) based on 2670 SNPs

Wang et al. 2009)에 적합한 형태로의 마커전환과 이를 이용하여 보다 폭 넓은 수박육성소재에 검증하는 과정이 향후 필요하리라 본다.

Table 2 Information for 55 SNPs selected in this study

SNP name	Chr	Pos	Ref	Allele	SN 1	SN 2	SN 3	SN 4	SN 5	SN 101	SN 102	SN 103	SN 104	SN 105	SN 106	SN 107	SN 601	SN 602	SN 603	SN 608	SN 610	SN 611	SN 612	SN 613	PIC
SNP 1-1	Chr1	924519	T	T/C	TT	TT	TT	CC	TT	CC	CC	TT	CC	CC	CC	TT	TT	CC	TT	CC	CC	TT	CC	TT	0.38
SNP 1-2	Chr1	7445569	C	C/T	TT	TT	CC	TT	TT	CC	CC	TT	CC	CC	TT	TT	TT	TT	CC	CC	TT	CC	CC	CC	0.38
SNP 1-3	Chr1	16953659	C	C/T	TT	CC	TT	CC	TT	TT	TT	CC	TT	CC	TT	TT	CC	CC	CC	TT	CC	CC	CC	0.38	
SNP 1-4	Chr1	21439311	A	A/G	GG	AG	AG	GG	AG	AA	AG	AG	AG	AG	AG	AA	AG	AA	AG	AG	AG	AG	AG	GG	0.38
SNP 1-5	Chr1	27192632	T	T/A	AA	AA	TT	TT	AA	TT	AA	AA	TT	TT	AA	TT	TT	AA	AA	AA	AA	TT	TT	TT	0.38
SNP 2-1	Chr2	2634833	T	T/C	CC	CC	TT	TT	TT	TT	CC	TT	TT	TT	TT	TT	CC	CC	CC	CC	CC	TT	TT	TT	0.36
SNP 2-2	Chr2	9524934	A	A/G	AG	AG	GG	AG	AG	AG	AG	AA	AA	AG	AA	AG	AA	AA	AG	GG	AG	AG	AG	AG	0.37
SNP 2-3	Chr2	14041748	T	T/A	AA	AA	TT	TT	TT	TT	TT	AA	TT	AA	AA	TT	TT	AA	AA	TT	AA	TT	TT	TT	0.36
SNP 2-4	Chr2	29438535	T	T/G	TT	GG	GG	TT	TT	GG	TT	TT	GG	GG	TT	GG	TT	TT	TT	GG	TT	TT	GG	TT	0.36
SNP 2-5	Chr2	34097022	T	T/C	TT	TT	TT	CC	TT	TT	CC	TT	TT	TT	CC	CC	CC	CC	CC	TT	CC	CC	TT	CC	0.38
SNP 3-1	Chr3	390925	G	G/A	GG	AA	GG	AA	AA	AA	AA	GG	AA	AA	AA	GG	GG	GG	AA	AA	GG	AA	AA	AA	0.35
SNP 3-2	Chr3	1764049	C	C/A	AA	CC	AA	AA	CC	CC	CC	AA	CC	CC	CC	AA	CC	AA	CC	CC	AA	CC	CC	CC	0.35
SNP 3-3	Chr3	8183604	G	G/T	TT	TT	TT	TT	GG	TT	GG	TT	TT	GG	GG	TT	GG	GG	GG	GG	TT	GG	TT	GG	0.38
SNP 3-4	Chr3	12860851	G	G/A	GA	GA	GA	GG	GA	GA	GA	AA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	0.38
SNP 3-5	Chr3	27674495	T	T/A	TT	TT	TT	TT	TT	AA	AA	AA	TA	AA	AA	AA	TT	AA	TT	AA	AA	TT	AA	AA	0.37
SNP4-1	Chr4	2997230	G	G/T	TT	GG	GG	GG	TT	TT	GG	GG	GG	GG	GG	GG	GG	GG	GG	TT	GG	TT	GG	TT	0.33
SNP 4-2	Chr4	4253120	A	A/G	AG	AG	AG	AG	AG	AG	AG	AG	AG	AG	AG	AA	AG	AG	AG	AG	AG	AG	AA	AG	0.37
SNP 4-3	Chr4	8031820	A	A/G	AG	AG	AG	AG	AG	AG	AG	AG	AG	AG	AG	AG	AA	AG	AG	AG	AG	AG	AA	AG	0.37
SNP 4-4	Chr4	13140272	C	C/T	CT	CT	TT	CT	CT	CC	TT	CC	CT	CT	CT	CT	CT	CT	CT	CT	TT	TT	CC	CT	0.38
SNP 4-5	Chr4	22127721	A	A/G	GG	GG	AA	AA	GG	AA	GG	AA	AA	AA	AA	AA	GG	AA	AA	GG	AA	AA	AA	AA	0.33
SNP 5-1	Chr5	340755	G	G/A	GG	GG	AA	AA	GG	AA	AA	GG	AA	GG	GG	AA	AA	GG	AA	GG	GG	AA	AA	AA	0.37
SNP 5-2	Chr5	8411720	A	A/T	TT	TT	TT	TT	AA	TT	AA	TT	TT	AA	TT	TT	AA	AA	AA	AA	AA	AA	TT	AA	0.38
SNP 5-3	Chr5	14683259	T	T/A	AA	AA	TT	TT	TT	TT	TT	AA	TT	TT	AA	AA	AA	AA	AA	TT	AA	AA	TT	TT	0.38
SNP 5-4	Chr5	23248157	C	C/T	TT	CC	CC	TT	TT	CC	TT	TT	TT	TT	CC	TT	TT	CC	CC	CC	CC	CC	TT	CC	0.37
SNP 5-5	Chr5	30020188	T	T/A	TT	TT	TT	TT	TT	AA	TT	TT	AA	AA	TT	AA	AA	TT	AA	AA	TT	TT	AA	AA	0.38
SNP 6-1	Chr6	190608	A	A/C	AA	CC	AA	CC	CC	CC	AA	AA	AA	AA	AA	CC	CC	CC	CC	AA	CC	AA	AA	AA	0.38
SNP 6-2	Chr6	1809686	A	A/G	GG	AA	AA	GG	GG	GG	GG	AA	AA	AA	GG	GG	AA	GG	GG	AA	GG	AA	AA	AA	0.38
SNP 6-3	Chr6	5931428	C	C/T	CT	CT	CT	CT	TT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CC	CT	CT	CT	CT	0.37
SNP 6-4	Chr6	14024887	T	T/C	TC	TC	TC	CC	TC	TC	TC	TC	TC	TC	CC	TT	TC	TC	CC	TC	TC	TC	TC	TC	0.37
SNP 6-5	Chr6	23095502	G	G/T	TT	TT	GG	GG	TT	GG	TT	GG	GG	TT	TT	TT	TT	TT	TT	GG	TT	GG	GG	GG	0.33
SNP 7-1	Chr7	2550025	A	A/C	CC	CC	CC	AA	CC	AA	AA	AA	AA	CC	AA	AA	CC	AA	AA	AA	AA	AA	AA	AA	0.35
SNP 7-2	Chr7	3899830	T	T/C	CC	CC	CC	TT	TT	TT	TT	TT	TT	TT	TT	TT	CC	CC	CC	TT	CC	TT	TT	TT	0.35
SNP 7-3	Chr7	9763127	C	C/T	TT	TT	CC	TT	TT	CC	TT	TT	CC	CC	CC	TT	TT	TT	CC	TT	TT	TT	CC	TT	0.33
SNP 7-4	Chr7	10476443	G	G/C	GG	GG	CC	GG	GG	CC	GG	GG	CC	CC	CC	GG	GG	GG	GG	GG	GG	GG	CC	GG	0.37
SNP 7-5	Chr7	29104809	G	G/A	AA	AA	GG	GG	AA	GG	AA	AA	GG	AA	AA	GG	AA	GG	AA	GG	AA	GG	AA	GG	0.38
SNP 8-1	Chr8	1679300	C	C/T	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CC	CT	TT	0.37
SNP 8-2	Chr8	2435109	C	C/T	CT	CT	CT	TT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CC	CT	CT	CC	CT	0.3
SNP 8-3	Chr8	10618786	A	A/G	AA	AA	AA	AA	AA	AA	AA	GG	AA	AA	AA	AA	GG	GG	GG	AA	GG	AA	AA	AA	0.3
SNP 8-4	Chr8	17220893	T	T/C	TT	TT	TT	TT	CC	TT	TT	TT	CC	TT	TT	TT	TT	TT	CC	TT	TT	TT	CC	CC	0.3
SNP 8-5	Chr8	24388243	A	A/G	AA	AA	GG	AA	AA	AA	AA	AA	GG	GG	AA	AA	AA	AA	AA	AA	AA	AA	GG	GG	0.37
SNP 9-1	Chr9	1640886	C	C/T	CC	TT	TT	TT	CC	TT	TT	TT	TT	TT	CC	TT	CC	CC	CC	CC	CC	TT	TT	CT	0.37
SNP 9-2	Chr9	6392426	A	A/T	TT	TT	AA	AA	TT	AA	AA	AA	AA	AA	TT	TT	TT	TT	TT	AA	TT	AA	AA	AA	0.37
SNP 9-3	Chr9	10426113	T	T/G	GG	TT	TT	TT	GG	TT	TT	GG	TT	TT	GG	GG	GG	GG	GG	TT	GG	TT	TT	TT	0.37
SNP 9-4	Chr9	22905153	A	A/G	GG	GG	GG	AA	AA	AA	AA	AA	AA	AA	GG	GG	GG	GG	GG	AA	GG	AA	AA	AA	0.37
SNP 9-5	Chr9	22905153	A	A/G	GG	GG	GG	AA	AA	AA	AA	AA	AA	AA	GG	GG	GG	GG	GG	AA	GG	AA	AA	AA	0.38
SNP 10-1	Chr10	951602	G	G/T	TT	TT	GG	TT	TT	GG	GG	GG	GG	GG	TT	TT	GG	GG	GG	TT	GG	GG	GG	GT	0.36
SNP 10-2	Chr10	1314394	A	A/G	GG	AA	GG	GG	AA	AA	AA	AA	AA	AA	GG	AA	GG	AA	GG	GG	GG	AA	AA	AA	0.37
SNP 10-3	Chr10	8849318	C	C/T	TT	CT	TT	TT	CT	TT	CT	CT	CC	CT	CT	CT	CT	CT	CT	CT	CT	CT	CC	CT	0.37
SNP 10-4	Chr10	17428398	T	T/C	TT	TT	CC	CC	TT	CC	TT	CC	CC	CC	TT	CC	TT	TT	TT	CC	TT	TT	CC	CC	0.38
SNP 10-5	Chr10	26857342	T	T/A	TT	TT	AA	AA	TT	AA	AA	TT	AA	AA	AA	TT	TT	TT	TT	TT	AA	AA	AA	AA	0.38
SNP 11-1	Chr11	3223095	T	T/C	TT	TC	TT	TC	TC	TT	TC	TC	TC	TC	TC	TC	TT	TC	TC	TT	CC	TC	TC	TC	0.36
SNP 11-2	Chr11	13331339	G	G/A	AA	AA	AA	GG	GG	GG	GG	AA	GG	GG	AA	GG	AA	AA	AA	AA	AA	GG	GG	GG	0.38
SNP 11-3	Chr11	17731991	A	A/C	AA	CC	CC	CC	AA	CC	AA	CC	CC	CC	AA	AA	CC	AA	AA	CC	AA	CC	CC	CC	0.36
SNP 11-4	Chr11	22208506	C	C/A	AA	AA	CC	CC	AA	CC	AA	AA	CC	CC	AA	CC	CC	CC	CC	CC	AA	CC	CC	CC	0.35
SNP 11-5	Chr11	25181007	T	T/C	TT	TT	TT	CC	TT	CC	TT	CC	TT	CC	CC	TT	TT	TT	CC	TT	CC	CC	CC	CC	0.37

Chr, chromosome; Pos, position; Ref, reference allele; PIC, polymorphic information content

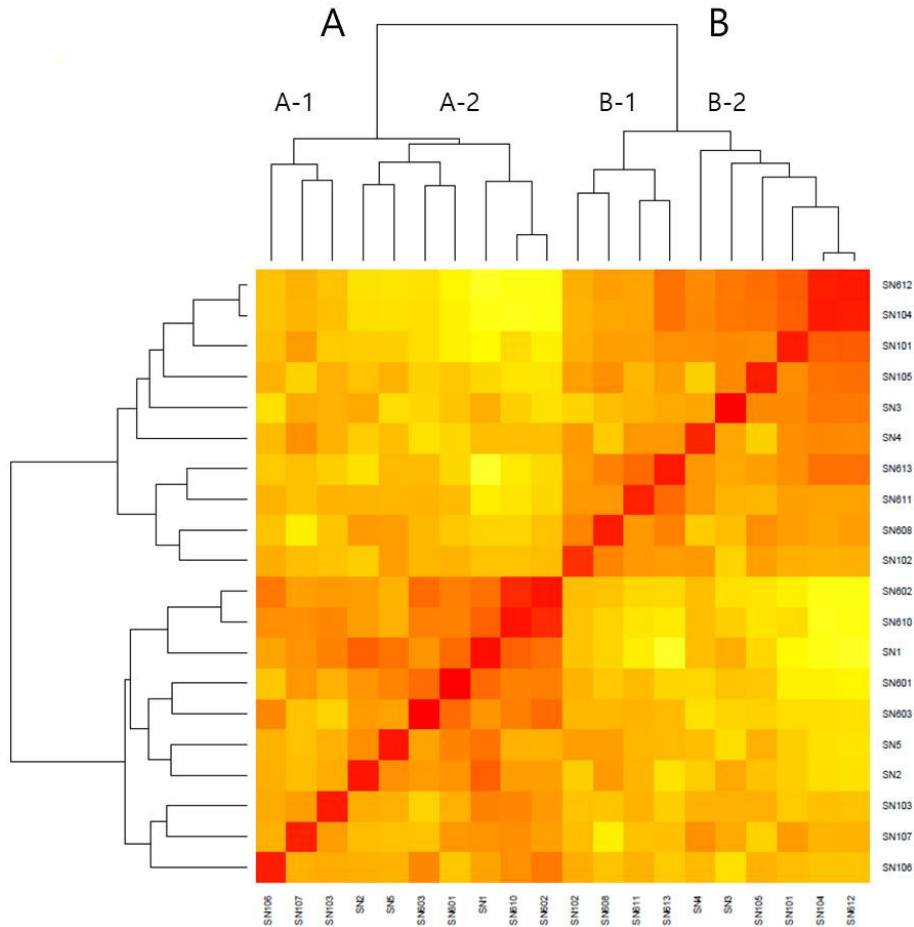


Fig. 5 Hierarchical clustering of 20 elite inbred lines using VanRaden kinship matrix based on the 55 SNPs in watermelon (*C. lanatus* L.)

적 요

본 연구는 국내 육종 회사에서 개발된 수박(*Citrullus lanatus* L.) 우량 육성계통 20종을 대상으로 Genotyping-by-sequencing (GBS) 분석을 통해 품종식별, 순도검정, 그리고 마커이용여 교잡(Marker-assisted backcross, MABC)용 SNP 세트를 개발하고자 수행되었다. GBS 분석 결과 총 1,100,000천개 raw read 중 77%가 수박 유전체에 mapping되었으며 평균 mapping region은 약 4,000 Kb로 2.3%의 genome coverage를 보였다. Filtering을 통해 평균 depth 31.57의 SNP 총 2,670개를 얻었으며, 20개 계통에 대한 이들의 Polymorphic information content (PIC) 값의 범위는 0.1 ~ 0.38 였다. 이 중 PIC 값이 0.3 이상이며 각 염색체 별로 5개씩 균등히 분포된 SNP 총 55개를 최종 선발하였다. 사용된 20개 계통의 유연관계분석을 위해 선발된 55개 SNP를 기반으로 한 주성분 분석(Principle component analysis, PCA) 결과 주성분 1 (52%)과 주성분 2 (11%)를 기준으로 4개의 그룹으로 분류 되었으며 각 계통 간 유전자형에 따른 뚜렷한 식별이 가능하였다. 계층적 군집화(Hierarchical clustering) 분석에서도 PCA에서와 유사한 분류양상을 관찰할 수 있었다. 따라서 본 연구에서 개발된 SNP 세트는 적용

가능성이 검증된 20개 계통뿐 만 아니라 향후 다양한 수박 육종소재 및 품종에 대한 품종식별, F1 순도검정 및 MABC에 활용될 수 있으리라 기대된다.

사 사

이 논문은 부산대학교 기본연구지원사업(2년)에 의하여 연구되었음.

References

- Bertrand CY, David J (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philos Trans R Soc Lond B Biol Sci* 363(1491): 557-572. doi: 10.1098/rstb.2007.2170
- Che K, Liang C, Wang Y, Jin D, Wang B, Xu Y, Zhang H (2003) Genetic assessment of watermelon germplasm using the AFLP technique. *Science* 38(1):81-84
- Choi YM, Hwang JH, Kim KW, Kang JS, Choi YH, Son BG, Park

- YH (2012) Application of EST-SSR Marker for purity test of watermelon F1 cultivars. *J Agric Life Sci* 46(4):85-92
- Edwards D, Batley J (2010) Plant genome sequencing: applications for crop improvement. *Plant Biotechnol J* 8(1):2-9. doi: 10.1111/j.1467-7652.2009.00459.x
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6(5):1-10. doi: 10.1371/journal.pone.0019379
- Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, Buckler ES (2014) TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PLoS One* 9(2). doi: 10.1371/journal.pone.0090346
- Gupta PK, Langridge P, Mir RR (2010) Marker-assisted wheat breeding: present status and future possibilities. *Mol Breed* 26(2):145-161. doi: 10.1007/s11032-009-9359-7
- Hasan MM, Rafii MY, Ismail MR, Mahmood M, Rahim HA, Alam MA, Latif MA (2015) Marker-assisted backcrossing: a useful method for rice improvement. *Biotechnol Biotechnol Equip* 29(2):237-254. doi: 10.1080/13102818.2014.995920
- He J, Zhao X, Laroche A, Lu ZX, Liu HK, Li Z (2014) Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Front Plant Sci* 5: 484. doi: 10.3389/fpls.2014.00484
- Jarret RL, Merrick LC, Holms T, Evans J, Aradhya MK (1997) Simple sequence repeats in watermelon (*Citrullus lanatus* (Thunb.) Matsum. & Nakai). *Genome* 40(4):433-441. doi: 10.1139/g97-058
- Kiruthika S, Padmanabha BV (2018) Analysis of hybrid purity in watermelon using microsatellite marker in comparison with field GOT. *Int J Res Eng Technol* 5(4):761-763
- Korean seed association (2017). Retrieved from <http://www.kosaseed.or.kr>
- Kwon YS, Hong JH, Kim DH, Kim DH (2015) Use of microsatellite markers derived from genomic and expressed sequence tag (EST) data to identify commercial watermelon cultivars. *HST* 33(5):737-750. doi: 10.7235/hort.2015.15045
- Lee SJ, Shin JS, Park KW, Hong YP (1996) Detection of genetic diversity using RAPD-PCR and sugar analysis in watermelon [*Citrullus lanatus* (Thunb.) Mansf.] germplasm. *Theor Appl Genet* 92(6):719-725. doi: 10.1007/BF00226094
- Levi A, Thomas CE (2007) DNA markers from different linkage regions of watermelon genome useful in differentiating among closely related watermelon genotypes. *HortScience* 42(2):210-214
- Nagy S, Poczai P, Cernák I, Gorji AM, Hegedus G, Taller J (2012) PICcalc: an online program to calculate polymorphic information content for molecular genetic studies. *Biochemical Genetics* 50(9-10):670-672. doi: 10.1007/s10528-012-9509-1
- Neeraja CN, Maghirang-Rodriguez R, Pamplona A, Heuer S, Collard BCY, Septiningsih EM, Mackill DJ (2007) A marker-assisted backcross approach for developing submergence-tolerant rice cultivars. *Theor Appl Genet* 115(6):767-776. doi: 10.1007/s00122-007-0607-0
- Nimmakayala P, Levi A, Abburi L, Abburi VL, Tomason YR, Saminathan T, Reddy UK (2014) Single nucleotide polymorphisms generated by genotyping by sequencing to characterize genome-wide diversity, linkage disequilibrium, and selective sweeps in cultivated watermelon. *BMC Genomics* 15(1). doi: 10.1186/1471-2164-15-767
- Park GR, Kim JY, Jin BK, Yang HB, Park SW, Kang SC, Park YH (2018) Genome-wide sequence variation in watermelon inbred lines and its implication for marker-assisted breeding. *HST* 36(2):280-291
- Pavan S, Marcotrigiano AR, Ciani E, Mazzeo R, Zonno V, Ruggieri V, Ricciardi L (2017) Genotyping-by-sequencing of a melon (*Cucumis melo* L.) germplasm collection from a secondary center of diversity highlights patterns of genetic variation and genomic features of different gene pools. *BMC Genomics* 18(1):1-10. doi: 10.1186/s12864-016-3429-0
- Reddy UK, Nimmakayala P, Levi A, Abburi VL, Saminathan T, Tomason YR, Karol A (2014) High-resolution genetic map for understanding the effect of genome-wide recombination rate on nucleotide diversity in watermelon. *G3 (Bethesda)* 4(11):2219-2230. doi: 10.1534/g3.114.012815
- Thomson MJ (2014) High-throughput SNP genotyping to accelerate crop improvement. *Plant Breed Biotechnol* 2(3): 195-212. doi: 10.9787/PBB.2014.2.3.195
- VanRaden PM (2008) Efficient methods to compute genomic predictions. *J Dairy Sci* 91(11):4414-4423. doi: 10.3168/jds.2007-0980
- Wang J, Lin M, Crenshaw A, Hutchinson A, Hicks B, Yeager M, Ramakrishnan R (2009) High-throughput single nucleotide polymorphism genotyping using nanofluidic dynamic arrays. *BMC Genomics* 10(1):1-13. doi: 10.1186/1471-2164-10-561
- Yang X, Ren R, Ray R, Xu J, Li P, Zhang M, Kilian A (2016) Genetic diversity and population structure of core watermelon (*Citrullus lanatus*) genotypes using DArTseq-based SNPs. *Plant Genet Resour* 14(03):226-233. doi: 10.1017/S1479262115000659