

GBS (Genotyping-by-sequencing) 알아보기

◆ 분자마커와 육종

‘농업의 반도체’ 혹은 ‘Golden seed’라고 불리는 종자주권을 둘러싸고 세계 각국의 경쟁이 뜨겁다. 경쟁력의 핵심은 시장의 요구에 빠르게 부응할 수 있는 우수한 신품종을 만드는 기술과 품질을 유지하는 것이다. 이를 해결하기 위한 가장 유용한 도구는 분자마커 개발 및 이의 적절한 활용이다.

분자마커(Molecular marker)는 식물 육종 연구에 유용한 도구로 사용되어 왔다. 교배·선택 방식 육종법을 이용하여 우수 품종을 육성할 때 분자마커와 같은 첨단 육종기술을 이용하면 육성기간 단축뿐만 아니라 조기선택을 통해 작물의 유지 및 관리에 소요되는 많은 경비를 줄일 수 있다. 이와 같은 실질적인 효과가 입증되면서 국외뿐만 아니라 국내 육종 현장에도 분자마커 활용 요구도가 급증하고 있는 추세다.

◆ 분자마커의 발달

분자마커는 지난 수십년동안 끊임없이 발전해 왔다. 대표적인 종류는 RFLP (restriction fragment length polymorphism)¹⁾, RAPD (randomly amplified polymorphic DNA)²⁾, STS (Sequence tagged sites)³⁾, AFLP (amplified fragment length polymorphism)⁴⁾, SNP (single nucleotide polymorphism)⁵⁾ 등이 있다.

최근 NGS(next generation sequencing) 장비의 발달로 유전체 내 구조변이(SNP, In/del) 탐색은 규모나 속도 면에서 급진적 발전을 이루고 있다. 특히 SNP 마커는 다른 마커들에 비해 유전체 내 발생빈도가 가장 빈번할 뿐만 아니라 high-throughput 기술에 보다 적합한 특성, 즉 탐색에서 분석에 이르는 전 과정이

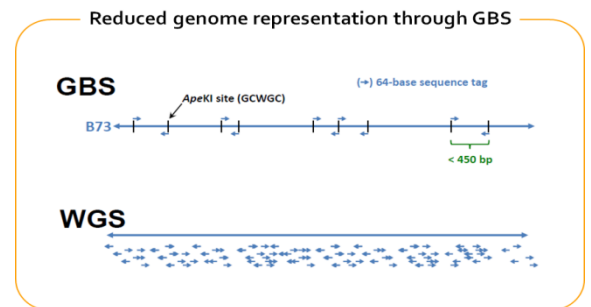
자동화 될 수 있다는 점이 다른 마커들보다 활용성을 높이고 있다.

NGS를 이용하여 SNP를 탐지하는 방법은 크게 유전체 전체를 읽는 Resequencing 방식(WGS)과 대상영역을 좁혀서 시퀀싱하는 방법(GBS, RAD-seq 등)이 있다. WGS 방식은 우수 자원이나 원종계통에 보편적으로 사용되고 있다. 유전체 모든 영역에 존재하는 SNP를 모두 확보 할 수 있는 장점이 있는 반면, 유전체 전체를 Sequencing 해야 한다는 비용적 부담이 있다. 또한 과도하게 많은 SNP 중 실제로 사용 가능한 SNP를 선별하는 일도 상당한 어려움을 일으키고 있다.

이로 인해 SNP 수를 적절한 수준으로 줄이면서 더 많은 수의 샘플이나 계통을 분석하고자 하는 요구가 발생하였다. 이에 GBS(Genotyping-by-sequencing) 기술이 향후 분자마커 개발을 가속화 할 수 있는 촉매제로 시장 내 거론되고 있다.

◆ GBS 기술의 특징 (차별성)

GBS 기술은 Resequencing과 같은 NGS 기반 방식이지만, 제한 효소로 genomic DNA를 절단하여 절단된 주변 서열만 부분적으로 해독할 수 있는 특징을 가진다. 유전체 전체를 Sequencing 하지 않고 부분적으로 sequencing 해서 유전체 전체 정보를 확보할 수 있기에 비용적인 부담을 확 줄일 수 있게 되었다.



Computational Biology Service Unit, Cornell University

< GBS와 WGS 비교 >

기존 WGS (Whole genome sequencing) 방식과 어떤 차별성을 지니고 있는가? 고추를 WGS 방식과 GBS 방식으로 Sequencing 한다고 가정해보자. 고추는 12개 chromosome으로 구성되며, 유전체 크기는 약 3Gb이다.

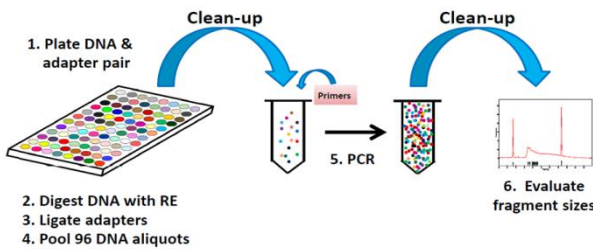
WGS 방식의 경우, 유전체 전체를 시퀀싱 하기에 3Gb 전체를 해독해야 하며, 결과의 정확성을 위해

- 1) RFLP는 가장 널리 사용되는 hybridization 기반한 분자마커로, 각 개체에서 DNA 절편 크기의 양상에 따른 차이를 나타내는 제한 효소를 기반으로 하는 마커이다.
- 2) RAPD는 중간 및 종 내의 유전적 근원관계를 연구하는데 많이 활용되고 있다.
- 3) 양쪽 끝 염기서열을 알고 있는 RFLP probe나 18~24 bp 크기의 염기서열을 알고 있는 primer를 가지고 어느 특정 유전체 부위 절편을 증폭시키는데 응용된다.
- 4) AFLP법의 원리는 인식 부위가 많지 않은 제한 효소로 절단된 DNA의 단편들에 adaptor를 붙인 다음, 표식 부위의 DNA 염기서열을 바탕으로 작성한 primer를 PCR의 primer로 사용하여 특정 제한 효소단편을 증폭시켜 그 산물에서 얻어지는 band 차이의 유무를 비교하는 것이다.
- 5) SNP는 개체간의 DNA에 존재하는 단일 염기 차이로서 DNA 다형성 중에서 가장 많이 존재하는 형태이다.

depth를 최소 10x 이상은 확보해야 한다. 따라서 유전체 해독 시, 샘플 당 30Gb를 Sequencing 하게 된다.

GBS 방식은 어떤 결과를 가져다줄까? GBS 방식의 경우 제한효소를 선정하여 유전체 전체를 절단한다. 기본 200 - 500bp 크기로 절단된 DNA 조각을 이용해 시퀀싱을 실시하게 된다. 제한효소의 종류에 따라 다르지만 유전체 전체의 10~30% 정도의 염기서열을 확인하게 된다. 샘플 당 생산하는 시퀀스 양이 줄어들면, 그에 따라 샘플수를 늘릴 수 있다는 결론이다. Depth 및 샘플 수의 증가는 연구의 깊이와 확장성을 더해주는 중요한 요소이다. 특별히 유전학을 기반으로 수십 혹은 수백 개체의 통계처리를 실시하는 경우에 매우 유용하다.

GBS 96-plex Protocol (<http://www.maizegenetics.net/gbs-overview>)



< GBS의 원리 > ⁶⁾

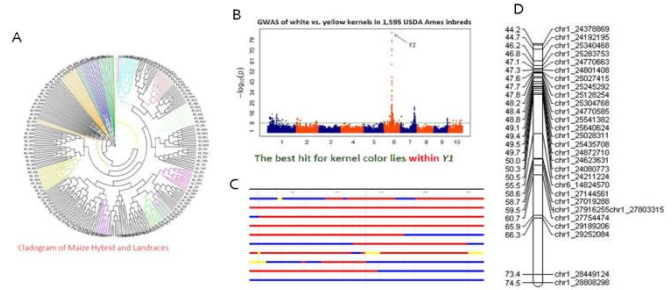
◆ GBS 연구의 시초

GBS 연구는 어디서 시작되었을까? 2011년 영국의 코넬(Cornell) 대학의 Sharone E. Mitchell, Robert J. Elshire이 옥수수를 소재로 GBS 연구를 진행하였고 논문을 발간하였다. 이후 보리, 밀, 콩을 소재로 국내외 연구진에 의해서 연구 성과가 보고되고 있다. 보리, 밀은 주요 작물이면서, 거대하고 복잡한 지놈의 대표적인 작물이므로 이는 연구 타당성을 입증할 신뢰할 만한 성과이다. 이후 다양한 종을 이용하여 적절한 제한효소를 찾고, 적용한 GBS 연구는 매우 빠르게 확장되고 있다.

◆ GBS 기술의 응용연구 사례들

GBS 기술은 여러 가지 목적에 맞게 개선되고 있을 뿐만 아니라 더욱 다양한 분야에서 시도되고 있다. 대표적인 사례를 살펴보면 GWAS, 연관지도 작성, MAB 마커 선발, 형질연관도 분석, 군집 구조 분석,

계통분석 등을 들 수 있다.



Dr. Ram Sharma, Rajneesh Paliwal, Roamy et al. 2013; Genome Biology

< GBS 응용분석 >

◆ GBS 기술이 육종과정에 어떻게 적용될까?

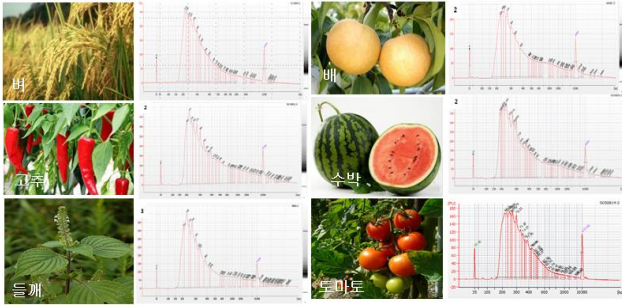
신품종 육성을 위해서는 우선 우수한 형질을 포함하고 있는 다양한 유전자원을 확보하고 평가해야 한다. 또한 교배 부모본으로 사용할 수 있는 우수친을 유지하고 평가하는 것이 필요하다. 수천개의 SNP로 수십 개체를 평가할 수 있는 GBS 기술은 phylogeny 혹은 kinship 분석 등을 통해 유전적 다양성 정도 혹은 유전적 고정 정도, 혼입여부 등을 평가하는 적용 사례가 꾸준히 증가할 것으로 예상되고 있다.

또한 다른 육종의 핵심역량 중 하나로, 다양한 형질 연관 분자마커 개발을 언급하고 있다. '육종 현장 내 다양한 요구(needs)에 맞는 분자마커를 빠르게 개발하여 공급할 수 있는가?' 이 질문에 대해 GBS 기술이 완벽한 모범 답안이 될 수는 없지만, 저렴한 비용과 분석의 단순성 등 GBS 기술이 지닌 장점으로 QTL의 연관지도 작성을 수주일 내에 가능하게 하는 등 분자마커 개발 속도에 박차를 가할 수 있는 촉매제라고 말할 수 있다. GWAS에도 활발하게 적용하고 있어, 향후 형질 연관 분자마커의 개발 속도가 더욱 빨라질 것으로 예상된다. MAS를 이용한 개체선발 뿐 만 아니라 MABC를 이용한 여교잡 선발도 매우 중요한 분야이다. GBS를 이용한 MABC도 매우 유망한 분야로 평가된다.

◆ 국내 필요성 증가 및 주요 질문

GBS 기술의 다양한 응용 가능성 및 가격 경쟁력으로 인해 시도하기를 원하는 연구자들이 많아지고 있다. 씨더스는 벼, 들깨, 수박, 토마토, 배, 고추 등 다양한 소재의 GBS 분석 경험을 통해 양질의 결과를 얻었고, 연구 결과에 대한 연구자들의 주요 질문과 그에 대한 답변은 다음과 같다.

6) Robert J. Elshire et al. (2011) A Robust, Simple Genotyping-by-sequencing(GBS) Approach for High Diversity Species. Plosone. 6(5). e19379



< GBS 분석 사례 >

1) 어떤 제한효소를 사용하는 것이 좋을까요?

상당히 다양한 제한효소가 개발되어 있다. 물론 각 유전체가 가지고 있는 염기서열의 특징뿐 아니라 얼마나 많은 부위를 시퀀싱 할지 등, 목적에 따라 다양하게 선택할 수 있다. 가장 대표적인 제한효소는 ApeKI (GCWGC), PstI (CTGCAG), EcoT22I (ATGCAT)이 대표적이다. 그 외의 다양한 제한효소는 TASSEL group의 자료를 참조하면 도움이 된다.

2) Genome 전체를 충분히 확인할 수 있나요?

제한효소의 종류와 GBS 라이브러리 작성에 미치는 여러 요인에 의해 영향을 받을 수 있다. 토마토, 배, 포도, 고추, 콩, 들깨 등의 시료에서 ApeKI를 이용해 실시한 경우, 상당히 성공적인 결과를 얻었다. 예를 들어, 토마토 자손집단 96개체를 대상으로 한 GBS 분석에서, 토마토 유전체의 75,929 영역(96개체 평균)에 약 7.70 X (depth)의 genome coverage를 확보하였다. 고정이 잘된 작물의 경우 충분히 분석이 가능하지만 정확도를 더 높이거나 heterozygosity가 높은 작물을 대상으로 할 경우에는 96개체 보다는 논문에서 48개체를 사용하는 이율을 추측할 수 있다.

3) 특정 염색체 영역에 집중 분포하진 않나요?

사과 자손집단 96개체의 GBS 분석 시, 17개 염색체에 시퀀싱 리드(read)가 고르게 mapping 되어 분포함을 확인하였다. 토마토 GBS 분석에서는 염색체 끝 즉, 유전자 밀집지역 (euchromatic region)에 조금 높은 비율의 시퀀싱 리드가 분포하였지만 전체 염색체 영역을 확인할 수 있었다.

4) SNP 수는 충분히 확보하나요?

목적이 무엇인지가 상당히 중요하다. 물론 근연관계가 멀수록 SNP의 수가 많아질 것이다. 일반적으로 유전체의 5~20% 정도를 확인하기 때문에 GBS를 통해서도 충분한 수의 SNP를 확보할 수 있다. 얻어진 SNP 중 사용 가능한 SNP를 정제하거나 결손된 영역을 어떻게 처리할지가 상당히 중요한 부분이다. 실제로 GBS 기반의 토마토 여교잡 개체 선발을 위한 부모친간 SNP 선별, 사과 Linkage map 작성을 위한 SNP 선별 결과에서 염색체 전체에 고루 분포하는 약 2,000 개 이상의 신뢰도 높은 SNP를 확보하여 우수한 결과를 얻었다.

5) 시퀀싱 reads는 충분히 사용되나요?

Hiseq2500을 이용하여 1 lane을 시퀀싱 할 경우 대략 36Gb의 데이터를 얻을 수 있다. 이곳에 96개체를 넣을지 48개체를 넣을지 실험자가 결정하게 된다. 여러 실험 결과를 살펴보면 GBS 라이브러리의 품질이 상당히 많은 영향을 미치는 것을 알 수 있었다. 상당히 많은 시퀀싱 리드를 잃어버린 경우가 많았다. Multiplex index에 시퀀싱 에러가 포함되어 해당 샘플로 할당할 수 없었기 때문이다. 물론 생물정보학적 기술로 개선하여 상당히 많은 리드를 구분하도록 하였으나 초기 GBS 라이브러리 품질이 최종 시퀀싱 리드 품질까지 상당한 영향을 미친다는 점을 유념해야 한다. 예로 토마토 자손집단 96개체 GBS 분석 시, 시퀀싱 리드의 50% 정도를 사용하였다. 즉 예상하는 양보다 적은 리드를 얻을 수 있다.

◆ GBS가 가능하다는 것은?

2011년 처음 GBS가 보고되고 활용되면서 이제는 국내에서도 여러 연구팀에 의해 적용되고 있다. 물론 좋은 결과들이 보고되면서 관심이 증가되고 더욱 많은 영역에 적용되고 있는 것은 기쁜 일이다. NGS가 유전체 영역뿐만 아니라 유전과 육종 등의 다양한 영역에 활용되어 실질적인 도움을 주고 있고, 이러한 변화는 향후 분자마커의 개발 및 다양한 유전연구에 획기적인 속도를 가해 변화될 미래를 상상하며 기대하게 한다.