

213006-
05-5-
CG100

채소작물의 종자개발을 위한 육종 특화 통합 DB 구축 및 운영 최종보고서

2022

농림축산식품부

농림식품기술기획평가원

보안 과제(), 일반 과제(O) / 공개(O), 비공개()발간등록번호(O)

Golden Seed 프로젝트사업 2단계 최종보고서

발간등록번호

11-1543000-003929-01

채소작물의 종자개발을 위한 육종 특화 통합 DB 구축 및 운영

2022. 3. 25.

프로젝트연구기관 / 충남대학교 산학협력단
세부프로젝트연구기관 / 충남대학교 산학협력단

농림축산식품부

(전문기관) 농림식품기술기획평가원

제 출 문

농림축산식품부 장관 귀하

본 보고서를 “Golden Seed 프로젝트 사업”(기간 : 2017. 1. ~ 2021. 12.)채소작물의 종자 개발을 위한 육종 특화 통합 DB 구축 및 운영 프로젝트의 최종보고서로 제출합니다.

2022. 3. 25.

프로젝트연구기관명 : 충남대학교 산학협력단 (대표자) 정종률 (인)
세부프로젝트연구기관명 : 충남대학교 산학협력단 (대표자) 정종률 (인)
참여기관명 : 충남대학교 산학협력단 (대표자) 정종률 (인)



프로젝트연구책임자 : 임용표
세부프로젝트연구책임자 : 임용표
참여기관책임자 : 임용표

국가연구개발사업의 관리 등에 관한 규정 제18조에 따라 보고서 열람에 동의합니다.

보고서 요약서

| | | | | | |
|------------------|-------------------|------------------------------------|-------------------------------|----------------------|---|
| 과제고유번호 | 213006-05-5-CG100 | 해당단계 연구기간 | 2017.01.01.~ 2021.12.31 | 단계구분 | 2/2 |
| 연구사업명 | 단위사업 | Golden Seed 프로젝트사업 | | | |
| | 사업명 | GSP채소종자사업단 | | | |
| 프로젝트명 | 프로젝트명 | 채소작물의 종자개발을 위한 육종 특화 통합 DB 구축 및 운영 | | | |
| | 세부프로젝트명 | 채소작물의 종자개발을 위한 육종 특화 통합 DB 구축 및 운영 | | | |
| 프로젝트책임자 | 임용표 | 해당단계 참여연구원 수 | 총: 120명 내부: 115명 외부: 5명 | 해당단계 연구개발비 | 정부: 2,892,000천원 민간: 천원 기관: 750,000천원 계: 3,642,000천원 |
| | | 총 연구기간 참여연구원 수 | 총: 202명 내부: 197명 외부: 5명 | 총 연구개발비 | 정부: 5,153,000천원 민간: 천원 기관: 750,000천원 계: 5,903,000천원 |
| 연구기관명 및 소속부서명 | 참여기업명 충남대학교 | | | 참여기업명 | |
| 국제공동연구 위탁연구 | 상대국명: 연구기관명: | | | 상대국 연구기관명: 연구책임자: | |

※ 국내외의 기술개발 현황은 연구개발계획서에 기재한 내용으로 같음

| | |
|----------------------|--------|
| 연구개발성과의 보안등급 및 사유 | 보안등급() |
|----------------------|--------|

9대 성과 등록·기탁번호

| 구분 | 논문 | 특허 | 보고서 원문 | 연구시 설·장비 | 기술요약 정보 | 소프트 웨어 | 화합물 | 생명자원 | | 신품종 | |
|-------------|----|----|-----------|-------------|------------|-----------|-----|----------|----------|-----|----|
| | | | | | | | | 생명 정보 | 생물 자원 | 정보 | 실물 |
| 등록·기탁 번호 | 29 | 12 | | | | | | | | | |

국가과학기술종합정보시스템에 등록된 연구시설·장비 현황

| 구입기관 | 연구시설·장 비명 | 규격 (모델명) | 수량 | 구입연월일 | 구입가격 (천원) | 구입처 (전화) | 비고 (설치장소) | NTIS 등록번호 |
|------|--------------|-------------|----|-------|--------------|-------------|--------------|--------------|
| | | | | | | | | |
| | | | | | | | | |

요약(연구개발성과를 중심으로 개조식으로 작성하되, 500자 이내로 작성합니다)

- 채소작물의 종자개발을 위한 육종 특화 통합 DB 구축 및 운영
- 유전체 정보 통합 DB 구축으로 종자 개발에 있어 분자유종의 효율화를 지원
- 각 프로젝트별(채소별)로 생산되는 통합해 제공함으로써 프로젝트간 시너지 효과 유발
- 종자 개발에 필요한 유전체, 유전분석을 수행하고 이를 통해 목표 형질 개체 선발 및 계통선발, 목표 형질 접근 방법에 효율화

보고서 면수

206

요약문

| | | | | | |
|------------------------------------|--|----------------------|-----------------------|---------------------------|------------------|
| 연구의 목적 및 내용 | <ul style="list-style-type: none"> - 채소작물의 종자개발을 위한 육종 특화 통합 DB 구축 및 운영 - 유전체 정보 통합 DB 구축으로 종자 개발에 있어 분자유종의 효율화를 지원 - 각 프로젝트별(채소별)로 생산되는 통합해 제공함으로써 프로젝트간 시너지 효과 유발 - 종자 개발에 필요한 유전체, 유전분석을 수행하고 이를 통해 목표 형질 개체 선발 및 계통선발, 목표 형질 접근 방법에 효율화 | | | | |
| 연구개발성과 | <p>○ 학술적 성과로 논문 29편(SCI: 28편, 비SCI:1편), 특허 출원 5건, 특허 등록 7건이 있다.</p> <ol style="list-style-type: none"> 1.육종 특화 데이터베이스 구축을 위한 기반 정보 수집 및 업데이트 <ul style="list-style-type: none"> - 수박, 고추, 파프리카, 무의 육종특화 데이터베이스 구축 및 정보 수집 - 염기서열, 전사체, 변이 분석 정보 수집 및 가공 - 수박 옹성불임 전사체 정보를 활용한 Indel 마커 개발 2.신규 분자 마커 개발을 위한 유전연구 <ul style="list-style-type: none"> - 전사체 정보를 활용한 마커 생산 방식 개발 및 네트워크 연구 - 배추 노화 형질에 대한 GWAS 분석 및 유용형질 QTL 분석 - 배추 유용형질 관련 신호전달경로/유전자 네트워크의 분석 기법 개발 - 저항성과 이병성 품종의 유전자 발현 프로파일, 네트워크 Pathway 기반 차이 분석 3.채소 작물의 유용 형질관련 분자마커의 범용화 <ul style="list-style-type: none"> - 수집단의 변이 정보 생산 방식 개선 및 annotation 방식의 확립 - 배추 표준 유전체 기반 정보의 생산 및 마커 개발 파이프라인 구성 및 평가 - 유용 표현형 관련 염기서열의 범용 마커로의 전환 4.분자마커의 high-throughput selection을 위한 마커 이전 서비스 <ul style="list-style-type: none"> - 마커 개발 및 high-throughput system 용 마커 개발 및 범용성 검증 - 카로티노이드 페스웨이 상의 변이 확보를 위한 파이프라인의 개발 5.종자산업 발전 및 마케팅 전략 구축을 위한 기반 연구 <ul style="list-style-type: none"> - 국제협력사업 관계 기관들과의 협력체 구축을 통한 채소 종자 수출 활성화 기반 조성 - 종자산업 발전 및 마케팅 전략 구축을 위한 기반연구 - 국내 종자산업 현황과 국제경쟁력 제고 방안 연구 | | | | |
| 연구개발성과의 활용계획 (기대효과) | <ol style="list-style-type: none"> 1.기술적 측면 <ul style="list-style-type: none"> - NGS를 통한 대량 분자마커 발굴을 이용하는 육종 기술의 진화 - 데이터베이스 구축, 형질에 관련 유전자 분석, 분자마커 개발 소요 시간 단축, 효율적 분자유종 시스템 구축 - 체계화된 분자마커 지원을 위한 사용자 중심의 데이터베이스 및 분석 지원 - 작물에 대한 분자마커 육종지원시스템 개발을 통해 분자마커를 이용하는 육종기술의 체계화 및 타 작물에 대한 유사 시스템 기술지원이 가능함 - 작물 육종 활동에 대한 QTL, GWAS 분석 방식의 접목은 다양한 표현형 중 특정 형질에 대한 신규 분자마커의 개발을 집단 작성 및 전개에 소요되는 시간과 비용을 크게 절감할 수 있음 2.경제적/산업적 측면 <ul style="list-style-type: none"> - 복잡 형질의 유전 현상을 시스템적으로 이해하고 예측 가능한 육종 계획을 세움으로 육종 기간을 단축하고 맞춤형종(breeding-by-design)을 구현할 수 있는 기반을 제공함. - 공통된 작물을 육종하는 여러 육종가에게 동일한 분자마커의 사용을 유도함으로써 육종 연한의 감소를 이룩함. - 다양한 작물 종을 대상으로 연구하는 분자유종가에게 공통의 분자마커 사용을 촉진함으로써, 신품종 육성 가속화로 인한 육종 기간 단축과 분자마커 개발 및 파생 기술을 이용한 산업화를 통하여 다가오는 바이오경제의 경쟁력을 확보할 수 있을 뿐만 아니라 관련 민간기업의 활성화에 기여할 수 있음. - 종자 개발의 효율성 극대화로 시장에서 경쟁력 확보 및 수출 증대 기대. - 빠르게 변화하는 시장의 소비자 요구에 부합하는 품종 개발의 효율성 증진 - 국내외 종자산업의 현황 분석과 마케팅전략 분석을 통하여 각 종자회사들의 국제경쟁력 제고 방안에 대한 효율적인 품종 개발 및 마케팅 방안 제공 | | | | |
| 국문핵심어 (5개 이내) | 데이터베이스 | 분자표지 | 분자유종 | 대용량 분석 시스템 | 종자산업 |
| 영문핵심어 (5개 이내) | Database | Molecular Markers | Molecular breeding | High-throughput system | Seed Industry |

< 목 차 >

| | |
|-----------------------------|-----|
| 1. 연구개발과제의 개요 | 6 |
| 2. 연구수행 내용 및 결과 | 9 |
| 3. 목표 달성도 및 관련 분야 기여도 | 192 |
| 4. 연구결과의 활용 계획 등 | 193 |
| 붙임. 참고 문헌 | 195 |

<별첨 1> 연구개발보고서 초록

<별첨 2> 연구성과 활용계획서

<별첨 3> 자체평가의견서

제 1 장 연구개발과제의 개요

제1절 연구개발의 목적

1. 최종목표

- 채소작물의 종자개발을 위한 육종 특화 통합 데이터베이스(DB)의 구축 및 운영
 - 유전체 정보 통합 DB 구축으로 종자 개발에 있어 분자유종의 효율화를 지원함.
 - 각 프로젝트별(채소별)로 생산되는 통합해 제공함으로써 프로젝트간 시너지 효과 유발
 - 종자 개발에 필요한 유전체, 유전분석을 수행하고 이를 통해 목표 형질 개체선발 및 계통선발, 목표 형질 접근 방법에 효율화를 꾀하고자함

2. 연구개발의 필요성

- 배추, 무, 고추는 벼와 함께 우리나라 3대 주요 작물
 - 위의 작물은 파프리카, 수박과 함께 우리나라를 비롯한 전 세계적으로 소비시장은 크지만 국내 종자 생산량은 급속한 감소 추세를 보이는 반면 종자 수입은 갈수록 증가함
- 표준 유전체의 염기서열 공개 및 이용 가능
 - 현재 세계 각국의 연구를 통해 배추, 무, 고추, 수박에 대한 유전체 염기서열 정보가 공개되어 이용이 가능한 상황
 - 표준 유전체의 완성과 동시에 수종 혹은 수집 종의 주요 육종 라인 혹은 중요 유전자원을 중심으로 re-sequencing 데이터의 생산이 기대됨.

| 작물 | 종명 | 유전체 크기 | 공개일 | 수행기관 |
|---------|--------------------------|----------------|---------------------------|-------------------------------|
| 배추 | <i>Brassica rapa</i> | 529 Mb | Nature Genetics; 2011.09. | MBGP |
| 무 | <i>Raphanus sativus</i> | 573 Mb | DNA research; 2014.05. | Kazusa DNA Research Institute |
| 수박 | <i>Citrullus lanatus</i> | 425 Mb (draft) | Nature Genetics; 2012.12. | IWGI |
| 고추/파프리카 | <i>Capsicum annuum</i> | 2700 Mb | Nature Genetics; 2014.01. | 서울대 |

- 치열한 종자시장 환경의 새로운 대안은 유전체 정보를 이용한 신속한 형질관련 분자마커 개발
 - 현재까지 형질관련 분자마커 개발은 단일 유전자에 의해 조절되는 형질 중심으로 문헌 정보를 이용하여 개발함
 - 세대단축을 위하여 여교잡 선발(Marker-assisted backcrossing: MAB)을 위한 교배조합별 이용 가능한 genome-wide SNP 마커 개발
- 육종과정에서 수집되는 작물의 정보는 개별적 혹은 수기로 작성되고 있어 시너지효과를 도출하고 있지 못함

- 유전체 분석기술(염기서열 해독 및 분석)은 비약적으로 발전하고 있으나 확보된 작물의 기본 자원의 활용은 미비해 개선이 대대적으로 필요한 상황
- 분자유종 지원을 위한 유전체 정보 통합 DB 구축 필요
 - 배추의 경우, 표준유전체 염기서열이 해독되고 DH, RIL, F₂, BC 집단 및 유전지도 확보 등 풍부한 유전자원이 수집되었으나, 연구단계로의 진입이 미비함.
- 유용 표현형 관련 분자마커의 개발을 위해서는 다양한 형질을 대상으로 한 유전 기초 연구가 필수적
 - 유용 표현형을 대상으로 한 mapping 집단의 작성과 표현형 검정으로 특정 표현형에 관련된 연관지도와 QTL을 얻을 수 있으며 이를 GWAS 분석 결과와 결합하여 신규 마커 개발 및 선발에 필요한 양질의 정보를 얻을 수 있음
- DB 구축 및 정보 공유만이 아닌 수집된 정보의 재생산 및 활용을 통해 신규 분자마커 개발의 가속화 및 선진화를 기대할 수 있음
 - 수집한 표현형 별 분자마커를 유전체 상에 비교적 고르게 분포하는 것으로 평가되는 SNP 마커로 전환하면 범용마커로서의 활용 가능성이 예상됨
 - 계통별 re-sequencing 데이터로부터 얻은 SNP 정보와 유용 표현형 데이터를 GWAS 방식에 따른 분석을 통해 특정 표현형에 대한 SNP마커 후보의 대량산출에 기반한 신규 마커의 개발 및 특허 등록 등과 같은 고부가가치 산물의 창출이 가능
 - High-throughput 검정의 도입은 기존의 분자마커 검정에 소요되는 비용과 노동력을 절감하고 신속한 분자마커 선발 가능성을 바탕으로 하여 앞으로 신품종 육성에 필수적인 플랫폼이 될 것으로 기대
- 형태 관련 표현형 중심의 육종을 넘어 병충해 저항성 및 기능성 2차대사 산물 집적에 초점을 맞춘 육종을 지원할 기반이 필요
 - 작물의 수확시 기대되는 수량 증가에 중점을 둔 육종에서 병저항성 및 2차대사 산물 중심의 기능성 육종으로의 이행은 농가의 생산비를 절감하고 시장으로 유통되는 농산물의 다변화를 이끌어 높은 수준의 경제적 효과를 야기할 것으로 기대
- 우리 종자산업의 세계화를 위해 산업 현황을 검토하고 이의 발전 및 마케팅 전략 구축을 위한 기반 연구가 필요한 시기

3. 연구개발 범위

- 데이터베이스 구축을 위한 기반 정보 수집 및 제공
 - 채소 작물의 목표 형질을 연구할 수 있는 육종 소재 및 형질 조사
 - 육종 소재의 활용성을 증대시키기 위해 육종과정에서 수집된 형질 조사 및 이미지를 데이터베이스화
 - 얻어진 데이터를 전산화하여 모든 데이터를 누적하고 통합분석의 기초를 확보함
 - 기존에 구축한 배추와 수박 DB의 업데이트와 고추, 파프리카, 무의 육종 특화 DB를 구축

- **목표 형질 관련 유전자, 분자마커 및 문헌조사**
 - 5대 채소 작물별로 목표형질을 설정하여 문헌조사 및 기연구 자료를 바탕으로 핵심 유전자 염기서열 및 기능 등 수집
 - 개발된 분자마커를 수집된 정보와 통합하여 활용성을 증대
- **채소 작물의 표준유전체, 비교유전체, 전사체 결과 수집**
 - 국제적으로 공개된 데이터베이스로부터 데이터 수집
 - 중요 형질을 포함한 계통의 경우 직접 생산 방식을 통해 데이터 확보
- **분자유종 활성화를 위한 특화된 데이터베이스 구축**
 - 목표형질 선발용 분자마커(MAS) 데이터베이스
 - 교배 조합별 이용 가능한 분자 마커 검색을 위한 데이터베이스 구축
 - Genetic map 및 형질 관련 사진, 해당 육종 소재, 유전양상 등 육종에 직접적으로 필요한 결과를 형질별 작물별 분류하여 이용할 수 있도록 함
- **유용 표현형 관련 분자마커 개발을 위한 유전 기초연구**
 - 수집한 계통 중 유용한 표현형을 나타내는 유전자원을 부모본으로 활용하여 집단을 양성하고 이를 통해 유전현상 해석 및 기초연구를 수행하여 신규 분자마커 개발에 이용
- **유용 표현형 지정 분자마커 개발을 위한 GWAS(Genome Wide Association Study)**
 - 수집한 계통의 re-sequencing 데이터로부터 얻은 변이 정보에 유용 표현형 관련 데이터를 GWAS방식으로 처리하여 기존의 유전연구 결과를 재확인하고 유용 표현형에 연관된 SNP를 대량 산출하여 신규 분자마커 개발에 이용하기 위한 시스템의 구축
- **채소 작물의 유용 표현형 연관 범용마커의 개발과 육종 공동체로의 보급**
 - 데이터베이스 업데이트를 통한 유용 표현형 및 분자마커의 지속적 수집
 - 수집 및 생산한 채소 작물의 계통별 re-sequencing 데이터를 활용하여 기존 분자마커를 범용 SNP 마커로 전환하고 보유한 계통에 적용하여 마커의 검정력 평가를 수행
- **여교잡 선발용 분자마커(MAB) 데이터베이스**
 - 비교 유전체를 통한 작물별 주요 교배친의 genome-wide SNPs 데이터 확보하고 이를 DB화
- **채소 작물 육종 지원 High-throughput 체계의 구성**
 - 데이터베이스 구축을 위한 자료의 수집과 생산과정에서 얻은 표현형 연관 변이 정보를 활용
 - 유용 표현형별 SNP set을 제작하여 육종 기업 및 주체에 보급하여 부가가치 창출과 국내 분자유종의 가속화 및 선진화를 도모
 - 종자산업진흥센터와의 연계를 통한 분자마커 분석시스템의 조기 구축을 위한 기초연구수행
- **종자산업 발전 및 마케팅 전략 구축을 위한 기반 연구**
 - 종자산업 현황을 검토하고 종자업계의 미래 방향을 검토할 수 있는 자료의 수집
 - 종자산업발전을 위한 기획연구 테마를 선정하고 이에 대한 심층연구를 수행
 - 종자산업의 글로벌화를 위한 구체적인 산업화 전략을 구상하고자 함

제 2 장 연구개발과제의 내용 및 결과

제1절 수박 분자유종 활성화를 위한 특화된 데이터베이스 구축 및 자료의 업데이트

1. 수박 유전자원의 염기서열 정보의 수집

수박(watermelon, 학명: *Citrullus lanatus* (Thunb.) Matsum. & Nakai)은 과실을 식용 부위로 이용하는 박과 작물(Cucurbitaceae)에 속하는 채소이다. 이는 전 세계적으로 거대한 시장을 형성하고 있다. 2013년에는 NGS(Next generation Sequencing)를 활용하여 425 Mb의 genome size와 11개의 chromosome에 대한 전장 유전체가 공개되었다. 골든씨드프로젝트 1단계(2013 ~ 2016) 작업에서는 채소육종 특화 데이터베이스(<http://vegetable.or.kr/>)의 기본 플랫폼에 NCBI에서 다운로드한 수박 20여개 계통의 Whole genome re-sequencing 데이터와 수박 표준 계통인 97103과 야생종 수박의 과피(rind)와 과육(flesh)에서 시기별 RNA-seq 데이터로부터 정보를 재생산한 결과를 업로드 하였다.

금년도의 작업에서도 수박의 육종특화 데이터베이스내의 가용 정보를 업데이트하기 위해 타 대학의 연구진 및 NCBI 등지에서 새로운 raw 데이터를 공여 및 수집하여 이를 재분석하였다. 이 과정에서 중앙대학교 생명공학대학 생명자원공학부 이금표 교수님 실험실로부터 수박 10개 계통에 대한 whole genome re-sequencing 데이터를 공여받았다. 또한 NCBI의 SRA(ncbi.nlm.nih.gov/sra)에서 전사체 관련 업로드 파일을 검색하여 수박의 저온처리 관련 8개의 전사체 raw RNA-seq data를 확보하였다. 수집된 raw sequence 데이터는 자체 분석 서버로 이관하여 구축된 분석 파이프라인을 통해 가공되어 수박의 마커개발 및 유전분석에 활용할 수 있는 정보의 생산에 활용되었다.

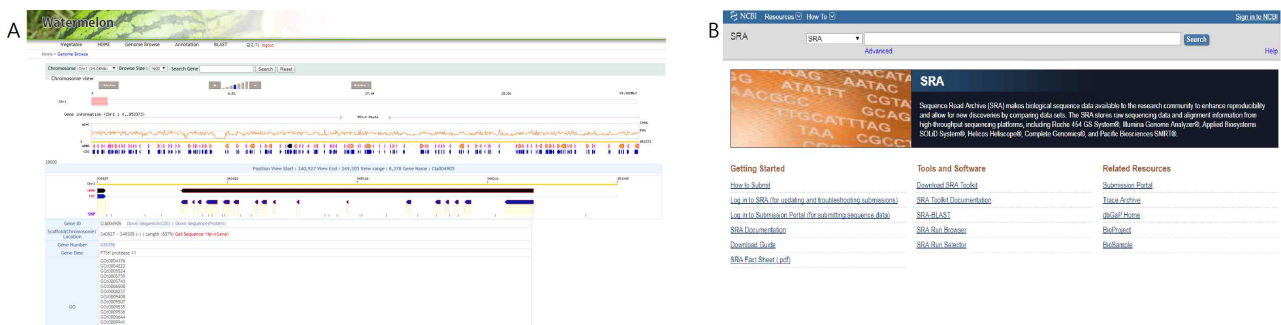


그림 1. 구축된 수박육종 특화 데이터베이스 (A: <http://168.188.15.201/vwatermelon/index.php/vege/browse>)와 공개된 NGS 정보에 접근이 가능한 NCBI SRA (B: <https://www.ncbi.nlm.nih.gov/sra>)

가. 수박 10개 계통의 Whole genome sequence(WGS) 정보 및 SNP 데이터의 구성

(1) 중앙대학교 공여 수박 10개 계통의 sequence 정보와 분석을 위한 전처리

중앙대학교로부터 공여받은 10개 수박 계통에 대한 raw sequence를 자체 분석 서버로 이관하여 fastqc를 통해 sequence quality를 확인할 수 있었다. 1차 QC(quality check) 결과를 정리하여 10개 계통이 reference genome의 assembled size인 353.5Mb의 30X 이상으로 시

퀵싱이 된 것을 확인할 수 있었으며 일부 계통의 raw sequence에 library 구성에 사용된 adapter 서열 등에 의해 염기서열상의 초기 15 bp의 GC contents가 일정하지 않은 것을 확인하여 Trimmomatic을 활용한 1차 Quality trimming 및 fastqc를 통한 2차 QC를 수행하였다 (표 1).

표 1. 중앙대학교 공여 수박 10개 계통 raw sequence 기본 정보

| Accession | Sequence file name | Sequencing method | Phred quality score | Number of sequence reads | Read length | Sequencing coverage (X) |
|-----------|--------------------|--------------------------|---------------------|--------------------------|-------------|-------------------------|
| 1902 | 1902_1.fastq | Sanger / Illumina 1.9 | Q33 | 40573780 | 151 | 34.66 |
| | 1902_2.fastq | Sanger / Illumina 1.9 | Q33 | 40573780 | 151 | |
| 1903 | 1903_1.fastq | Sanger / Illumina 1.9 | Q33 | 35219423 | 151 | 30.09 |
| | 1903_2.fastq | Sanger / Illumina 1.9 | Q33 | 35219423 | 151 | |
| 1906 | 1906_1.fastq | Sanger / Illumina 1.9 | Q33 | 35147285 | 151 | 30.03 |
| | 1906_2.fastq | Sanger / Illumina 1.9 | Q33 | 35147285 | 151 | |
| 1909 | 1909_1.fastq | Sanger / Illumina 1.9 | Q33 | 37154905 | 151 | 31.74 |
| | 1909_2.fastq | Sanger / Illumina 1.9 | Q33 | 37154905 | 151 | |
| 1912 | 1912_1.fastq | Sanger / Illumina 1.9 | Q33 | 44370592 | 151 | 37.91 |
| | 1912_2.fastq | Sanger / Illumina 1.9 | Q33 | 44370592 | 151 | |
| 1913 | 1913_1.fastq | Sanger / Illumina 1.9 | Q33 | 37230802 | 151 | 31.81 |
| | 1913_2.fastq | Sanger / Illumina 1.9 | Q33 | 37230802 | 151 | |
| 6200 | 6200_1.fastq | Sanger / Illumina 1.9 | Q33 | 67256082 | 101 | 38.43 |
| | 6200_2.fastq | Sanger / Illumina 1.9 | Q33 | 67256082 | 101 | |
| 6201 | 6201_1.fastq | Sanger / Illumina 1.9 | Q33 | 64063631 | 101 | 36.61 |
| | 6201_2.fastq | Sanger / Illumina 1.9 | Q33 | 64063631 | 101 | |
| 6202 | 6202_1.fastq | Sanger / Illumina 1.9 | Q33 | 65985958 | 101 | 37.71 |
| | 6202_2.fastq | Sanger / Illumina 1.9 | Q33 | 65985958 | 101 | |
| 6203 | 6203_1.fastq | Sanger / Illumina 1.9 | Q33 | 90924829 | 101 | 51.96 |
| | 6203_2.fastq | Sanger / Illumina 1.9 | Q33 | 90924829 | 101 | |

(2) 기존 수집 수박 계통의 SNP 정보의 재분석 및 신규 공여 계통의 변이 정보 생산

기존에 수박육종 특화 데이터베이스에 1차적으로 입력된 수박 계통들에 대한 변이 정보는 2013년 출판된 수박의 표준 유전체 서열이 공개된 논문의 supplementary 데이터를 활용하여 구성이 되었다. 본 작업에서 계통별 SNP 정보를 구성을 위해 활용된 수박의 reference genome(계통명 97103)은 과거 수집 계통과 금년도 공여 받은 10개 계통에 모두 같이 적용되었다. 그러나 과거 수집 계통의 sequence로부터 SNP 정보를 생산할 때 사용한 알고리즘 및 프로그램의 버전 정보가 현행 사용되고 있는 방식에서 상당 수준 개선이 되었기 때문에

정확한 수집 계통을 이루는 집단의 SNP의 개발을 위해서는 보완된 방식의 SNP 구성 방식을 채택할 필요가 있었다. Re-sequencing 방식을 기반으로 한 수많은 작물의 SNP 개발 사례에서 BWA-mem 방식을 통한 Read sequence data의 reference genome으로의 alignment와 이 결과로부터 변이 데이터 추출이 samtools의 mpileup 기반으로 이루어지고 있다는 경향을 확인하였다. 따라서 이 방식을 채택하여 과거 수집 계통과 신규 공여 계통의 SNP 정보의 생산 과정을 재구축하고 이를 수행하였다.

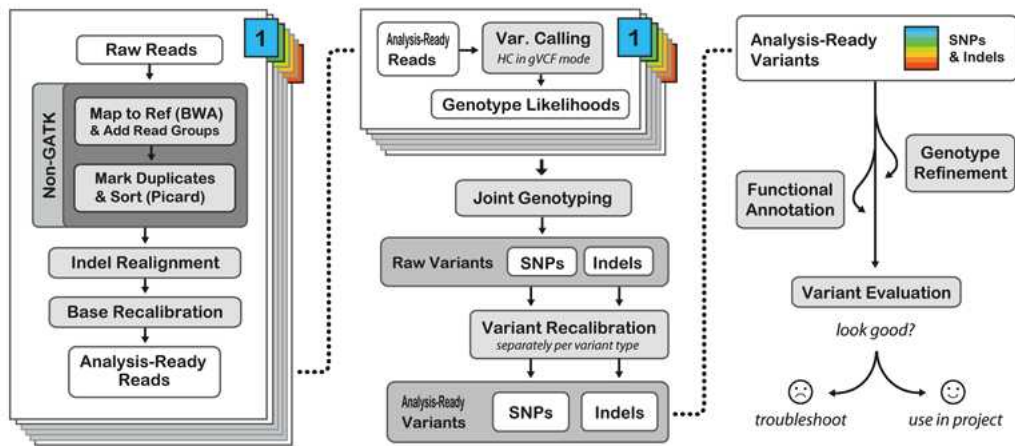


그림 2. BWA-mem 방식 기반의 계통별 SNP 생산 및 집단의 SNP matrix 구성을 위한 일반적인 개요도

나. 수집 계통으로부터의 SNP 정보 생산

(1) 기존 수박 수집 계통의 정보

이전 단계의 수박 육종특화 데이터베이스 구성을 위한 수박의 계통별 SNP 정보를 얻기 위해 해당 정보를 NCBI SRA에서 수집하였다. 해당 자료는 수박의 Whole genome sequencing project에서 계통 97103을 reference로 삼아 re-sequencing가 수행된 20계통의 genome re-sequencing 파일의 ftp상의 경로에 접근하여 얻을 수 있었다. 이를 분석 서버로 이관하여, GATK기반으로 구축한 파이프라인을 통해 가공하여 수집된 계통들의 reference genome인 97103에 대한 변이 정보를 생산하였으며 수집된 계통과 그 sequence 정보는 표 2와 같다.

표 2. 기존 수집한 NCBI SRA 유래 20개 수박 계통의 raw sequence 기본 정보

| Accession name | SRA ID | Sequencing method | Phred quality score | Number of sequence reads | read length | Sequencing coverage (X) |
|----------------|-----------|-----------------------|---------------------|--------------------------|-------------|-------------------------|
| JX-2 | SRR494422 | Sanger / Illumina 1.9 | Q33 | 34493590 | 45 | 9 |
| JXF | SRR494424 | Sanger / Illumina 1.9 | Q33 | 34493590 | 45 | 9 |
| RZ-901 | SRR494425 | Sanger / Illumina 1.9 | Q33 | 34778793 | 45 | 18 |
| | | | Q33 | 34778793 | 45 | |
| XHBFGM | SRR494426 | Sanger / Illumina 1.9 | Q33 | 26449711 | 45 | 7 |
| Black Diamond | SRR494427 | Sanger / Illumina 1.9 | Q33 | 26449711 | 45 | 7 |
| Calhoun Gray | SRR494428 | Sanger / Illumina 1.9 | Q33 | 28252539 | 45 | 7 |
| Sugarlee | SRR494429 | Sanger / Illumina 1.9 | Q33 | 28252539 | 45 | 7 |
| Sy-904304 | SRR494430 | Sanger / Illumina 1.9 | Q33 | 35086723 | 75 | 9 |
| | | | Q33 | 35086723 | 45 | |
| RZ-900 | SRR494431 | Sanger / Illumina 1.9 | Q33 | 34183058 | 45 | 18 |
| | | | Q33 | 34183058 | 45 | |
| PI482271 | SRR494432 | Sanger / Illumina 1.9 | Q33 | 34183058 | 45 | 9 |
| PI189317 | SRR494433 | Sanger / Illumina 1.9 | Q33 | 34776783 | 90 | 9 |
| PI500301 | SRR494434 | Sanger / Illumina 1.9 | Q33 | 34776783 | 90 | 9 |
| PI595203 | SRR494441 | Sanger / Illumina 1.9 | Q33 | 34620800 | 45 | 18 |
| | | | Q33 | 34620800 | 45 | |
| PI249010 | SRR494444 | Sanger / Illumina 1.9 | Q33 | 32233478 | 90 | 8 |
| PI248178 | SRR494439 | Sanger / Illumina 1.9 | Q33 | 32233478 | 90 | 8 |
| PI482276 | SRR494440 | Sanger / Illumina 1.9 | Q33 | 26180760 | 45 | 14 |
| | | | Q33 | 26180760 | 45 | |
| PI482303 | SRR494443 | Sanger / Illumina 1.9 | Q33 | 27742489 | 45 | 7 |
| PI296341-FR | SRR494446 | Sanger / Illumina 1.9 | Q33 | 27742489 | 45 | 14 |
| | | | Q33 | 28840419 | 45 | |
| PI482326 | SRR494437 | Sanger / Illumina 1.9 | Q33 | 28840419 | 45 | 7 |
| JLM | SRR494438 | Sanger / Illumina 1.9 | Q33 | 15370989 | 88 | 4 |

다. 수박 표준 유전체 97103 기준의 수집 및 공여 계통으로부터의 SNP 정보 생산

수집한 계통 별로 BWA - mem 방식을 통한 genome sequence read의 수박 표준 유전체에 대한 alignment를 수행하였다. 그리고 alignment 결과내의 duplicate read data의 제거 및 indel 영역의 보정을 위한 GATK pipeline의 수행 결과 각 계통별로 SNP 정보를 얻을 수 있었다. 또한 각 계통 별로 homozygous한 위치만을 선별하여 이에 대한 위치정보를 하나로 합친 후 중복된 위치 정보는 단일화하여 수집한 집단 내에서 중복이 없는 SNP 위치 정보에 대한 목록을 구성 후 joint variant calling을 위한 input으로 활용하였다. joint variant calling 이후 신뢰성이 있는 SNP 만을 남기기 위한 filtering을 진행하였으며 그 결과로 산출된 SNP list에 대한 VCF file을 구성하였다. 최종적으로 표준 유전체에서의 각 SNP position에서 alternative allele의 발생이 야기하는 효과를 평가하기 위해 snpEff 소프트웨어를 활용한 SNP annotation을 수행하였다.

(1) 수박의 수집 집단의 계통별 SNP 산출 결과

구축된 pipeline을 통해 동일 알고리즘이 적용된 계통별 SNP를 얻을 수 있었다. 단 JLM은 paired-end type이 아닌 single-end type으로 라이브러리의 구성 및 sequencing이 진행되어 현재 사용한 pipeline에 일괄적용을 할 수가 없어서 작업 과정에서 제외되었다. JLM을 제외한 나머지 계통들에 대해서는 SNP calling 이후 homozygous한 allele만을 filtering하는 과정을 리눅스 운영체제에 기본적으로 제공되는 sed 명령어를 통해 진행하였다. 이를 통해 계통별 homozygous SNP를 갖는 vcf 파일을 구성 및 계통별 SNP의 밀도를 확인할 수 있었다 (표 3).

표 3. 수박 29계통의 homozygous SNP의 산출 결과

| Lines | Source | Total homozygous SNPs | SNP density against reference genome (cv. 97103) | Resequencing coverage |
|---------------|----------|-----------------------|--|-----------------------|
| 1902 | CAU | 160,073 | 45.28 SNPs/100kb | 34.66 |
| 1903 | CAU | 141,876 | 40.13 SNPs/100kb | 30.09 |
| 1906 | CAU | 178,797 | 50.58 SNPs/100kb | 30.03 |
| 1909 | CAU | 122,363 | 34.61 SNPs/100kb | 31.74 |
| 1912 | CAU | 205,346 | 58.09 SNPs/100kb | 37.91 |
| 1913 | CAU | 186,067 | 52.64 SNPs/100kb | 31.81 |
| 6200 | CAU | 245,359 | 69.41 SNPs/100kb | 38.43 |
| 6201 | CAU | 196,881 | 55.69 SNPs/100kb | 36.61 |
| 6202 | CAU | 197,458 | 55.86 SNPs/100kb | 37.71 |
| 6203 | CAU | 170,746 | 48.30 SNPs/100kb | 51.96 |
| Black_Diamond | NCBI SRA | 115,379 | 32.64 SNPs/100kb | 6.73 |
| Calhoun_gray | NCBI SRA | 111,319 | 31.49 SNPs/100kb | 7.19 |
| JX-2 | NCBI SRA | 54,505 | 15.42 SNPs/100kb | 8.78 |
| JXF | NCBI SRA | 91,092 | 25.77 SNPs/100kb | 8.78 |
| PI189317 | NCBI SRA | 71,590 | 20.25 SNPs/100kb | 8.85 |
| PI248178 | NCBI SRA | 90,623 | 25.64 SNPs/100kb | 8.21 |
| PI249010 | NCBI SRA | 115,570 | 32.69 SNPs/100kb | 8.21 |
| PI296341-FR | NCBI SRA | 1,250,541 | 353.76 SNPs/100kb | 14.41 |
| PI482271 | NCBI SRA | 143,331 | 40.55 SNPs/100kb | 8.7 |
| PI482276 | NCBI SRA | 1,533,140 | 433.70 SNPs/100kb | 13.33 |
| PI482303 | NCBI SRA | 1,703,448 | 481.88 SNPs/100kb | 7.06 |
| PI482326 | NCBI SRA | 1,295,137 | 366.38 SNPs/100kb | 7.34 |
| PI500301 | NCBI SRA | 164,362 | 46.50 SNPs/100kb | 8.85 |
| PI595203 | NCBI SRA | 61,985 | 17.53 SNPs/100kb | 17.63 |
| RZ-900 | NCBI SRA | 118,042 | 33.39 SNPs/100kb | 17.64 |
| RZ-901 | NCBI SRA | 112,139 | 31.72 SNPs/100kb | 17.71 |
| Sugarlee | NCBI SRA | 95,931 | 27.14 SNPs/100kb | 7.19 |
| Sy-904304 | NCBI SRA | 169,563 | 47.97 SNPs/100kb | 8.93 |
| XHBFMG | NCBI SRA | 96,530 | 27.31 SNPs/100kb | 6.73 |

(2) 수박의 수집 집단의 SNP annotation

각 계통별로 homozygous allele만 다루는 vcf 파일을 구성하고 SNP의 위치 정보만을 추출하였다. 이를 sorting 한 후 중복되는 지역은 하나만 남긴 리스트를 구성하여 수집한 29개 수박 계통 집단 내의 unique한 SNP의 위치 리스트를 구성하였다. 이 리스트를 기준으로 29개 계통에 대한 joint variant calling을 수행 결과를 filtering하여 최종적인 수박 수집 집단에 대한 SNP matrix를 구성할 수 있었다. 그리고 snpEff를 통해 SNP의 효과 및 분포 양상을 분석할 수 있었다.

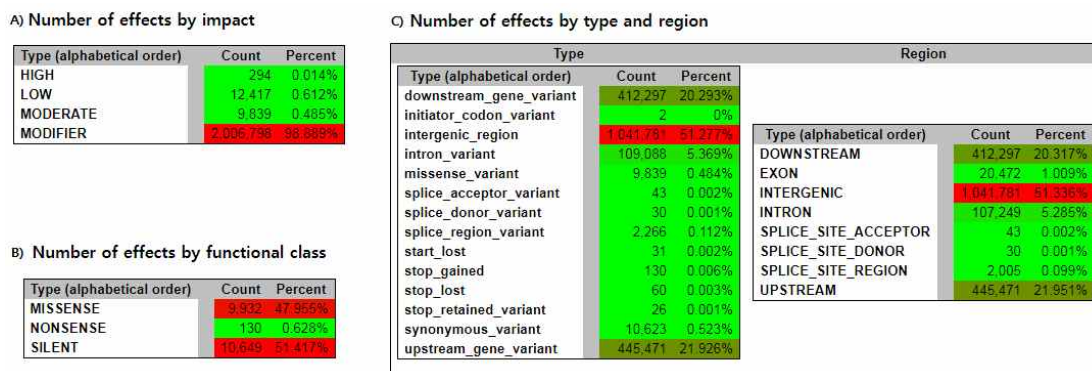


그림 3. 수박 29개 계통의 집단에 대한 snpEff 분석 결과

본 작업의 결과물은 현재 운영 중인 수박 육종특화 데이터베이스의 수박 집단의 기본 변이 정보로 update될 것이며 genome browser를 매개로 하여 데이터베이스 내의 다른 contents와 연동되어 제공될 계획이다. 또한 새로 작성된 SNP의 발생에 따른 효과를 데이터베이스 내에 반영할 수 있도록 schema를 구성하고 이를 추가하여 차년도 작업에서 반영할 수 있도록 계획 중에 있다.

2. 수박에 대한 특정 처리 및 조직 특이적 발현 유전자 정보 생산을 위한 전사체 정보의 수집

현재 주어진 샘플의 특정 조건하에서의 RNA 발현 양상의 확인은 NGS(Next-generation sequencing) 방식으로 수행되어가고 있는 추세이다. 과거에는 마이크로어레이(microarray) 방식으로 다양한 종을 대상으로 특정 조건하에서의 유전자 발현 양상의 식별을 위해 많은 실험 및 연구가 진행되었으나 분석의 다양성 및 raw 데이터의 활용의 범용성 등의 이점을 갖춘 NGS 기반의 RNA-seq을 활용한 연구가 늘어난 상황이다 (그림 3).

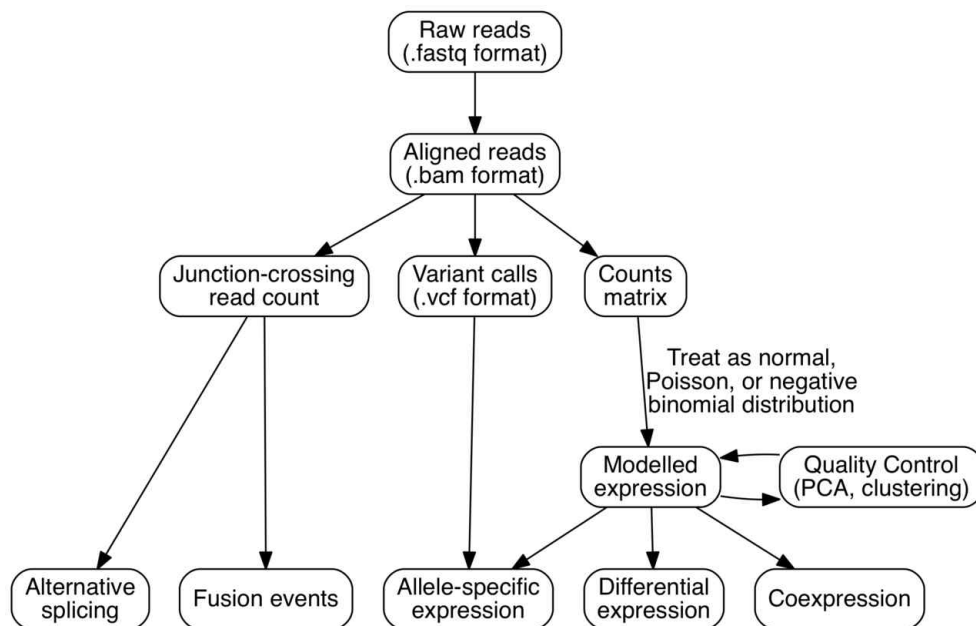


그림 4. RNA-seq의 input data의 형식과 일반적인 분석 절차와 기대 결과

본 작업에서는 수박의 저온 처리에 관련된 문헌 정보를 수집하고 해당 결과의 supplementary data로서 NCBI SRA(<https://www.ncbi.nlm.nih.gov/sra>)에 업로드 되어 있는 RNA-seq sequence data(SRP07821)를 다운로드 하였으며 해당 전사체 데이터에 대한 개략적인 정보는 표 4와 같다.

표 4. 수박 저온 관련 전사체 raw sequence의 기본 정보

| Accession | Sequence file name | Sequencing method | Phred quality score | Number of sequence reads | Read length | Property |
|------------|--------------------|-----------------------|---------------------|--------------------------|-------------|------------------------------|
| SRR3822369 | SRR3822369_1.fastq | Sanger / Illumina 1.9 | Q33 | 20163296 | 125 | Case of control 1 |
| | SRR3822369_2.fastq | Sanger / Illumina 1.9 | Q33 | 20163296 | 125 | |
| SRR3822370 | SRR3822370_1.fastq | Sanger / Illumina 1.9 | Q33 | 19021925 | 125 | Case of control 2 |
| | SRR3822370_2.fastq | Sanger / Illumina 1.9 | Q33 | 19021925 | 125 | |
| SRR3822371 | SRR3822371_1.fastq | Sanger / Illumina 1.9 | Q33 | 18852319 | 125 | Melatonin treated 1 |
| | SRR3822371_2.fastq | Sanger / Illumina 1.9 | Q33 | 18852319 | 125 | |
| SRR3822372 | SRR3822372_1.fastq | Sanger / Illumina 1.9 | Q33 | 18244757 | 125 | Melatonin treated 2 |
| | SRR3822372_2.fastq | Sanger / Illumina 1.9 | Q33 | 18244757 | 125 | |
| SRR3822373 | SRR3822373_1.fastq | Sanger / Illumina 1.9 | Q33 | 18061393 | 125 | Cold treated 1 |
| | SRR3822373_2.fastq | Sanger / Illumina 1.9 | Q33 | 18061393 | 125 | |
| SRR3822374 | SRR3822374_1.fastq | Sanger / Illumina 1.9 | Q33 | 17976978 | 125 | Cold treated 2 |
| | SRR3822374_2.fastq | Sanger / Illumina 1.9 | Q33 | 17976978 | 125 | |
| SRR3822375 | SRR3822375_1.fastq | Sanger / Illumina 1.9 | Q33 | 16422503 | 125 | Melatonin and Cold treated 1 |
| | SRR3822375_2.fastq | Sanger / Illumina 1.9 | Q33 | 16422503 | 125 | |
| SRR3822376 | SRR3822376_1.fastq | Sanger / Illumina 1.9 | Q33 | 16631311 | 125 | Melatonin and Cold treated 2 |
| | SRR3822376_2.fastq | Sanger / Illumina 1.9 | Q33 | 16631311 | 125 | |

다운로드한 SRA format 파일은 fastq-dump 명령어를 통해 paired-end 형식의 fastq 파일로 전환하였으며 fastqc 소프트웨어를 이용하여 각 fastq 파일에 대한 quality check(QC)를 수행하였다. 그 결과, 모든 시퀀스에서 no hit 그룹에 속하는 overrepresented sequence가 나타나 이를 fasta format으로 정리하여 각 샘플에 대한 contaminant sequence DB를 만들었다. 이를 각 샘플의 paired-end sequence와 함께 input으로 활용하여 Trimmomatic에 의한 sequence quality filtering 및 contaminant 서열의 제거를 실시하였다. 작업 이후, 각 샘플 상에서 contaminant 서열이 제거된 것을 확인하였다. 그러나 QT가 끝난 모든 샘플의 초기 15 bp가 GC contents에 큰 요동을 보였으며 SRR3822376_2.fastq와 SRR3822369.fastq가 각각 잔여 overrepresented sequence와 70 bp의 GC contents 요동을 보였다. 이를 개선하기 위해 seqtk로 요동을 보이는 염기서열의 영역을 각 샘플별 sequence에서 제거하였으며 잔존 overrepresented sequence를 갖는 샘플은 해당 염기서열을 반영하여 Trimmomatic을 다시 수행하였다. QC 및 filtering이 완료된 각 샘플별 fastq 파일에 대해 tophat 소프트웨어를 통해 수박의 공개된 표준 유전자 위치 정보를 활용한 reference alignment 작업을 진행하였다.

Tophat에 의한 read의 alignment가 완료되면 각 샘플별로 bam(binary alignment map) file 형식의 결과가 만들어지게 되는데 이를 DESeq에서 제공하는 파이프라인에 적용 (<https://bioconductor.org/packages/devel/bioc/vignettes/DESeq/inst/doc/DESeq.pdf>) 하여 각 샘플별로 수박 표준 유전자 상에 mapping 된 read의 수를 산출한 유전자 발현량 table을 구할 수 있었다.

3. 수박 응성불임 전사체 정보를 활용한 Indel 마커 개발

응성불임성(male sterility)은 대량 F₁ 종자 생산 산업을 가능하게 하는 작물이 가진 형질 중 하나이다. 이는 육종가가 의도하는 교배 조합의 수분 수정을 통제하기 위한 제웅 작업 절차를 생략하게 하여 육종가 및 육종회사로 하여금 제웅에 필요한 인력 및 비용을 절감할 수 있는 효과를 기대할 수 있다.

본 작업에서는 기존에 수집한 수박의 응성불임성에 차이를 보이는 두 근동질 계통(NIL line: Near isogenic line)인 DAH3615-MS(응성불임)와 DAH3615(응성가임)의 RNA-seq raw sequence로부터 자체적으로 구축한 파이프라인을 활용하여 indel 마커를 구성하고자 하였다. 이 sequence raw data를 생산한 문헌에서 두 계통의 응성불임성이 유전자적 응성불임(GMS: Genetic Male sterility)에 기인한 것이라 기술되어 있는 것을 확인하였으며, 이를 기반으로 하여 두 계통의 응성불임성에 대한 표현형적 차이를 나타내는 유전변이가 있을 것이라는 가정을 세웠다 (그림 5).

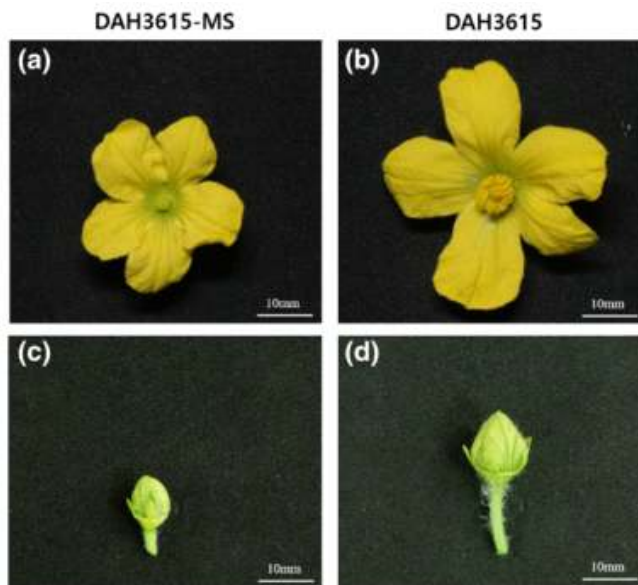


그림 5. 응성불임성에 대하여 차이를 보이는 수박의 두 근동질 계통인 DAH3615-MS(응성불임)과 DAH3615(응성가임)의 표현형적 차이 (Li et al., 2015)

가. RNA-seq raw data의 처리

해당 샘플에 대한 연구를 진행한 문헌상에서 관련 염기서열의 다운로드에 활용할 수 있는 NCBI상의 ID를 확인할 수 있었다 (SRP058526). 본 문헌상의 연구에서 DAH3615-MS(응성불임)와 DAH3615(응성가임)에 대해 화기(flower) 조직 전체와 화아(bud) 조직에 대해 mRNA에 대한 시퀀싱이 Illumina의 HiSeq 2000을 통해 수행되었다. NCBI의 FTP 서버에서 해당 ID에 대한 경로를 검색하고 관련된 총 4개의 RNA-seq에 대한 SRA 파일을 자체 보유한 분석서버에 직접 다운로드 받아 관련된 작업 디렉토리를 구성하였다.

표 5. SRP058526의 구성 RNA-seq 샘플의 정보

| Accession | Sequence file name | Sequencing method | Phred quality score | Number of sequence reads | Read length | Property |
|------------|--------------------|-----------------------|---------------------|--------------------------|-------------|---------------------|
| SRR2033940 | SRR2033940_1.fastq | Sanger / Illumina 1.9 | Q33 | 11834220 | 75-151 | Male fertile bud |
| | SRR2033940_2.fastq | Sanger / Illumina 1.9 | Q33 | 11834220 | 75-151 | |
| SRR2033941 | SRR2033941_1.fastq | Sanger / Illumina 1.9 | Q33 | 13845808 | 75-151 | Male fertile flower |
| | SRR2033941_2.fastq | Sanger / Illumina 1.9 | Q33 | 13845808 | 75-151 | |
| SRR2033942 | SRR2033942_1.fastq | Sanger / Illumina 1.9 | Q33 | 12411872 | 75-151 | Male sterile bud |
| | SRR2033942_2.fastq | Sanger / Illumina 1.9 | Q33 | 12411872 | 75-151 | |
| SRR2033943 | SRR2033943_1.fastq | Sanger / Illumina 1.9 | Q33 | 12489412 | 75-151 | Male sterile flower |
| | SRR2033943_2.fastq | Sanger / Illumina 1.9 | Q33 | 12489412 | 75-151 | |

다운로드 받은 데이터는 SRA 포맷으로 이를 sra-toolkit의 fastq-dump 명령어를 이용하여 paired-end 형식의 fastq 포맷으로 전환하였다. 전환된 파일은 fastqc를 통해 데이터의 trimming 여부 및 정도를 결정할 수 있었다. 1차 Quality check 결과 전반적으로 모든 raw sequence 데이터가 초기 15bp, 후기 10bp가 RNA-seq을 위한 라이브러리 구성에 사용된 sequencing vector 및 adapter에 의한 contaminant 서열에 의해 GC contents의 요동이 확인되었으며 후기 염기서열로 갈수록 base pair quality가 저하되는 것을 알 수 있었다. (그림 6) 따라서 Trimmomatic을 통해 각 샘플 별로 contaminant 서열 정보를 수집하여 이를 adapter 서열로 간주하여 제거와 동시에 phred score 30 기준의 quality trimming을 수행하였다. 그리고 염기서열 좌우 말단의 GC 요동을 제거하기 위해 segtk를 통해 좌우 서열을 15bp, 10bp씩 제거하였다. 이 경우 trimmomatic 필터링 과정에서 전체 서열 중에서 상당 부분이 제거된 서열이 segtk 필터링 과정에서 서열 자체가 사라지거나 30 bp 이하의 염기 서열을 가질 수 있기 때문에 다시 Trimmomatic을 통해 2차 필터링 이후 forward와 reverse 서열이 전부 생존한 pair sequence만 산출할 수 있도록 필터링 과정을 계획하였다. 본 과정은 작업의 효율성을 위해 shell script로 작성 후 진행되었다.

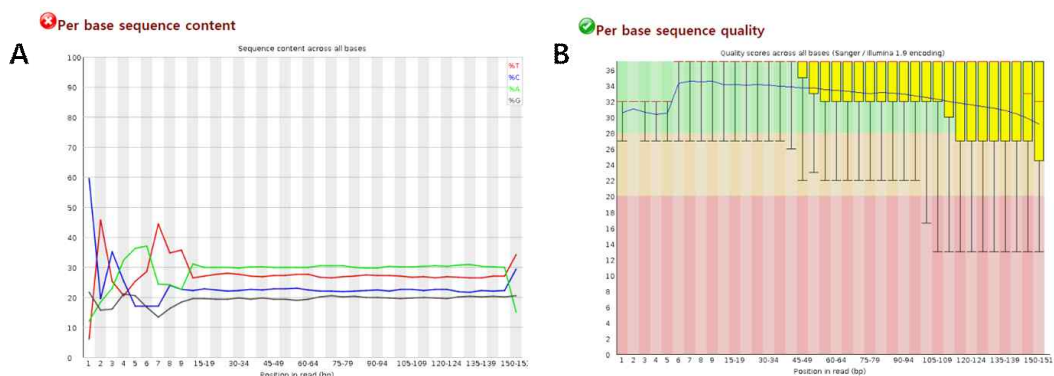


그림 6. SRR2033940_1.fastq의 1차 QC 결과. A) fastqc 결과 html 파일 내의 Per base sequence content B) Per base sequence quality 결과

shell script에 기재된 절차에 따른 adapter 서열의 제거 및 quality trimming 작업 이후 다시 fastqc를 통해 2차 QC를 진행하였으며 그림 6와 같이 양호한 결과의 quality를 갖는 fastq 파일을 확보할 수 있었다. 이후 웅성불임 및 웅성가임 별로 fastq 파일을 cat 명령어

를 통해 이를 형질 별로 하나로 합친 fastq 파일을 재생산하였다. 그리고 이를 input 데이터로 삼아 Trinity를 통한 옹성가임성과 옹성불임성 조직 별로 de novo assembly를 수행하였다. 이 과정을 통해 두 계통의 화기 관련 조직에서 발현한 mRNA의 염기서열을 재현할 수 있었으며 이는 계통별로 차이를 보이는 유전적 영역 중에서 indel만을 target으로 삼는 파이프라인을 통해 indel 마커를 생산하게 된다.

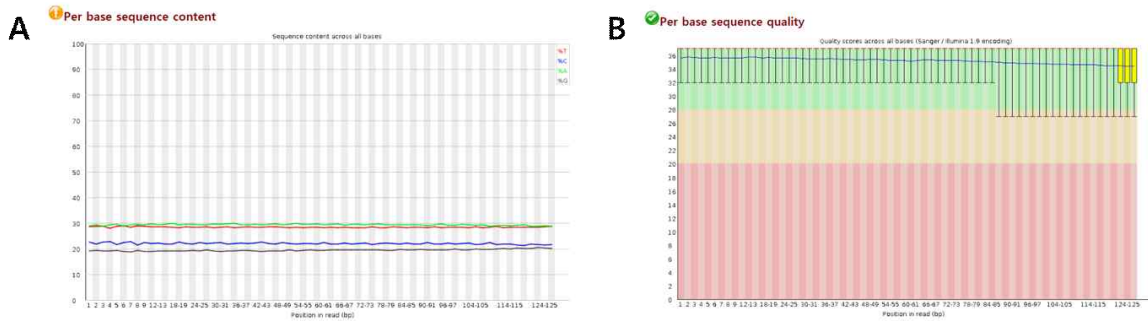


그림 7. SRR2033940_1.fastq의 2차 QC 결과. A) fastqc 결과 html 파일 내의 Per base sequence content B) Per base sequence quality 결과

나. Indel 마커 서열의 생산과 마커의 표적 유전자 관련 정보

구축한 파이프라인에 의해 생산된 indel 마커는 옹성가임 조직에서 de novo assembly를 통해 산출된 transcript 서열 ID를 기준으로 구성되었다 (표 6).

표 6. 옹성가임 조직의 transcript 기준으로 산출된 indel 마커의 기본 정보

| Marker ID | Forward primer | Reverse primer | Product size(bp) | Target gene ID |
|----------------------------------|----------------------|-------------------------|------------------|----------------|
| Male_fertility_51690_in_2_INS9 | TCGTCGTATTCGCTTCATCC | TCCTTGAATTTCTCCTCGTCC | 168 | Cla016690 |
| Male_fertility_51690_in_3_DEL3 | GGATGACAAATCCATCCCCT | GGTCTTGCTCGCTTCAACAC | 172 | Cla016690 |
| Male_fertility_57098_in_2_DEL8 | TCTCAGGACCAATTTGTGCG | TCTTGGCTTTCATCCTCGTC | 102 | Cla002818 |
| Male_fertility_69984_in_3_DEL12 | GTAAGAGCAGCACCACCTGC | ACCGTAGCATCAAATGGCAA | 176 | Cla019051 |
| Male_fertility_70430_in_4_DEL11 | GGTGCCCTTTTCTCTCTTCA | CATGCCCAAATAGCTTCCAA | 207 | Cla013521 |
| Male_fertility_72774_in_6_DEL3 | ACCAGCAGTCTAGTGGGCCT | GATCGGGAGAAGATTGAGCA | 210 | Cla013267 |
| Male_fertility_72841_in_1_DEL23 | ATCCTGGTTCCTGGTGGAAG | GGATCAGATTTAAGCAACGAAGC | 151 | Cla016120 |
| Male_fertility_74461_in_2_DEL15 | GATGGCACTGTGAAGACGTG | AAACCTCCCCTCCAAGACCT | 145 | Cla019832 |
| Male_fertility_74565_in_2_DEL13 | GATCGGGACATTTCAAAGCA | CGAACGATGGATACCCAC | 176 | Cla007102 |
| Male_fertility_74565_in_3_DEL13 | CCTTGGTCATTTGCCGTTTA | GCATTTTCCATGCTTTCACC | 129 | Cla007102 |
| Male_fertility_25230_in_2_DEL16 | ATCCAGGTTTTGTGCAGCC | TGCCCATTTCTACTGATGCT | 105 | Cla004685 |
| Male_fertility_26567_in_15_INS11 | CAATGGTGCTTCCAATGCTT | GTGCCTACAGCGACCAAAC | 199 | Cla020235 |

이를 통해 옹성불임성에 차이를 보이는 두 근동질 계통의 유전체의 염기서열에서 직접적으로 차이를 보이는 indel 영역을 targeting 하는 총 12개의 분자마커를 구성할 수 있었다. 또한 이 분자마커들이 수박의 표준 유전자가 coding되는 영역에서 직접적으로 산출된 결과이기 때문에 자연스럽게 각 분자마커가 지정하는 수박의 표준 유전자 또한 식별할 수 있었다.

표 7. 구성된 indel marker의 target 유전자의 기본 정보

| Target gene ID | Gene position | Gene description | Interpro domain |
|----------------|---------------------------|--|--|
| Cla010990 | Chr1:18011988 - 18038596 | Sister chromatid cohesion protein PDS5 homolog B | IPR016024 Armadillo-type fold |
| Cla016690 | Chr11:23855617 - 23867697 | WD-40 repeat protein-like (Fragment) | IPR017986 WD40 repeat, region |
| Cla013267 | Chr2:30834951 - 30839882 | CCR4-NOT transcription complex subunit 2 | IPR007282 NOT2/NOT3/NOT5 |
| Cla013521 | Chr2:28443462 - 28448510 | Membrane transporter D1 | IPR003663 Sugar/inositol transporter |
| Cla019832 | Chr2:26726918 - 26738658 | WD repeat-containing protein 48 | IPR017986 WD40 repeat, region |
| Cla020235 | Chr2:21834759 - 21841132 | Heavy metal-transporting ATPase | IPR006403 ATPase, P type, cation/copper-transporter |
| Cla007102 | Chr5:15635702 - 15645819 | Uridine kinase | IPR000764 Uridine kinase |
| Cla019051 | Chr6:24689385 - 24694714 | Nuclear transport factor 2 and RNA recognition motif domain-containing protein | IPR018222 Nuclear transport factor 2, Eukaryote |
| Cla002818 | Chr9:14959564 - 14962079 | Monoxygenase FAD-binding | IPR003953 Fumarate reductase/succinate dehydrogenase flavoprotein, N-terminal |
| Cla004685 | Chr9:32090650 - 32097483 | Methionyl-tRNA synthetase | IPR002304 Methionyl-tRNA synthetase, class Ia |
| Cla016120 | Chr9:13854736 - 13917880 | Vacuolar sorting protein | IPR001849 Pleckstrin homology |

위의 표에서 확인할 수 있는 수박 표준 유전자 중에서 Cla013521은 sugar/inositol transporter라는 단백질을 coding 하는 유전자이다. 이 유전자는 화분(pollen)의 발달에 필요한 당의 공급에 관계되어 있으며 이 단백질의 구조 변화는 화분의 임성에 영향을 미친다는 결과가 알려져 있다. 구축된 파이프라인을 통해 개발된 indel 마커는 서로 다른 두 계통이 exon상에서 차이를 보이는 영역만을 대상으로 구성되기 때문에 직관적으로 개발된 마커가 두 계통간의 표현형 차이를 유도할 수 있는 영역을 표적으로 하는 것을 기대할 수 있다. 이는 마커 개발 과정에서 indel이 있는 표준 유전자의 annotation을 확인하여 어떠한 유형의 단백질을 coding하는 지를 확인하는 것을 용이하게 한다. 문헌 조사 등을 통해 해당 유전자가 연구 대상인 표현형에 밀접하게 관련되어 있는 것으로 알려진 것이라면 본 과정을 통한 indel marker의 개발은 표현형 구분 마커 개발의 효율적인 사례라고 할 수 있다.

제2절 고추, 파프리카의 육종 특화 구축을 위한 기반 정보 수집 및 기존 정보 업데이트

1. 고추(Pepper, *Capsicum annuum*) 관련 육종 관련 정보의 산출이 가능한 데이터의 출처 탐색 및 데이터의 보유 현황 확인

NCBI SRA는 전세계의 연구진으로부터 제출된 유전현상에 대한 raw 데이터로서의 시퀀싱 기반 데이터들을 하나로 통합하고 접근할 수 있게 하는 플랫폼이다. 이를 통해 다양한 목적으로 특정 계통 및 처리로부터 생산된 시퀀싱 데이터들이 그것을 필요로 하는 연구 공동체에 개방되어 활용됨으로써 이차적인 연구 및 분석을 기대할 수 있게 하였다. NCBI SRA는 Roche 454 GS System®, Illumina Genome Analyzer®, Applied Biosystems SOLiD System®, Helicos Heliscope®, Complete Genomics®, and Pacific Biosciences SMRT®와 같은 다양한 NGS(Next generation sequencing) 기반 sequencer 기종의 sequencing 결과를 취급하며 이는 NCBI SRA에서의 검색 활동을 통해 사용자가 원하는 정보를 선별적으로 선택하여 활용이 가능하다. 그러나 NGS 기반의 데이터를 다루는 경험이 부족한 연구자 및 육종 관련 기업체의 경우 이와 같은 활동을 직접 수행하는데 어려움을 겪을 수밖에 없으며, 데이터를 획득하더라도 데이터 분석 과정에서 수반되는 중간 생성파일의 용량 및 컴퓨팅 리소스를 감당하기 어려워 의미 있는 분석 결과를 얻기가 어렵다. 따라서 본 작업에서는 현재 NCBI에서 활용이 가능한 특정 작물의 sequencing 데이터에 접근하여 효율적으로 이를 획득하고 육종에 활용할 수 있는 데이터를 구성할 수 있는 절차의 구축 및 이를 채소종자육종데이터베이스에 입력할 수 있는 형태로 가공하여 일반 연구자 및 육종가가 이를 활용할 수 있도록 하고자 한다.

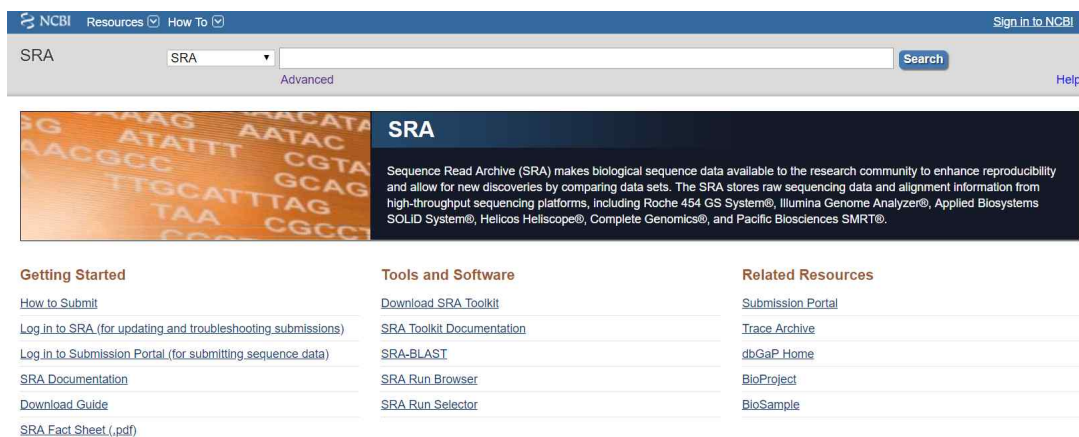


그림 1. NCBI SRA 메인 페이지(<https://www.ncbi.nlm.nih.gov/sra>)

가. NCBI SRA(Sequence Read Archive)의 정보의 접근 및 활용 방안의 구체화

그림 1과 같은 NCBI SRA의 메인 페이지에서 검색을 통해 특정 keyword를 입력하면 해당 keyword를 타이틀에 포함하는 모든 검색 결과가 출력되게 된다. 실제 Pepper를 대상으로 검색을 진행하여 출력되는 결과를 확인하였을 때, 고추에서 유래한 DNA 혹은 RNA 관련 파일이 아닌 결과가 일부 확인된다. 또한 한 페이지에서 확인 가능한 결과의 개수가 최대 200개이기 때문에 한 번에 어떤 파일이 업로드 되어 있는지를 확인할 수 없다. 이와 같은 경우 화면 우측의 “Send to”를 클릭하고 첫 번째 체크 박스에서 File을 선택한다. 이

경우 자동으로 셀렉트 박스가 체크 박스 하단에 구성이 되며 여기서 Runinfo를 선택하도록 하면 해당 검색어에 대하여 csv 포맷으로 정리된 결과를 다운로드 받을 수 있다(그림 2).



그림 2. keyword를 Pepper로 설정하여 검색한 결과의 csv 포맷 형식 출력

해당 결과는 엑셀 소프트웨어로 열람이 가능하며 첫 번째 행의 여러 컬럼을 기준으로 정렬 및 필터링을 통해 연구자가 원하는 조건에 맞는 결과만을 추출할 수 있다. 결과를 정리 하였을 때 2018년 10월 기준으로 총 37개의 서로 다른 Bioproject에서 620여개의 sequencing 데이터(Total 2.39 Tb)를 확인할 수 있었다. 또한 해당 sequencing 데이터들은 크게 miRNA, transcriptome 그리고 genome 상의 변이 식별을 위해 생산되었으며 주로 Illumina의 Hiseq 시리즈가 데이터구성에 사용된 것을 알 수 있었다 (그림 3).

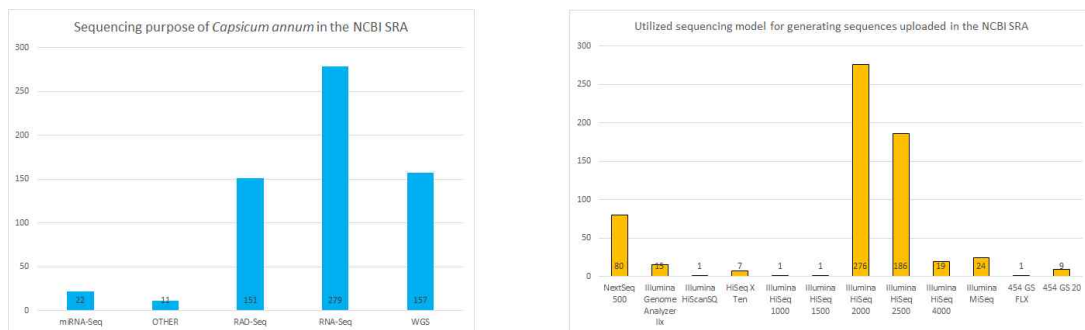


그림 3. NCBI SRA에 업로드되어 있는 Pepper(*Capsicum annuum*) sequence raw 데이터의 목적 및 sequence 데이터를 생산에 사용된 sequencer 기종의 분포

다운로드한 csv 포맷 형식의 파일에는 각 sequencing 결과에 ID를 부여하고 NCBI 플랫폼 내에서 다운로드가 가능한 경로를 지정하고 있다. 이 정보를 활용하여 리눅스 서버 내에서 각 경로로부터 sequencing 파일을 SRA 포맷의 형태로 다운로드 받고 이를 fastq 파일로 전환하여 후속 분석 및 데이터 저장에 이용할 수 있다. 그리고 csv 파일의 BioProject 열의 ID를 NCBI의 같은 이름의 플랫폼에서 (<https://www.ncbi.nlm.nih.gov/bioproject>) 검색한 결과에서 해당 BioProject를 포함하는 연구를 통해 출판된 논문을 확인하면, 검색한 SRA 파일을 포함하는 sequencing project가 어떠한 목적으로 수행이 되었는가를 확인할 수 있어 이를 연구자들이 연구 목적에 부합하는 sequencing 데이터를 선별 및 활용할 수 있다.

표 1. SRA 검색 파일의 BioProject 열을 기준으로 재검색된 논문의 예시

| No. | Title | PMID | Bioproject in SRA |
|-----|--|----------|----------------------------|
| 1 | De novo assembly of the pepper transcriptome (<i>Capsicum annuum</i>): a benchmark for <i>in silico</i> discovery of SNPs, SSRs and candidate genes. | 23110314 | PRJNA163071 |
| 2 | Development of molecular markers tightly linked to Pvr4 gene in pepper using next-generation sequencing. | 25798050 | PRJNA262545 |
| 3 | Dynamics of the chili pepper transcriptome during fruit development | 24555715 | PRJNA235215 |
| 4 | Genome sequence of the hot pepper provides insights into the evolution of pungency in Capsicum species | 24441736 | PRJNA223222 PRJNA298503 |
| 5 | Genome-wide comparative analysis in Solanaceous species reveals evolution of microRNAs targeting defense genes in <i>Capsicum</i> spp | - | PRJNA398663 |
| 6 | Genomewide genotyping of a novel Mexican Chile Pepper collection illuminates the history of landrace differentiation after <i>Capsicum annuum</i> L. domestication | - | PRJNA472885 |
| 7 | New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication | 29089032 | PRJNA223222 |
| 8 | The evolution of inflorescence diversity in the nightshades and heterochrony during meristem maturation | 27821409 | PRJNA343677 |
| 9 | The hot pepper (<i>Capsicum annuum</i>) microRNA transcriptome reveals novel and conserved targets: a foundation for understanding MicroRNA functional roles in hot pepper.. | 23737975 | PRJNA177852 |
| 10 | Transcriptome analysis of symptomatic and recovered leaves of geminivirus-infected pepper (<i>Capsicum annuum</i>). | 23185982 | PRJNA186495 |
| 11 | Transcriptome profiling and molecular marker discovery in red pepper, <i>Capsicum annuum</i> L. TF68 | 21706160 | PRJNA140069 |
| 12 | Transcriptome Sequencing and De Novo Analysis of a Cytoplasmic Male Sterile Line and Its Near-Isogenic Restorer Line in Chili Pepper (<i>Capsicum annuum</i> L.) | - | PRJNA401344 PRJNA186921 |
| 13 | Whole-genome sequencing of cultivated and wild peppers provides insights into Capsicum domestication and specialization | 24591624 | PRJNA193077 PRJNA193078 |

나. NCBI SRA(Sequence Read Archive)의 FastQ format 전환 및 NGS 데이터 처리과정에서의 유의 사항을 반영한 파이프라인(Pipeline)의 구성

일반적으로 Whole genome sequencing(WGS)를 통한 유전변이의 식별 분석에서는 NGS 기반의 SNP calling을 통해 수행된다. 전반적인 작업은 sequencing 결과인 fastq 파일의 quality check 및 trimming, Read alignment into reference genome, GATK pipeline을 통한 bam 파일의 편집 및 variant calling으로 구성되어진다. 현재 전세계 적으로 sequencing 분석을 주도하고 있는 것은 Illumina 사의 Hiseq platform으로 DNA sample로부터 sequencing library를 구성하고 flowcell 상에서 PCR duplication를 통해 증폭하여 형광의 세기를 sequence의 센서가 검출이 가능한 수준으로 높게 된다. 이 과정에서 동일한 read sequence가 variant calling의 단계에서 false positive variant를 만들 수 있는 유의미한 수준으로 본래의 개수에 비해 크게 증가되어진다. 이러한 optical duplicate를 variant calling process에서 검출하고 제거하는 것을 mark duplicates이라 하는데 이 단계는 주로 sequence를 reference genome에 alignment한 결과인 sam file을 bam format으로 전환 후 sorting 까지 진행한 이후에 수행하게 된다. 이러한 mark duplicates는 본래 fastq format file내의 read가 가진 ID가 갖는 정보를 바탕으로 식별이 가능하다. 그러나 NCBI에서 다운로드 받을 수 있는 sequencing 파일은 이러한 read ID에 대한 정보가 누락되어 통상적인 방법으로 GATK 파이프라인 내에서의 raw sequence file에 포함된 불필요한 duplicates의 제거가 어렵다. 따라서 통상적인 optical duplicates의 제거방식이 아닌 fastq 파일 내에서 sequence가 정확하게 일치하는 read sequence를 PCR duplicates로 간주하여 제거하는 방식을 채택하여 이를 통해 리눅스 작업환경에서 GATK 파이프라인에서의 MarkDuplicates 단계를 대체하는 절차를 구성하였다. 구성된 작업은 다운로드 대상인 파일의 NCBI FTP상의 저장 경로를 반영한 자동 download script와 다운로드 결과인 SRA format file을 SRAtoolkit의 vdb-validate를 통하여 다운로드 과정에서 문제점이 없었는지를 확인 이후, 같은 package의 fastq-dump를 활용하여 fastq format file를 준비한다. 이후 fastqc software를 통해 추가적인 필터링이 필요한지에 대한 여부를 확인하고 문제가 없을 경우 fastq format file내의 PCR duplicates를 제거하는 작업을 bmap package의 Dedupe.sh 프로그램을 통해 수행한다. 해당 작업은 paired end sequence를 대상으로 진행하였을 때

input file을 하나로 합치므로 같은 package내의 reformat.sh 프로그램을 이용하여 다시 paired end 형태로 파일을 분리하는 작업을 실시한다. 이와 같은 과정으로 optical duplicates 제거 기반의 MarkDuplicates 단계를 대체하여 Variant calling 작업의 후속작업을 진행할 수 있다.

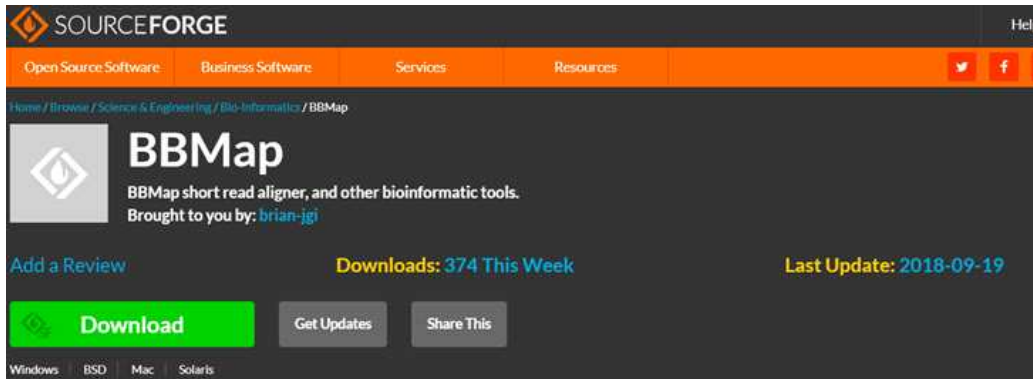


그림 4. 생물정보분석 관련 software 및 script를 지원하는 SOURCEFORGE의 bbmap package 관련 webpage(<https://sourceforge.net/projects/bbmap/>)

1. Download SRA file and converting fastq format
2. Remove PCR duplicates by dedup.sh in BBmap
3. FASTQC and remove adapter sequence
4. Read alignment by BWA - mem algorithm
5. Follow GATK bam motification process except for MardDuplicate and base recalibration
6. SNP calling by Samtools/Bcftools combined command

그림 5. NCBI SRA에서 획득한 fastq 파일 기반 variant calling을 위해 구축한 파이프라인

2. 채소종자데이터베이스 플랫폼의 고추 항목을 위한 표준 유전체 정보의 확보 및 데이터화

고추는 남아메리카 원산의 가지과 작물로서 1493년 스페인에 도입된 이후, 15세기에 영국과 중부 유럽, 그리고 17세기에 중국으로 전파된 이력을 가진 전 세계적인 시장을 가지고 있는 작물이다. 국내의 경우 이용면에서 건과용 고추와 풋고추용을 구분하여 재배하고 있다. 또한 고추는 국내에서 소비되는 김치의 주재료로서 거대한 시장을 형성하고 있으며 2014년에는 서울대학교의 연구팀에 의해 고추의 전장 유전체가 해독되어 공개된 상태이다. 다양한 경로를 통해 확보한 고추의 계통별 raw 데이터로부터 변이정보 유용 유전자의 발현량 및 분자마커의 개발등을 위해서는 진술한 고추의 표준 전장유전체 및 예측된 유전자의 정보가 필요하다. 예측된 유전자의 경우 GO term, KEGG pathway 등과 같은 기준에 구축된 유전자의 일반적 특성을 다루는 데이터베이스의 정보와 연동이 되어야 추후 고추의 변이 및 발현량 데이터에 부가적인 정보를 부여할 수 있다.

가. 고추 표준 유전체 정보의 확인

2018년 기준으로 고추에 대한 전장 유전체 sequencing 주요한 성과는 국내의 서울대학교와 중국의 쓰촨 농업대학의 연구에서 확인할 수 있었다(Kim et al., 2014, Cheng et al., 2014). 이 중 서울대학교의 연구는 contig로부터 구성한 scaffold들을 통해 12개의 고추의 크로모솜을 구성하고 크로모솜 상의 유전자까지 예측하였다. 해당 결과는 웹데이터베이스 (<http://peppergenome.snu.ac.kr>)에 공개되어 전세계의 연구자들에 의해 고추의 표준유전체로서 활용되고 있다(그림 6).

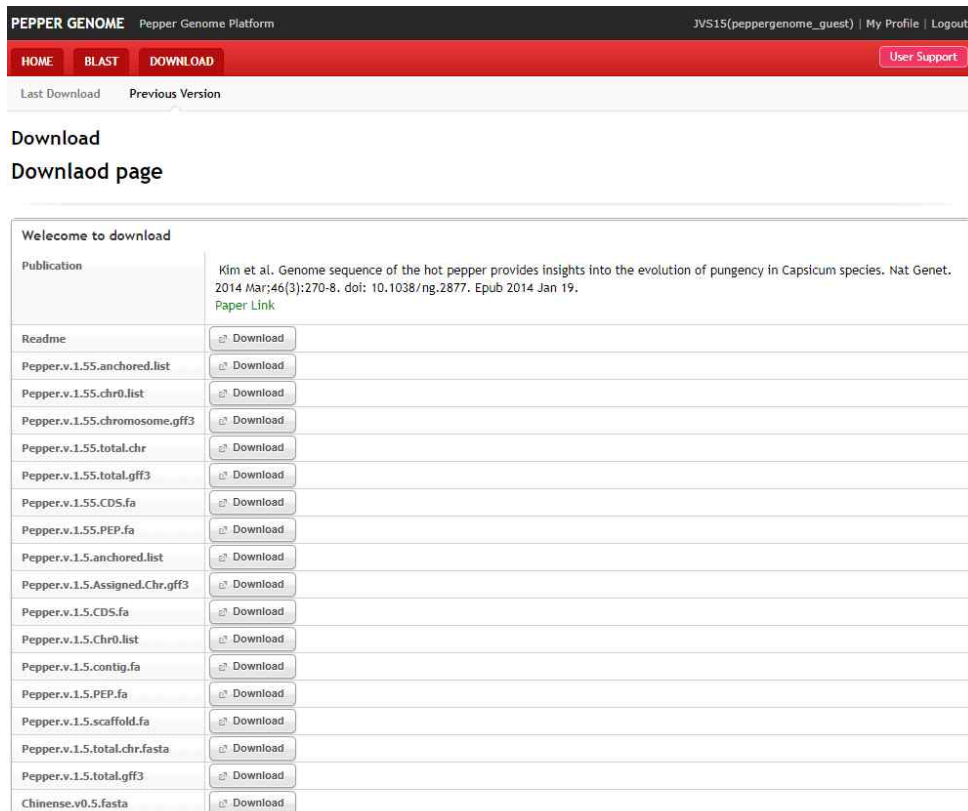


그림 6. Pepper genome database의 다운로드 페이지

해당 데이터베이스에 접속한 후, 수집 및 공여받은 고추계통의 fastq 파일을 대상으로 variant calling을 진행하기 위하여 reference genome의 버전 1.55에 해당하는 고추의 12개 크로모솜에 대한 시퀀스 정보를 가진 fasta 파일 및 예측된 유전자의 크로모솜 상의 위치를 기록한 gff 파일을 다운로드하여 채소종자데이터베이스가 탑재된 메인서버에 저장하였다. 전체 유전체의 크기는 2.51 Gb에 달하고 총 30,242개의 유전자가 유전체상에 그 위치가 예측이 되어있다.

나. 고추 표준 유전체 상의 예측된 유전자의 annotation

주어진 집단에 대한 sequence alignment 이후 variant calling을 통해 indel 혹은 SNP를 산출할 수 있다. 이와 같이 산출된 변이가 유전자 상에 있는가를 식별하기 위해서는 reference genome 상에 예측된 유전자의 위치정보가 필요하다. 특히 GWAS의 경우 SNP에 연관 유전자의 정보를 연결하게 되는데 이때 추가적으로 수행해야 하는 사항이 해당 유전자는 어떠한 생물학적 정보를 보유하고 있는가이다. 따라서 고추에서 예측된 유전자의 특성을 분류하기 위하여 고추 표준 유전자의 아미노산 서열을 query로 하여 NCBI non-redundant database에 BLASTP를 수행하였다. 고추 표준 유전체 ver 1.55에는 현재 총 34,899개의 유전자가 예측되어 있다. BLASTP의 결과 이 중 28,270개의 유전자에 대해 NCBI의 RefSeq ID가 부여되었다. 해당 RefSeqID를 sorting 하고 중복되지 않는 리스트를 얻으면 총 14,286개의 RefSeq ID가 고추 표준 유전자에 대하여 부여된 것을 확인할 수 있었다. 이를 UNIPROT database(www.uniprot.org)의 Retrieve/ID mapping 플랫폼에 입력하여 해당 유전자에 대해 표 2에 기재된 부여할 수 있는 모든 정보를 customize하여 출력하여 고추의 유전육종 데이터에 활용할 수 있도록 데이터베이스화 하는 작업을 수행하였다.

표 2. NCBI RefSeqID를 통해 매개된 현행 웹데이터베이스의 정보 일람

| Property for gene annotation | Source | Web address |
|------------------------------|---------|-----------------------------|
| RefSeq ID | NCBI | www.ncbi.nlm.nih.gov/refseq |
| PubMed ID | NCBI | www.ncbi.nlm.nih.gov/pubmed |
| EC number | ExpASY | enzyme.expasy.org |
| Geneontology | GO | www.geneontology.org |
| Protein interaction | STRING | string-db.org |
| Pathway | KEGG | www.genome.jp/kegg |
| Domain information | Pfam | pfam.xfam.org |
| Protein classification | PANTHER | www.pantherdb.org |
| Conserved domain | NCBI | www.ncbi.nlm.nih.gov/cdd |

3. 고추 RIL(recombinant inbred line) 집단의 genotype 데이터의 구성

서울대학교 원예학과 원예작물유전육종학 연구실(<http://hortbreeders.snu.ac.kr/>)로부터 총 122개의 고추 계통에서 얻은 whole genome resequencing 결과를 fastq 파일 형식으로 전달 받았다.

가. 집단의 sequencing 데이터 구성 정보

집단을 이루는 각 계통의 Sequencing 데이터는 Illumina의 Hiseq2000 기종으로 생산되었으며 Paired End 형식으로 약 500bp 내외의 insert size가 적용되어 구성되었다. 하나의 집단을 이루는 122개 계통 중 2개의 계통(Perennial, Dempsey)는 각각 집단의 부모본에 해당 하며 그 외 120개의 계통은 이 교배 조합에서 유래한 RIL(Recombinant inbred line)이다. 부모본의 경우 고추 표준유전체 1.55 ver (2.51 Gb)에 대하여 15 X 수준으로 sequencing이 되었고 RIL 계통들의 경우 평균 약 1.22 X 수준으로 low-depth whole genome sequencing 이 이루어졌다 (표 3).

표 3. 서울대학교 공여 고추(Pepper) RIL 집단의 sequence 정보

| Number | Library Strategy | Bases | Read coverage(X) | Library Name | Info |
|--------|------------------|---------|------------------|--------------|-------------------------|
| 1 | WGS | 39.18Gb | 15.61 | Dempsey | Parental line |
| 2 | WGS | 42.01Gb | 16.74 | Perennial | Parental line |
| 3 | WGS | 2.61Gb | 1.04 | PDRIL_1 | Recombinant Inbred line |
| 4 | WGS | 3.65Gb | 1.46 | PDRIL_2 | Recombinant Inbred line |
| 5 | WGS | 2.88Gb | 1.15 | PDRIL_3 | Recombinant Inbred line |
| 6 | WGS | 2.91Gb | 1.16 | PDRIL_4 | Recombinant Inbred line |
| 7 | WGS | 2.27Gb | 0.9 | PDRIL_6 | Recombinant Inbred line |
| 8 | WGS | 3.32Gb | 1.32 | PDRIL_7 | Recombinant Inbred line |
| 9 | WGS | 3.41Gb | 1.36 | PDRIL_8 | Recombinant Inbred line |
| 10 | WGS | 3.25Gb | 1.3 | PDRIL_10 | Recombinant Inbred line |
| 11 | WGS | 6.36Gb | 2.53 | PDRIL_11 | Recombinant Inbred line |
| 12 | WGS | 2.16Gb | 0.86 | PDRIL_13 | Recombinant Inbred line |
| 13 | WGS | 2.78Gb | 1.11 | PDRIL_15 | Recombinant Inbred line |
| 14 | WGS | 2.99Gb | 1.19 | PDRIL_16 | Recombinant Inbred line |
| 15 | WGS | 2.59Gb | 1.03 | PDRIL_17 | Recombinant Inbred line |
| 16 | WGS | 3.02Gb | 1.2 | PDRIL_19 | Recombinant Inbred line |
| 17 | WGS | 2.32Gb | 0.92 | PDRIL_21 | Recombinant Inbred line |
| 18 | WGS | 2.56Gb | 1.02 | PDRIL_23 | Recombinant Inbred line |
| 19 | WGS | 3.53Gb | 1.41 | PDRIL_25 | Recombinant Inbred line |
| 20 | WGS | 2.82Gb | 1.13 | PDRIL_26 | Recombinant Inbred line |
| 21 | WGS | 2.72Gb | 1.09 | PDRIL_27 | Recombinant Inbred line |
| 22 | WGS | 2.56Gb | 1.02 | PDRIL_28 | Recombinant Inbred line |
| 23 | WGS | 1.67Gb | 0.67 | PDRIL_30 | Recombinant Inbred line |
| 24 | WGS | 2.15Gb | 0.85 | PDRIL_34 | Recombinant Inbred line |
| 25 | WGS | 1.24Gb | 0.49 | PDRIL_35 | Recombinant Inbred line |
| 26 | WGS | 3.32Gb | 1.32 | PDRIL_36 | Recombinant Inbred line |
| 27 | WGS | 3.74Gb | 1.49 | PDRIL_37 | Recombinant Inbred line |
| 28 | WGS | 1.85Gb | 0.74 | PDRIL_38 | Recombinant Inbred line |
| 29 | WGS | 3.35Gb | 1.33 | PDRIL_39 | Recombinant Inbred line |
| 30 | WGS | 3.29Gb | 1.31 | PDRIL_40 | Recombinant Inbred line |
| 31 | WGS | 3.12Gb | 1.24 | PDRIL_42 | Recombinant Inbred line |
| 32 | WGS | 3.9Gb | 1.56 | PDRIL_43 | Recombinant Inbred line |
| 33 | WGS | 1.96Gb | 0.78 | PDRIL_44 | Recombinant Inbred line |
| 34 | WGS | 1.79Gb | 0.71 | PDRIL_46 | Recombinant Inbred line |
| 35 | WGS | 3.37Gb | 1.34 | PDRIL_47 | Recombinant Inbred line |
| 36 | WGS | 3.71Gb | 1.48 | PDRIL_48 | Recombinant Inbred line |

| | | | | | |
|----|-----|--------|------|-----------|-------------------------|
| 37 | WGS | 3.48Gb | 1.39 | PDRIL_49 | Recombinant Inbred line |
| 38 | WGS | 1.82Gb | 0.73 | PDRIL_50 | Recombinant Inbred line |
| 39 | WGS | 7.89Gb | 3.14 | PDRIL_52 | Recombinant Inbred line |
| 40 | WGS | 3.03Gb | 1.21 | PDRIL_53 | Recombinant Inbred line |
| 41 | WGS | 1.31Gb | 0.52 | PDRIL_56 | Recombinant Inbred line |
| 42 | WGS | 3.13Gb | 1.25 | PDRIL_58 | Recombinant Inbred line |
| 43 | WGS | 3.21Gb | 1.28 | PDRIL_59 | Recombinant Inbred line |
| 44 | WGS | 3.47Gb | 1.38 | PDRIL_62 | Recombinant Inbred line |
| 45 | WGS | 3.66Gb | 1.46 | PDRIL_63 | Recombinant Inbred line |
| 46 | WGS | 3.17Gb | 1.26 | PDRIL_64 | Recombinant Inbred line |
| 47 | WGS | 5.54Gb | 2.21 | PDRIL_66 | Recombinant Inbred line |
| 48 | WGS | 2.52Gb | 1 | PDRIL_67 | Recombinant Inbred line |
| 49 | WGS | 2.77Gb | 1.1 | PDRIL_70 | Recombinant Inbred line |
| 50 | WGS | 2.74Gb | 1.09 | PDRIL_71 | Recombinant Inbred line |
| 51 | WGS | 3.52Gb | 1.4 | PDRIL_74 | Recombinant Inbred line |
| 52 | WGS | 4.07Gb | 1.62 | PDRIL_76 | Recombinant Inbred line |
| 53 | WGS | 2.96Gb | 1.18 | PDRIL_77 | Recombinant Inbred line |
| 54 | WGS | 3.74Gb | 1.49 | PDRIL_78 | Recombinant Inbred line |
| 55 | WGS | 1.87Gb | 0.75 | PDRIL_80 | Recombinant Inbred line |
| 56 | WGS | 2.12Gb | 0.84 | PDRIL_82 | Recombinant Inbred line |
| 57 | WGS | 1.68Gb | 0.67 | PDRIL_83 | Recombinant Inbred line |
| 58 | WGS | 1.74Gb | 0.69 | PDRIL_84 | Recombinant Inbred line |
| 59 | WGS | 2.38Gb | 0.95 | PDRIL_86 | Recombinant Inbred line |
| 60 | WGS | 1.68Gb | 0.67 | PDRIL_87 | Recombinant Inbred line |
| 61 | WGS | 2.39Gb | 0.95 | PDRIL_88 | Recombinant Inbred line |
| 62 | WGS | 3.86Gb | 1.54 | PDRIL_91 | Recombinant Inbred line |
| 63 | WGS | 6.9Gb | 2.75 | PDRIL_92 | Recombinant Inbred line |
| 64 | WGS | 1.64Gb | 0.66 | PDRIL_93 | Recombinant Inbred line |
| 65 | WGS | 1.94Gb | 0.77 | PDRIL_95 | Recombinant Inbred line |
| 66 | WGS | 2Gb | 0.8 | PDRIL_99 | Recombinant Inbred line |
| 67 | WGS | 5.35Gb | 2.13 | PDRIL_100 | Recombinant Inbred line |
| 68 | WGS | 2.85Gb | 1.14 | PDRIL_101 | Recombinant Inbred line |
| 69 | WGS | 3.2Gb | 1.27 | PDRIL_103 | Recombinant Inbred line |
| 70 | WGS | 3.09Gb | 1.23 | PDRIL_104 | Recombinant Inbred line |
| 71 | WGS | 3.08Gb | 1.23 | PDRIL_105 | Recombinant Inbred line |
| 72 | WGS | 3.11Gb | 1.24 | PDRIL_107 | Recombinant Inbred line |
| 73 | WGS | 3.1Gb | 1.23 | PDRIL_109 | Recombinant Inbred line |
| 74 | WGS | 3.66Gb | 1.46 | PDRIL_112 | Recombinant Inbred line |
| 75 | WGS | 3.93Gb | 1.57 | PDRIL_113 | Recombinant Inbred line |
| 76 | WGS | 3.4Gb | 1.35 | PDRIL_114 | Recombinant Inbred line |
| 77 | WGS | 3.01Gb | 1.2 | PDRIL_116 | Recombinant Inbred line |
| 78 | WGS | 2.45Gb | 0.98 | PDRIL_120 | Recombinant Inbred line |
| 79 | WGS | 3.24Gb | 1.29 | PDRIL_122 | Recombinant Inbred line |
| 80 | WGS | 2.94Gb | 1.17 | PDRIL_123 | Recombinant Inbred line |
| 81 | WGS | 3.25Gb | 1.29 | PDRIL_125 | Recombinant Inbred line |
| 82 | WGS | 3.38Gb | 1.35 | PDRIL_126 | Recombinant Inbred line |
| 83 | WGS | 3.07Gb | 1.22 | PDRIL_128 | Recombinant Inbred line |
| 84 | WGS | 2.79Gb | 1.11 | PDRIL_130 | Recombinant Inbred line |
| 85 | WGS | 3.86Gb | 1.54 | PDRIL_133 | Recombinant Inbred line |
| 86 | WGS | 2.85Gb | 1.14 | PDRIL_134 | Recombinant Inbred line |
| 87 | WGS | 3.82Gb | 1.52 | PDRIL_135 | Recombinant Inbred line |
| 88 | WGS | 2.92Gb | 1.16 | PDRIL_136 | Recombinant Inbred line |
| 89 | WGS | 2.56Gb | 1.02 | PDRIL_137 | Recombinant Inbred line |
| 90 | WGS | 2.73Gb | 1.09 | PDRIL_139 | Recombinant Inbred line |
| 91 | WGS | 2.42Gb | 0.96 | PDRIL_141 | Recombinant Inbred line |
| 92 | WGS | 2.49Gb | 0.99 | PDRIL_143 | Recombinant Inbred line |
| 93 | WGS | 3.93Gb | 1.56 | PDRIL_144 | Recombinant Inbred line |
| 94 | WGS | 2.41Gb | 0.96 | PDRIL_148 | Recombinant Inbred line |
| 95 | WGS | 2.28Gb | 0.91 | PDRIL_149 | Recombinant Inbred line |

| | | | | | |
|-----|-----|--------|------|-----------|-------------------------|
| 96 | WGS | 3.22Gb | 1.28 | PDRIL_151 | Recombinant Inbred line |
| 97 | WGS | 4.56Gb | 1.82 | PDRIL_155 | Recombinant Inbred line |
| 98 | WGS | 2.69Gb | 1.07 | PDRIL_156 | Recombinant Inbred line |
| 99 | WGS | 2.98Gb | 1.19 | PDRIL_157 | Recombinant Inbred line |
| 100 | WGS | 3.18Gb | 1.27 | PDRIL_158 | Recombinant Inbred line |
| 101 | WGS | 2.76Gb | 1.1 | PDRIL_159 | Recombinant Inbred line |
| 102 | WGS | 2.95Gb | 1.18 | PDRIL_160 | Recombinant Inbred line |
| 103 | WGS | 3.86Gb | 1.54 | PDRIL_162 | Recombinant Inbred line |
| 104 | WGS | 2.72Gb | 1.08 | PDRIL_163 | Recombinant Inbred line |
| 105 | WGS | 3.87Gb | 1.54 | PDRIL_171 | Recombinant Inbred line |
| 106 | WGS | 4.26Gb | 1.7 | PDRIL_176 | Recombinant Inbred line |
| 107 | WGS | 3.24Gb | 1.29 | PDRIL_181 | Recombinant Inbred line |
| 108 | WGS | 3.1Gb | 1.23 | PDRIL_185 | Recombinant Inbred line |
| 109 | WGS | 2.22Gb | 0.88 | PDRIL_186 | Recombinant Inbred line |
| 110 | WGS | 2.76Gb | 1.1 | PDRIL_187 | Recombinant Inbred line |
| 111 | WGS | 2.85Gb | 1.14 | PDRIL_188 | Recombinant Inbred line |
| 112 | WGS | 3.07Gb | 1.22 | PDRIL_191 | Recombinant Inbred line |
| 113 | WGS | 3.79Gb | 1.51 | PDRIL_192 | Recombinant Inbred line |
| 114 | WGS | 2.57Gb | 1.02 | PDRIL_194 | Recombinant Inbred line |
| 115 | WGS | 3.16Gb | 1.26 | PDRIL_196 | Recombinant Inbred line |
| 116 | WGS | 2.75Gb | 1.1 | PDRIL_200 | Recombinant Inbred line |
| 117 | WGS | 3.12Gb | 1.24 | PDRIL_201 | Recombinant Inbred line |
| 118 | WGS | 4.42Gb | 1.76 | PDRIL_203 | Recombinant Inbred line |
| 119 | WGS | 2.67Gb | 1.06 | PDRIL_204 | Recombinant Inbred line |
| 120 | WGS | 3.01Gb | 1.2 | PDRIL_205 | Recombinant Inbred line |
| 121 | WGS | 2.8Gb | 1.11 | PDRIL_206 | Recombinant Inbred line |
| 122 | WGS | 2.81Gb | 1.12 | PDRIL_208 | Recombinant Inbred line |

나. Sequencing 데이터의 전처리 및 Variant calling

전달받은 fastq 파일은 fastqc를 통해 sequence 상에 잔존한 adapter의 유무, phred score가 30이하로 추가적인 필터링이 필요 여부를 확인하였으나 모든 파일이 이미 관련 필터링이 진행되어 있어 별도의 filtering 절차가 없이 후속 작업인 read alignment를 고추 reference genome 1.55ver에 BWA mem 알고리즘(ver 0.7.12)을 활용하여 진행하였다. 계통별로 Read alignment로부터 생산된 sam(sequence alignment map) 파일을 samtools 소프트웨어(ver 1.3)를 통해 bam으로 전환한 후 alignment가 수행된 지놈상의 위치를 기준으로 sorting하였다. 각 fastq 파일은 이미 optical duplicates 방식이 아닌 sequence 비교방식으로 duplicate가 이미 제거가 되었으므로 GATK 파이프라인의 picard tool을 이용한 MarkDuplicate를 거치지 않고 바로 GATK 파이프라인 상의 Indel realignment 단계를 수행하였다. 이후 Read group을 편집중인 bam file에 추가하는 과정을 거쳐 공여받은 고추 계통별로 read alignment의 편집 작업을 완료하고 SNP 및 Indel 영역의 식별을 위한 variant calling을 Vcftools을 이용하여 진행하였다.

제3절 무의 육종 특화 DB 구축을 위한 기반 정보 수집 및 생산

1. 무(Radish, *Raphanus sativus*)의 육종 특화 DB 구축을 위한 표준 유전체 정보의 출처 탐색

가. 무 현행 reference genome assembly의 활용을 위한 관련 정보의 download

○ 무의 reference genome assembly인 wonkwang(ver 1.0)을 Radish genome database (<http://www.radish-genome.org/>)로부터 다운로드하였다. 본 reference genome의 정보는 gene annotation을 통해 예측된 protein의 기능 정보 연결 및 무 6 계통으로부터 생산된 NGS 데이터에서의 변이데이터의 생산에 활용되었다(그림1).



그림 1. Radish genome database 메인 페이지(<http://www.radish-genome.org/>)

나. 무의 predicted protein의 기능 확인

○ Wonkwang genome assembly에서 총 46,514여개의 단백질이 genome 상에서 예측이 되었고 그 coding region 및 염기 및 단백질 서열의 활용이 가능하다. 그러나 예측된 단백질 서열의 서열 정보만이 활용이 가능한 실정이다. 이러한 서열 정보의 활용성을 증대하기 위하여 무의 예측된 유전자와의 염기서열 유사도가 높으며 기타 기능 정보를 보유한 다른 서열의 정보를 유용하는 방법을 제시할 수 있다. 따라서 UniProt 데이터베이스에서 제공하는 protein database 파일 및 protein 별 기능 검색 service를 활용하여 무에서 예측된 46,514개 gene model의 기능 정보를 연결하는 작업을 수행하였다(그림2).

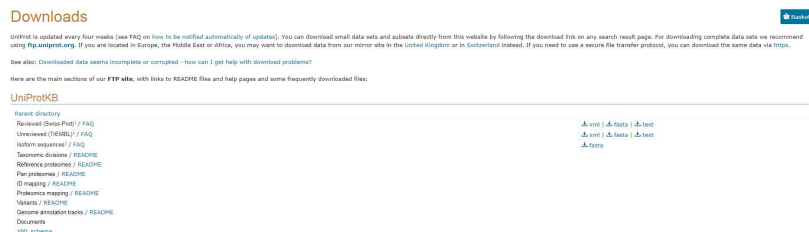


그림 2. UniProt의 protein database 다운로드 웹 페이지(<https://www.uniprot.org/downloads>)

○ 다운로드가 완료된 UniProt protein database를 makeblastdb를 통해 BLASTP를 위한 데이터베이스 파일을 구성하고 무의 예측된 단백질 서열 46,514개를 query로 활용한 BLASTP 분석을 수행하였으며 e-value threshold는 10^{-6} 으로 적용되었다. BLASTP 결과는 in-house perl script 및 shell script로 parsing하여 multiple mapping case를 제거하였다. 또한 query 및 subject coverage를 계산하고 두 coverage 수치가 70%를 넘는 결과만을 필터링 하였다. 하나의 무 gene ID가 복수의 uniprot sequence와 관계를 갖는 경우 필터링 결과를 query 및 subject coverage의 내림차순 정렬 후 가장 최상의 uniprot ID를 homologue로서 채택하였다. 그 결과 무의 예측 유전자 23,086개에 대하여 12,376개의 UniProt ID를 연결시킬 수 있었다. 해당 Uniprot ID이 연결되어 있는 기타 public database의 정보의 연결을 통해 표 1과 같은 무의 예측 유전자의 기능 및 단백질에 관련된 정보의 확장이 가능하다(표 1).

표 1. Uniprot ID와 연결된 public database의 정보 일람

| Contents | Database | Address of web-database |
|------------------------|----------|--|
| RefSeq ID | NCBI | www.ncbi.nlm.nih.gov/refseq |
| PubMed ID | NCBI | www.ncbi.nlm.nih.gov/pubmed |
| EC number | ExPASy | enzyme.expasy.org |
| Gene ontology | GO | www.geneontology.org |
| Protein interaction | STRING | string-db.org |
| Pathway | KEGG | www.genome.jp/kegg |
| Domain information | Pfam | pfam.xfam.org |
| Protein classification | PANTHER | www.pantherdb.org |
| Conserved domain | NCBI | www.ncbi.nlm.nih.gov/cdd |

2. 무 6개 계통의 변이데이터 생성

가. 무 변이 데이터 구성을 위한 sequencing 데이터의 생산

○ 집단을 이루는 각 계통의 Sequencing 데이터는 Illumina의 Hiseq2000 platform으로 생산되었으며 Paired End 형식으로 약 500 bp 내외의 insert size가 적용되어 구성되었다(표 2).

표 2. 충남대학교 무 sequencing 6개 계통의 기본 정보

| Number | Accession ID | Total read number | Total base pair | Sequencing coverage(X) |
|--------|--------------|-------------------|-----------------|------------------------|
| 1 | 835 | 125,904,906 | 12,716,395,506 | 24.94 |
| 2 | B2 | 198,896,772 | 20,088,573,972 | 39.39 |
| 3 | YR4-1 | 168,290,502 | 16,997,340,702 | 33.33 |
| 4 | YR31-1 | 189,443,916 | 19,133,835,516 | 37.51 |
| 5 | YR39-1 | 190,042,688 | 19,194,311,488 | 37.63 |
| 6 | YR106-2 | 208,986,506 | 21,107,637,106 | 41.39 |

나. Sequencing 데이터의 filtering 및 reference genome assembly상의 alignment

○ Illumina sequencing platform을 통해 구성된 계통별 sequence 서열을 seeders의 ftp 서버를 경유하여 충남대학교 서버로 다운로드를 완료하였다. 그리고 fastqc software를 사용하여 각 계통별로 생산된 sequence read의 coverage외에 read의 quality 정보와 같은 기본적인 사항을 확인할 수 있었다(그림 3, 4).

○ Fasqc 이후 Trimmomatic software를 통해 각 계통의 read sequence에 대한 quality filtering을 수행하였다. 그리고 BWA를 활용하여 filtering이 완료된 cleaned read sequence를 wonkwang reference genome에 대한 alignment를 수행하였다.

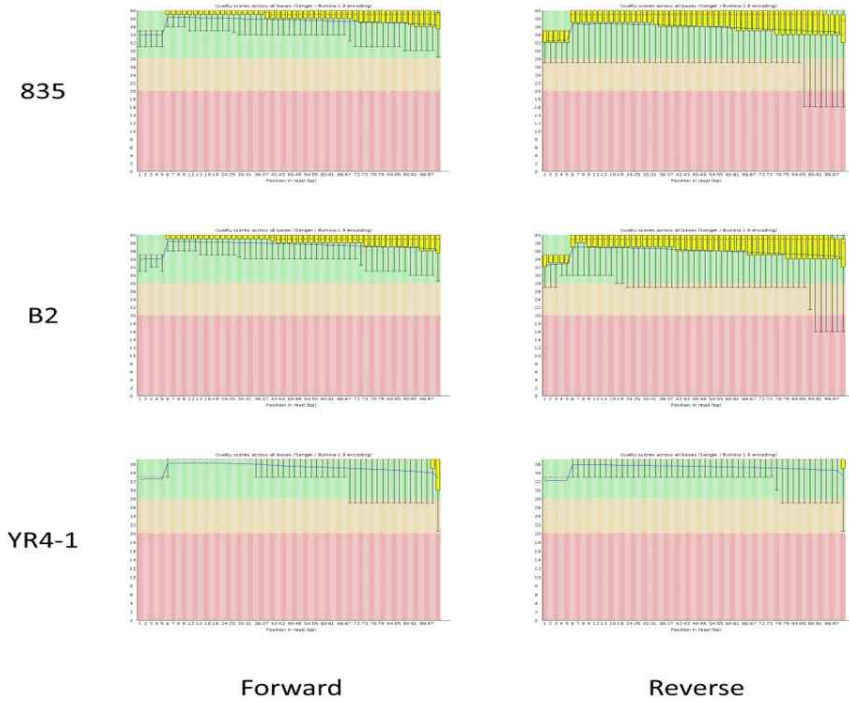


그림 3. 무 paired end sequence의 read error rate (계통명 835, B2, YR4-1)

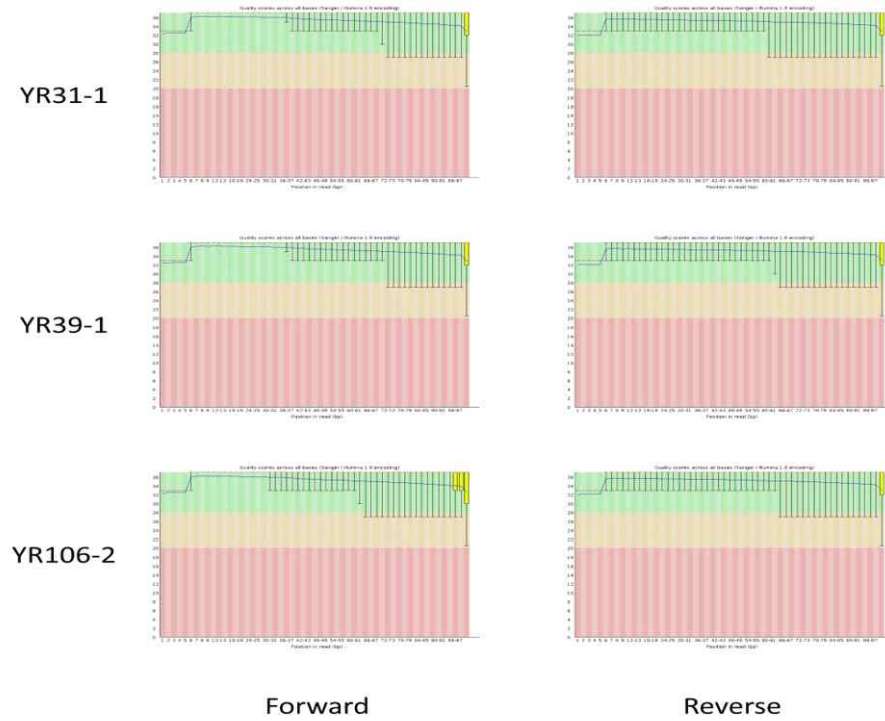


그림 4. 무 paired end sequence의 read error rate (계통명 YR31-1, YR39-1, YR106-2)

다. 무 6개 계통에 대한 변이 데이터를 위한 alignment 결과의 처리

○ Alignment 이후 생성된 sam 파일은 samtools의 view option을 통해 bam 파일로 전환하였으며, 파일 전환 과정에서 MAPQ가 30 이하인 mapped read는 제거되었다. 이후 picard tools의 MarkDuplicate 기능을 통해 mapped read 중 PCR duplicate는 marking 및 bam 파일 내에서 제거되었다. 이후, GATK의 package를 통해 read가 mapping된 region 상에서 indel이 나타난 위치를 식별 후 이의 위치정보를 구성한 다음, 해당 영역의 read를 다시 alignment를 하는 과정을 거치는 작업을 수행하였다. 그리고 각 bam 파일의 sample 및 group 정보의 부여는 GATK의 AddOrReplacement를 통해 수행되었다.

○ 이와 같은 과정을 거친 후 variant calling에 활용할 bam 파일을 최종적으로 구성하였으며 각 계통별 bam 파일이 나타내는 sequence read의 mapping 양상은 qualimap software를 통해 해당 통계 자료 및 시각화된 자료의 형태로 구성할 수 있었다. 자료의 구성에 반영된 6개 계통의 bam파일의 mapping quality 평가를 통해 6개 계통에서 공통적인 genome assembly 상의 위치에서 높은 수준의 mapped read count가 발생하는 것을 확인할 수 있었으며 scaffold에 mapping 된 read들은 특히 높은 수준의 mapping count와 동시에 낮은 mapping quality를 나타내었다. 이는 무의 genome 자체에서 나타나는 중복된 영역이 나타내는 특성이라 사료되며, 특히 reference genome 파일에서 나타나는 scaffold들이 높은 수준의 반복서열로 인해 genome assembly 과정에서 chromosome으로 통합되지 못한 것이라는 사실을 통해 해당 결과를 해석할 수 있다(그림 5, 6).

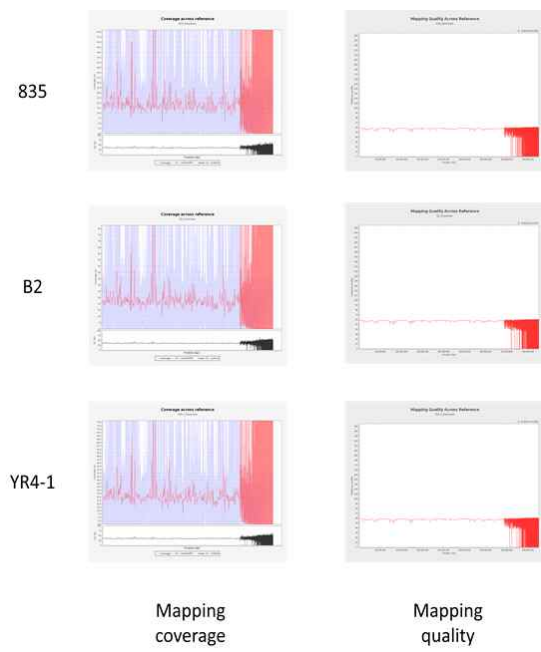


그림 5. 계통별 wonkwang genome assembly에 대한 bam 파일의 mapping rate(계통명 835, B2, YR4-1)

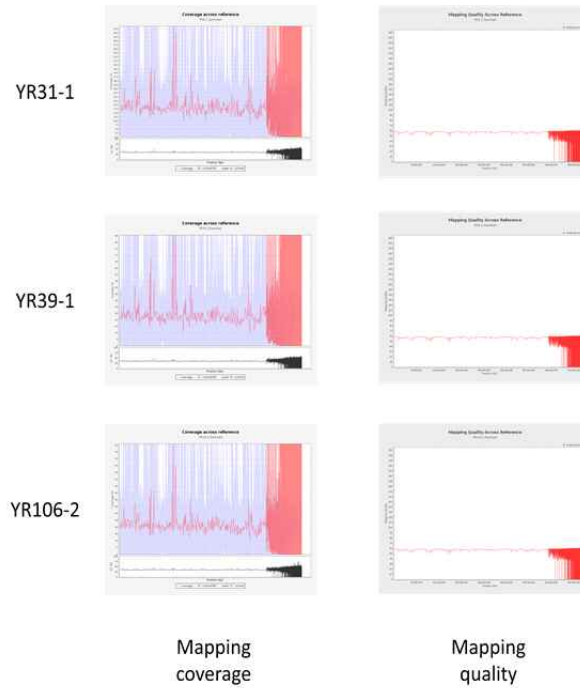


그림 6. 계통별 wokkwang genome assembly에 대한 bam 파일의 mapping rate (계통명 YR31-1, YR39-1, YR106-2)

라. 무 6개 계통에 대한 SNP calling

○ PCR duplicate의 제거 및 Indel realignment가 완료된 무 6개 계통의 bam 파일로부터 Samtools와 Bcftools를 pipe로 결합한 명령어를 통해 variant calling을 수행하였다. 1차적으로 각 계통에서 구성된 raw VCF 파일은 vcftools를 통해 homozygous alternative SNP only, biallelic SNP only, minDP 5, minGQ 30, 및 minQ 30 조건으로 필터링 되었다. 그리고 in-house perl script를 통해 각 계통에서 나타난 모든 filtered SNP position을 추출하고 6개 계통 내에서 발생이 가능한 모든 SNP의 리스트를 text 파일로 구성하였다.

○ 무 6개 계통에서 나타날 수 있는 모든 SNP의 리스트와 각 계통의 bam 파일이 위치한 server 내의 경로를 input 파일로서 구성하였다. 계통별 SNP calling과 같이 Bcftools와 Samtools를 결합한 명령어를 통해 무 6개 계통에서 나타날 수 있는 SNP position들에 한정된 joint variant calling을 수행하였다.

○ 6개 계통의 joint calling 결과는 biallelic site only 및 minDP 5 조건으로 최종 필터링 되었으며 genome assembly 상의 각 chromosome별 length에 따라 SNP의 밀도를 계산하였다. 그 결과, 무 wokkwang genome assembly ver 1.0에 대하여 총 30,80,560개의 SNP를 산출하였으며 1000bp 당 8.95개의 SNP를 갖는 것으로 나타났다 (표 3).

표 3. 무 6개 계통의 wonkwang genome assembly에 대한 SNP 산출 결과

| Chromosome | Assembly size (bp) | SNP number | SNP density (SNPs/1kb) |
|------------|--------------------|------------|------------------------|
| R1 | 26,309,735 | 278,304 | 10.58 |
| R2 | 43,799,612 | 396,603 | 9.05 |
| R3 | 29,132,933 | 276,409 | 9.49 |
| R4 | 50,002,108 | 389,546 | 7.79 |
| R5 | 45,943,323 | 403,011 | 8.77 |
| R6 | 53,636,577 | 541,080 | 10.09 |
| R7 | 27,187,321 | 234,740 | 8.63 |
| R8 | 29,681,327 | 246,948 | 8.32 |
| R9 | 38,354,807 | 313,919 | 8.18 |
| Total | 344,047,743 | 3,080,560 | 8.95 |

3. 무 열근 현상의 유전분석

가. 무의 열근(cracking) 현상

○ 열근(cracking) 현상은 근권부가 가식부로서 활용되는 무의 원예적 가치를 저해하는 요소이다. 이는 무의 재배과정에서 뿌리가 비대하는 과정에서 발생하는 현상으로 수확기가 늦거나 부적절한 관수가 주된 원인으로서 나타나는 현상이다. 주로 뿌리의 어깨, 몸통 및 꼬리 부분이 가로 혹은 세로 방향으로 갈라지는 양상을 보이며 뿌리 내부 조직의 비대가 외부 조직의 비대보다 빠를 때 발생한다. 또한 열근 부위를 중심으로 조직의 코르크화가 진행되어 상품으로서의 가치를 잃게 되며 가공성 또한 저하되는 문제를 유발한다(그림 7).

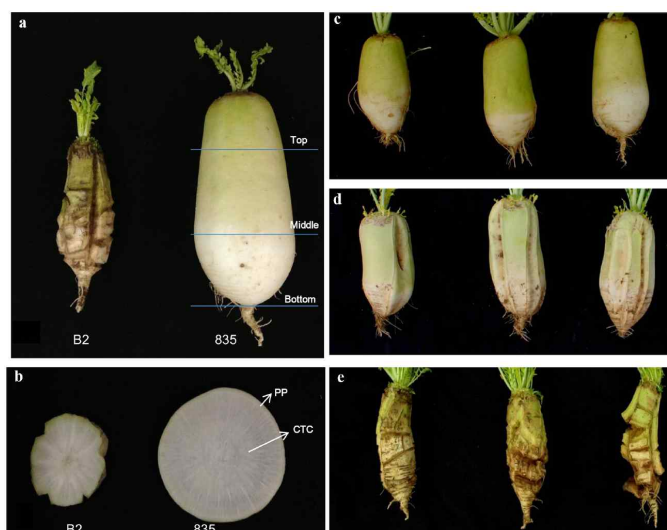


그림 7. 무의 열근 현상의 유전분석에 활용된 양친 (B2, 835)

나. 무의 열근(cracking) 현상의 유전분석

○ 무의 열근 현상의 유전분석을 위해 열근 저항성인 835와 감수성인 B2를 양친으로 활용한 bi-parental mapping population을 구성하였다. F_{2:3} 집단의 표현형 분석과 genotyping을 WinQTLcart 2.5 software로 처리하여 무의 linkage map 상에서 열근 형질에 연관된 QTL을 식별하였다. (표 4, 그림 8)

표 4. 무의 열근 형질의 연관된 QTL의 정보

| No. of loci | Loci name | Linkage group | Year | Block region | Confidence interval (cM) | LOD | R ² (%) | Additive | Dominant |
|-------------|-------------|---------------|------|--------------|--------------------------|------|--------------------|----------|----------|
| 1 | <i>qCr1</i> | 1 | 2012 | R | 13.7–16.9 | 3.43 | 4.72 | 0.12 | 0.42 |
| 2 | | 1 | 2013 | R | 14.9–27.3 | 3.47 | 18.11 | 0.25 | 0.20 |
| 3 | | 1 | 2014 | R | 12.3–18.5 | 3.55 | 4.47 | 0.25 | 0.05 |
| 4 | <i>qCr2</i> | 2 | 2012 | | 121.5–126.4 | 3.09 | 2.16 | -0.05 | -0.43 |
| 5 | <i>qCr3</i> | 2 | 2014 | J | 34.1–42.9 | 3.32 | 3.75 | -0.02 | 0.26 |
| 6 | <i>qCr4</i> | 4 | 2014 | S | 68.8–77.4 | 3.65 | 10.07 | 0.13 | -0.18 |
| 7 | <i>qCr5</i> | 5 | 2013 | M | 109–113.4 | 4.56 | 7.77 | 0.34 | 0.02 |
| 8 | <i>qCr6</i> | 8 | 2012 | I | 27.3–38.7 | 2.93 | 1.41 | -0.25 | -0.22 |
| 9 | <i>qCr7</i> | 8 | 2012 | R | 99.3–107.5 | 2.68 | 9.76 | 0.07 | -0.89 |
| 10 | <i>qCr8</i> | 9 | 2014 | A | 54.5–56.7 | 3.09 | 6.09 | -0.03 | 0.28 |

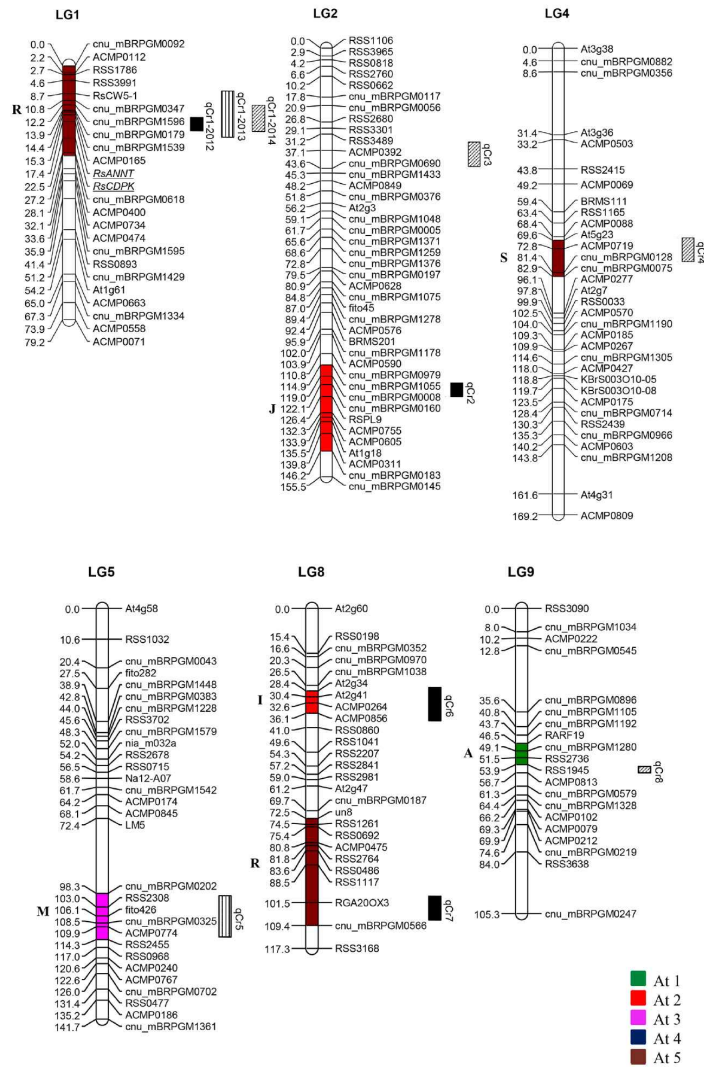


그림 8. 무의 linkage map 상의 열근 형질의 연관된 QTL의 시각화

4. 무의 육종 특화 DB의 구축

가. 무의 육종 특화 DB의 플랫폼 구성

○ 기존의 채소육종특화데이터 베이스(<http://168.188.15.201/vege/>)의 플랫폼에 금년도의 작업 결과를 반영하기 위한 별도의 플랫폼을 개발하였다(<http://168.188.15.201/vradish>). 현재 genome browser를 기반으로 한 전체 정보의 통합 검색 체제를 구축하였다. 해당 genome browser를 통해 무의 wonkwang genome assembly(ver 1.0) 상의 예측된 모든 유전자의 위치 정보 및 염기/단백질 서열의 출력이 가능하며 gene model의 경우 intron 및 exon의 위치 또한 식별이 가능하다 (그림 9, 그림 10).



그림 9. 무의 육종 특화 데이터베이스의 genome browser platform

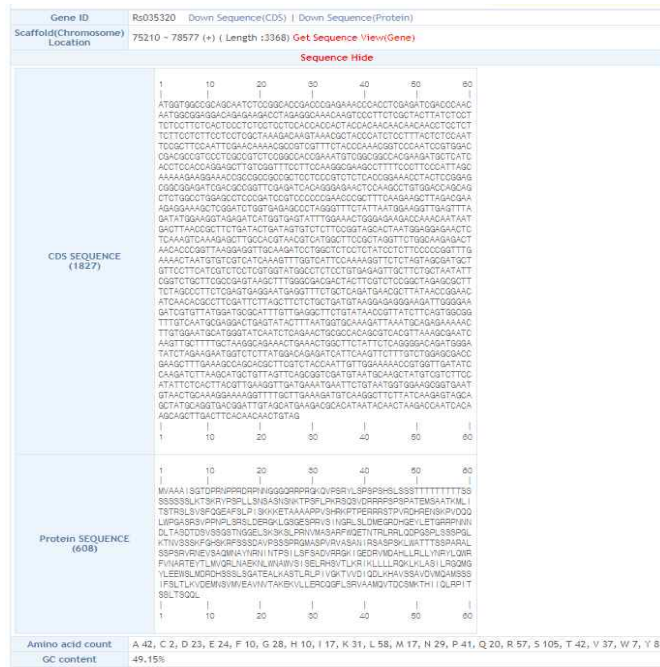


그림 10. Genome browser를 통해 출력된 무의 gene model Rs035320의 기본 정보의 출력 결과

나. BLAST 기능 탑재를 통한 상동 유전자의 검색 플랫폼의 구성

○ 데이터베이스 플랫폼 상에 web BLAST tool을 탑재하여 사용자가 보유한 염기 및 단백질 서열과 상동인 유전자의 검색 체계를 구현하였다. 출력된 결과는 genome browser 플랫폼에 입력된 정보와 직접적으로 연결된다. 이를 통해 연구자가 보유한 sequence의 상동 유전자 및 해당 유전자에 대해 무 육종 데이터베이스가 보유할 예정인 무의 수집된 변이 및 유전자 발현량과 같은 육종에 활용 가능한 정보에 대한 접근성을 높였다(그림 11).

Radish

Vegetable HOME Genome Browse BLAST 관리자 [logout](#)

Home > Blast Search

BLAST Search
Choose program to use and database to search:

Program Database

Enter sequence below in FASTA format

Or load it from disk [파일 선택](#) 선택된 파일 없음

Set subsequence: From To

The query sequence is filtered for low complexity regions by default.
Filter Low complexity Mask for lookup table only

Expect Matrix Perform ungapped alignment

Query Genetic Codes (blastx only)

Database Genetic Codes (tblast[nx] only)

Frame shift penalty for blastx

Other advanced options:

Alignment view

Descriptions Alignments Color schema

Comments and suggestions to: < blast-help@ncbi.nlm.nih.gov >

Last modified: Jan 11, 2002

그림 11. 무의 육종 특화 데이터베이스의 탑재된 BLAST search tool

5. 배추 NGS 데이터를 활용한 특정 패스웨이 유전자 상의 변이 대량 발굴

가. NovaSeq 6000 platform을 활용한 배추 NGS 데이터의 업데이트

○ 구축한 배추 수집단내에서 일부 sequencing depth가 낮은 계통의 데이터 업데이트 및 배추 육종사에서 중요한 계통의 도입을 목적으로 추가적인 sequencing을 통해 총 26개 계통에 대한 paired-end 형식의 short read sequencing data를 생산하였다(표 5).

표 5. 신규 도입 및 업데이트된 배추 26개 계통 정보

| # | Accession ID | Species | Description | Institute/university | Notice |
|----|--------------|---|-----------------------------|----------------------|--------|
| 1 | csr188-2 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 권심배추 | 도호쿠 대학 | New |
| 2 | csr187-3 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 아이치 배추 | 도호쿠 대학 | New |
| 3 | csr184-5 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 미치릴리 | 농촌진흥청 | New |
| 4 | csr181-2 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 하두 | 농촌진흥청 | New |
| 5 | csr178-2 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 팔월육 | 농촌진흥청 | New |
| 6 | csr185-1 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 백색포두련 | 도호쿠 대학 | New |
| 7 | csr183-2 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 산동채 | 농촌진흥청 | New |
| 8 | csr169-2 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 청방 핵도문 | 충남대학교 | New |
| 9 | csr186-5 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 화심배추 | 도호쿠 대학 | New |
| 10 | csr180-2 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 경도 3호 | 농촌진흥청 | New |
| 11 | csr167-2 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 무쌍 (F1) | 충남대학교 | New |
| 12 | csr190-2 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 결구 지부배추 | 도호쿠 대학 | New |
| 13 | csr168-2 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 춘추 (F1) | 충남대학교 | New |
| 14 | csr179 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 포두련 | 농촌진흥청 | New |
| 15 | csr158-2 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 叶長 黑殼 疊抱 晚抽臺 内叶黄色 抗病 青白帮 | 충남대학교 | Update |
| 16 | csr175 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 시모야마 치토세 | 농촌진흥청 | New |
| 17 | csr51-5 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | ZS072518 / 내흔계 | 충남대학교 | Update |
| 18 | csr159-4 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 舒心 黄心口 白帮 叶色淺 | 충남대학교 | Update |
| 19 | csr124-2 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 小獅子頭 | 충남대학교 | Update |
| 20 | csr128-2 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 錦424-S8 | 충남대학교 | Update |
| 21 | csr177-2 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 60일 | 농촌진흥청 | Update |
| 22 | csr111 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 秋玉一号F1 | 충남대학교 | Update |
| 23 | csr030 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 09-FK18(대형가락) | 충남대학교 | Update |
| 24 | csr118-5 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 丰盈 | 충남대학교 | Update |
| 25 | csr032-5 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 09-FK24(경도2호) | 충남대학교 | Update |
| 26 | csr176-2 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 노자키 2호 | 농촌진흥청 | New |

○ 신규 계통의 도입으로 인하여 배추 수집단의 구성 계통은 156개에서 173개로 증가하였으며 전체 173개 계통 중 배추 유전체 크기 대비 30X 이상으로 sequencing이 이루어진 계통은 120개가 되었으며 기존의 low depth를 갖던 9개 계통의 데이터가 업데이트 되었다(표 6, 그림 12). 신규 데이터 업데이트에 따라 일본 육성 계통 및 중국 북방계에 해당하는 계통의 수가 가장 많은 수의 증가를 나타내었다. 반면 국내 재래종 및 비배추 그룹에 속하는 계통에는 변화가 없었다.

표 6. 배추 신규 NGS 데이터의 업데이트 결과

| Category | 2019 | 2020 |
|-------------------|------|------|
| Chinese - | 42 | 43 |
| Chinese northern | 10 | 14 |
| Chinese shandong | 1 | 4 |
| Chinese southern | 5 | 6 |
| Japanese breeding | 6 | 12 |
| Korean breeding | 72 | 74 |
| Korean landrace | 5 | 5 |
| Non-pekinensis | 15 | 15 |
| Total | 156 | 173 |

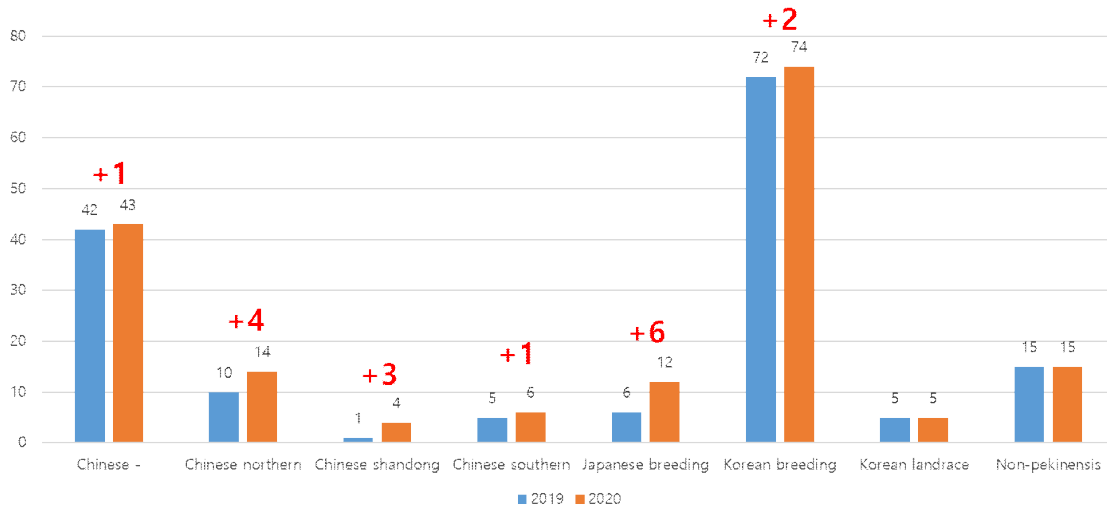


그림 12. 신규 데이터 업데이트에 따른 배추 수집단 내의 소분류에 따른 변화

나. 신규 NGS 데이터를 활용한 배추 수집단의 변이 정보의 산출

○ 4차년도에 총 173개 계통으로 이루어진 집단의 크기가 증가된 데이터 세트로부터 변이 데이터의 산출을 다시 진행하여 기존의 SNP 산출 결과 또한 업데이트를 진행하였다. 변이 데이터의 산출은 이전에 구축한 파이프라인을 통해 수행하였으며 계통별 변이의 산출 이후 전체 집단에서 발생이 가능한 모든 위치의 SNP에 대하여 joint variant calling 및 후속 filtering 작업을 진행하였다(표 7).

표 7. Filtering에 따른 집단 SNP 개수의 변화

| Step | Process | Number of SNPs in the population |
|------|--|----------------------------------|
| 1 | Joint variant calling | 8,235,625 |
| 2 | Multi-allelic position filtering | 7,860,835 |
| 3 | Tassel 5.0 (MAF, Missing rate, Heterozygosity) | 2,514,850 |
| 4 | LD pruning | 49,433 |

집단의 분석이 가능한 최대 SNP 개수를 갖는 Tassel 5.0에 의한 필터링 결과를 활용하여 배추 표준유전체(ver 3.0) 상에서의 변이의 밀도는 표8 와 같다.

표 8. 배추 수집단의 표준유전체(ver 3.0) 상에서의 식별된 SNP 및 변이의 밀도

| # | Size of Genome assembly (<i>B. rapa</i> annotation ver 3.0) | Number of identified SNPs | Density (SNPs/1kb) |
|-------|---|---------------------------|--------------------|
| A01 | 29,595,527 | 266,056 | 8.99 |
| A02 | 31,442,979 | 276,227 | 8.79 |
| A03 | 38,154,160 | 322,666 | 8.46 |
| A04 | 21,928,416 | 188,976 | 8.62 |
| A05 | 28,493,056 | 234,658 | 8.24 |
| A06 | 29,167,992 | 272,410 | 9.34 |
| A07 | 28,928,902 | 239,437 | 8.28 |
| A08 | 22,981,702 | 190,969 | 8.31 |
| A09 | 45,156,810 | 354,253 | 7.84 |
| A10 | 20,725,698 | 169,198 | 8.16 |
| Total | 296,575,242 | 2,514,850 | 8.48 |

○ Tassel 5.0 필터링 결과와 LD pruning 결과의 SNP 밀도를 구하고 이를 배추 유전체 상의 repeat 발생 빈도와 함께 시각화를 수행하였을 때, 필터링 방식의 관계 없이 배추 SNP 밀도의 패턴은 10개 크로모솜에서 유사하게 나타났으며 repeat의 빈도가 높아지는 영역 및 동원체가 예측된 영역에서는 SNP의 밀도가 낮게 나타났으며 이는 3번 크로모솜의 말단 부분에서 명확하게 드러났다(그림 13).

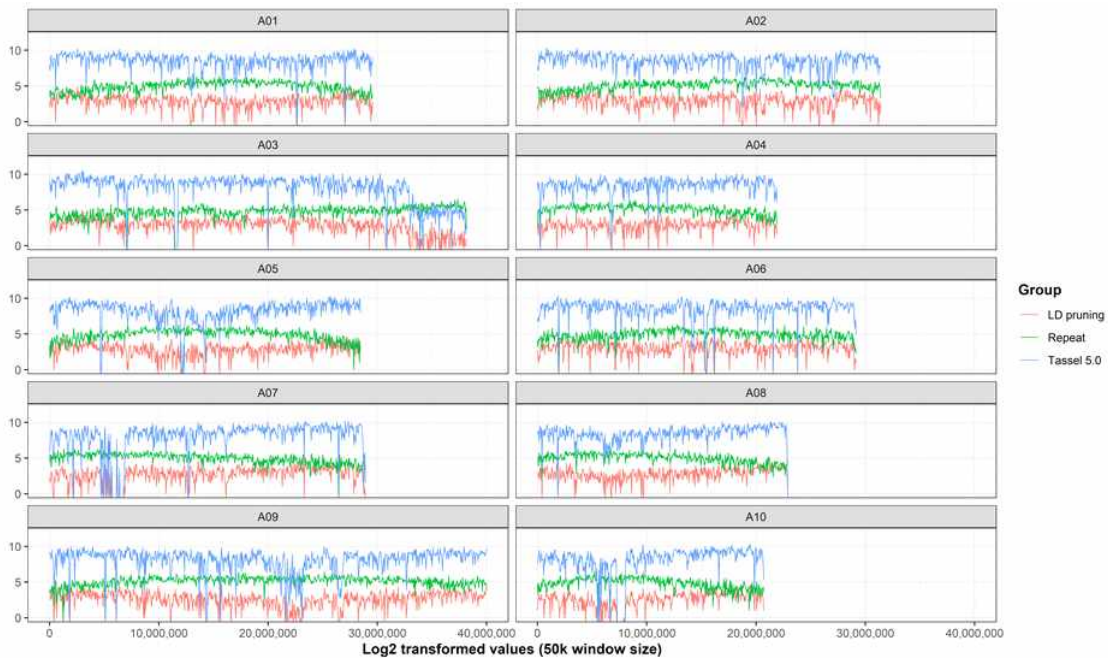


그림 13. 배추 표준유전체의 10개 크로모솜에서의 수집단의 SNP 밀도

○ Tassel 5.0을 활용하여 계통별 SNP로부터 distance matrix를 산출하고 이를 활용하여 MDS plot을 구성한 결과 배추의 분류 및 육종 시기에 따라 집단내 계통들이 클러스터링 되는 것을 확인할 수 있었다(그림 14).

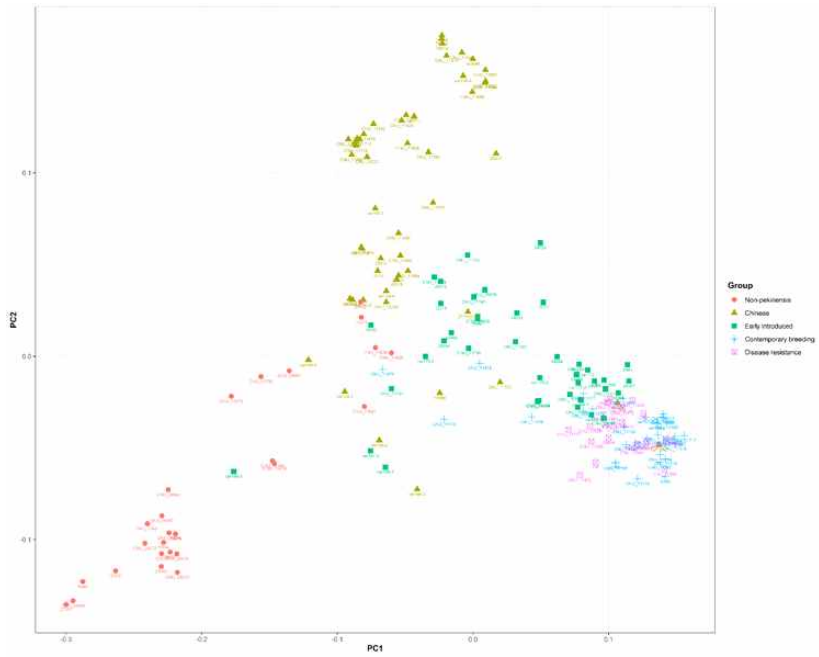


그림 14. 배추 수집단을 구성하는 계통의 클러스터링 결과

○ PopLDdecay를 활용하여 배추 수집단을 구성하는 계통의 도입 및 육종시기에 따른 linkage disequilibrium(LD)을 시각화하여 비배추, 중국계 및 초기 도입 계통, 국내 육성 계통 및 병저항성 계통 순으로 LD decay가 느리게 나타나는 것을 확인할 수 있었다. 이를 통해 육종활동에 의한 집중적인 선발을 받은 집단일수록 높은 수준의 LD를 나타내는 것을 확인할 수 있었다(그림 15).

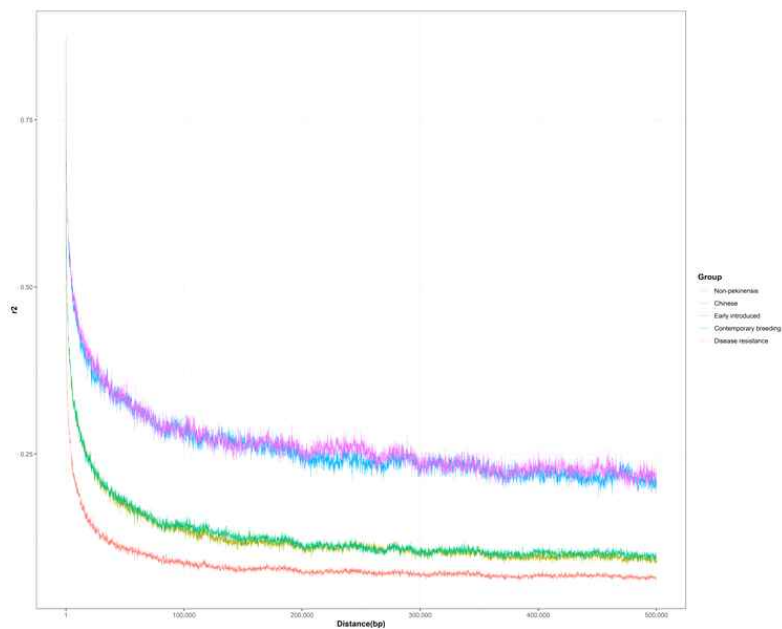


그림 15. 배추 수집단 내의 소그룹에 따른 LD decay

○ StRauto를 활용하여 배추 수집단 내의 소그룹을 예측하였을 때 집단은 크게 2개(K=2)의 분류로 나누어지며(배추, 비배추) K=6에서 배추 수집단 내의 다양한 기원지의 배추를 반영하는 것으로 나타났다(그림 16).

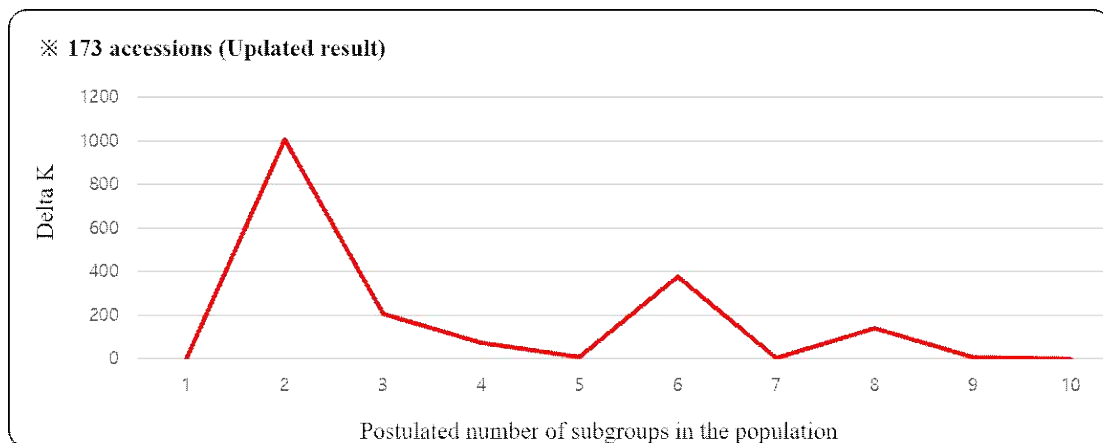


그림 16. 배추 수집단을 이루는 173개 계통이 이루는 소분류의 식별

○ 그러나 SNPhylo를 활용하여 173개 계통의 변이 정보를 활용한 계통도를 작성하였을 때 중국계 배추가 나타나는 소그룹에서 상대적으로 후대에 육성된 일본계 배추의 그룹이 삽입되는 것을 확인할 수 있었다 (그림 17).

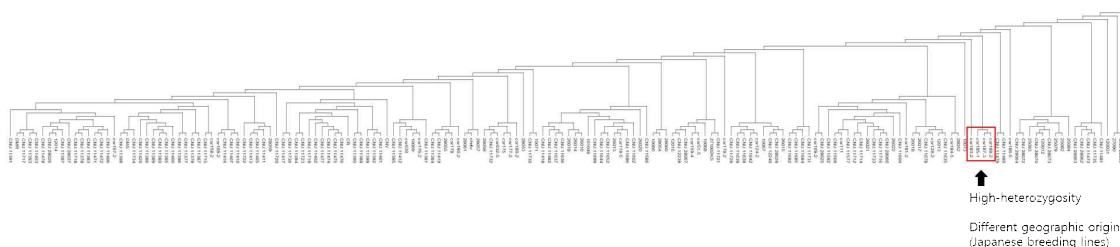


그림 17. 계통도 내에서 비배추와 중국계 배추가 이루는 clade 사이에 삽입된 신규 도입 계통의 소그룹

○ 계통도내에서 예상과 다른 영역에 clade가 나타난 계통들의 신규 계통과 기존 계통의 heterozygosity를 비교한 결과 신규 계통의 heterozygosity가 매우 높은 수준으로 나타난 것을 확인할 수 있었다(그림 18). 이는 충남대학교에서 육성된 배추 수집단의 경우, 반복적인 자가수분(selfing) 및 반수체배가화(haploid doubling)를 통해 육성이 되었으나 신규 도입 계통의 경우, 계통의 육성 및 유지 방식이 이와 달라 높은 수준의 heterozygosity를 나타내게 된 것으로 보인다. 이와 같이 heterozygosity가 큰 차이를 보이는 계통들이 하나의 집단으로 분석이 되었을 때 집단 내 계통의 클러스터링 분석 결과 및 해석에 큰 영향을 미칠 수 있을 것으로 사료된다.

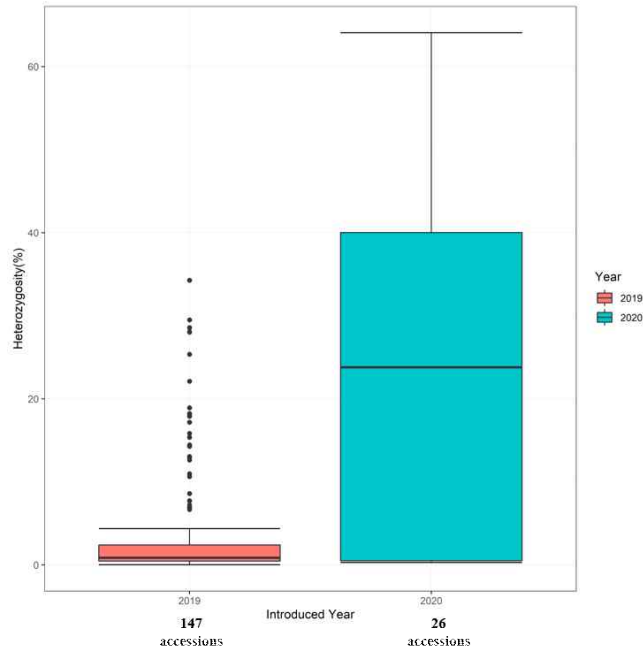


그림 18. 배추 수집단의 기존 계통 및 신규 계통의 heterozygosity 비교

나. NCBI SRA를 활용한 작물의 resequencing 데이터 수집 및 가공

○ NCBI SRA에서 배추(*Brassica rapa*), 무(*Raphanus sativus*), 고추(*Capsicum annuum*), 그리고 수박(*Citrullus lanatus*) 4개 작물에 대하여 등록된 NGS 데이터를 활용하기 위하여 검색을 수행하였다. 불필요한 양식의 데이터의 검색을 차단하기 위하여 keyword는 DNA, Illumina, Paired-end, fastq 파일로 한정하여 검색을 수행하였으며, 검색 결과를 csv 파일로 형식으로 다운로드한 이후 해당 파일을 BioProject 별로 집계한 결과 4개 작물에 대하여 총 1,019개의 resequencing 데이터가 활용 가능한 것을 확인하였다(표 9).

표 9. SRA로부터 수집된 4개 작물의 resequencing 데이터

| # | Species | BioProject | Description | Number of SRA files |
|----|----------------------|-------------|---|---------------------|
| 1 | <i>Brassica rapa</i> | PRJNA546028 | Whole-genome sequencing of Chinese cabbage inbred line CT001, a line widely used in traditional or molecular breeding, was conducted. Also, the spontaneous mutation rate of inbred line was estimated using resequencing data. | 3 |
| 2 | <i>Brassica rapa</i> | PRJNA542546 | High-density mapping of QTL for the main flower stalk length was performed based on the deep resequencing of parental lines and specific locus-amplified fragment sequencing (SLAF-Seq) of individual recombination inbred lines. The results of this study provide an important platform for gene/QTL mapping and marker-assisted selection (MAS) breeding for bolting-resistant Chinese cabbage | 151 |
| 3 | <i>Brassica rapa</i> | PRJNA564416 | Genome-wide identification of SNPs in Chinese cabbage | 4 |
| 5 | <i>Brassica rapa</i> | PRJNA533728 | Genome divergence in <i>Brassica rapa</i> subspecies revealed by whole genome analysis on doubled-haploid line of Korean Ganghwa turnip | 3 |
| 6 | <i>Brassica rapa</i> | PRJNA486260 | Chinese cabbage is the most consumed leafy crop in east Asian countries. Continuous low temperatures induce premature bolting, which severely decreases yield and edibility; therefore limiting the planting season and geographic distribution of the Chinese cabbage. In the past ~40 years, spring Chinese cabbage with strong winteriness has been selected to meet the temporal-spatial demand of Chinese cabbage. Here, we focused our study on the following questions to elucidate the evolutionary and selective history of Chinese cabbage: (1) which type of Chinese cabbage served as the progenitor and (2) where was the geographic origin of spring Chinese cabbage, and (3) what is the genetic basis during spring Chinese cabbage selection. | 192 |
| 7 | <i>Brassica rapa</i> | PRJNA428769 | <i>Brassica napus</i> is a worldwide multiuse crop, especially as a very important edible oil resource. It is an allopolyploid species naturally hybridized from <i>Brassica rapa</i> (AA=20) and <i>Brassica oleracea</i> (CC=19) about 7,500 years ago. Now, it mainly has three groups: <i>B. napus</i> subsp. <i>oleifera</i> , <i>B. napus</i> subsp. <i>rapifera</i> and <i>B. napus</i> subsp. <i>pabularia</i> . To study its origin and diversification, we use high-throughput RNA-seq and GSS methods sequenced hundreds of <i>B. napus</i> and its two parental species. | 79 |
| 8 | <i>Brassica rapa</i> | PRJNA385301 | The #174-12-26 line is a BC3F2 lineage, hybrid introgressed line derived from a crossing of commercial cultivar <i>Brassica oleracea</i> var. Reiho No. 1 and commercial cultivar <i>Brassica rapa</i> L. <i>pekinensis</i> var. CR Kanki that was then repeatedly backcrossed to parental <i>B. rapa</i> cultivar. These sequenced reads were used to verify the IntroMap software's ability to detect introgressed genomic regions from a recurrent parental line, in this case <i>B. oleracea</i> genomic regions introgressed into a <i>B. rapa</i> genomic background. | 4 |
| 9 | <i>Brassica rapa</i> | PRJNA385249 | To further elucidate the mechanisms behind heterosis in Chinese cabbage, we re-sequenced the parental inbred lines (S11, R09) of the commercial F1 hybrid cultivar of Chinese cabbage, "W77", that exhibits increased yield as compared to its parental lines. | 2 |
| 10 | <i>Brassica rapa</i> | PRJNA281682 | <i>Brassica rapa</i> includes some of the most important vegetables worldwide as well as oilseed crops. The complete annotated genome sequence confirmed its paleohexaploid origins and provides opportunities for exploring the detailed process of polyploid genome evolution. We generated a genome-wide DNA methylation profile for <i>B. rapa</i> using a modified reduced representation bisulfite sequencing (RRBS) method. This sampling represented 2.24% of all CG loci (2.5×10^5), 2.16% CHG (2.7×10^5) and 1.68% CHH loci (1.05×10^5). It was found that genic regions of single copy genes had significantly higher methylation compared to those of two or three copy genes. Differences in degree of genic DNA methylation were observed in a hierarchical relationship corresponding to the relative age of the three ancestral subgenomes, primarily accounted by single-copy genes. RNA-seq analysis revealed that overall the level of transcription was negatively correlated with mean gene methylation content depending on copy number or association with the different subgenomes. These results provide new insights into the role epigenetic variation plays in polyploid genome evolution, and suggest an alternative mechanism for duplicate gene loss. | 1 |
| 11 | <i>Brassica rapa</i> | PRJNA174067 | Non-heading Chinese cabbage. | 2 |
| 12 | <i>Brassica rapa</i> | PRJNA273786 | Genomic sequencing for two cultivars of <i>Brassica rapa</i> subspecies turnip crops, | 2 |

'Tsuda' and 'Yurugi Akamaru'. To identify the mutations between turnip and Chinese cabbage.

| | | | | | |
|----|--------------------------|-------------|--|-------|------|
| 13 | <i>Raphanus sativus</i> | PRJEB10925 | Radish (<i>Raphanus sativus</i> L.) has a long history of domestication and its origin is controversial. High-depth resequencing analysis of 17 accessions revealed that Asian cultivated types are closely related to wild Asian species but distinct from European cultivated types, supporting the notion that Asian cultivars were domesticated from wild Asian species. SNP comparison between Asian genotypes coupled with expression profiling identified domestication-related gene candidates across the genome, many of which were related to root architecture, cell wall, and sugar metabolism. Overall, this study provides novel insight into evolution of the mesohexaploid genomes of the tribe Brassiceae and the selection of genes during domestication. | 45 | |
| 14 | <i>Raphanus sativus</i> | PRJNA278500 | To date, studies about repetitive elements of radish, especially the satellite repeats, are still limited. To better understand genome structure of radish, we undertook a study to evaluate the proportion of repetitive elements and their distribution in a Chinese radish inbred line. | 1 | |
| 15 | <i>Casicum annuum</i> | PRJNA223222 | Pepper Genome Sequencing | 9 | |
| 16 | <i>Casicum annuum</i> | PRJNA298503 | Resequencing of <i>Capsicum annuum</i> 'Perennial', 'Dempsey', and recombinant inbred lines (RILs) | 122 | |
| 17 | <i>Casicum annuum</i> | PRJNA186921 | Comparative study of domesticated pepper and its progenitor genome reveal signature of artificial selection. <i>C. annuum</i> var. <i>annuum</i> L. and <i>C. annuum</i> var. <i>glabriusculum</i> | 25 | |
| 18 | <i>Casicum annuum</i> | PRJNA262545 | Potyvirus resistance 4 (Pvr4) locus in pepper confers resistance to three pathotypes of Potato Virus Y (PVY) and to pepper mottle virus (PepMoV). We describe the use of next generation sequencing technology to generate molecular markers tightly linked to Pvr4. | 2 | |
| 19 | <i>Citrullus lanatus</i> | PRJNA527790 | We resequenced genomes of watermelon accessions representing all seven extant species in the <i>Citrullus</i> genus to perform phylogenomic and population genomic analyses, as well as genome-wide association studies (GWAS) for important fruit quality traits. | 337 | |
| 20 | <i>Citrullus lanatus</i> | PRJNA532463 | Fruit characteristics of dessert watermelon have been formed largely by human selection. Here we report an improved watermelon reference genome and a comprehensive genome variation map generated by resequencing 415 wild and cultivated accessions representing all extant species in the <i>Citrullus</i> genus. Genome-wide association studies identified a number of loci associated with fruit quality traits. Population genomic analyses reveal the evolutionary history of <i>Citrullus</i> , suggesting a close relationship between <i>C. lanatus</i> and <i>C. mucospermus</i> , while independent evolution in <i>C. amarus</i> . Our findings indicate that in watermelon different fruit size loci have been under selection during evolution, domestication and improvement. A non-bitter allele arose in the progenitor of dessert watermelon has been under selection during domestication and largely fixed in <i>C. lanatus</i> cultivars. Selection of flesh sweetness has been started in the progenitor of <i>C. lanatus</i> and continued during modern breeding through genes controlling raffinose catabolism and sugar transport from cytoplasm to vacuole. Fruit flesh coloration and sugar accumulation might have co-evolved through shared genetic components including the sugar transporter gene, CITST2. | 1 | |
| 21 | <i>Citrullus lanatus</i> | PRJNA532308 | Fine mapping a dwarf gene in watermelon | 4 | |
| 22 | <i>Citrullus lanatus</i> | PRJNA576063 | Bsa analysis of yellow rind (WT4) and green rind (WM102) of watermelon | 2 | |
| 23 | <i>Citrullus lanatus</i> | PRJNA551784 | watermelons with yellow and green rind | 2 | |
| 24 | <i>Citrullus lanatus</i> | PRJNA476464 | Genetic mapping using SNPs, which were generated from susceptible and resistant watermelon lines and their progeny, to find out a causal region enabling anthracnose resistance in watermelon. | 2 | |
| 25 | <i>Citrullus lanatus</i> | No info | | 24 | |
| | | | | Total | 1019 |

○ 현재 1,019개 re-sequencing 데이터에 대한 다운로드를 진행 중이며 채소종자데이터베이스에 업로드된 각 작물의 표준유전체에 대해서 variant calling을 수행 중에 있다. 또한 최근, pan-genome 연구의 대두로 인하여 하나의 작물 내의 다양한 아종(subspecies)을 대상으로 genome assembly 및 annotation이 수행 중에 있다. 따라서 추후, 각 작물별 활용이 가능한 표준유전체를 확보가 이들에 대한 변이 정보를 재생산을 수행할 계획이다.

다. NCBI SRA를 활용한 작물의 transcriptome 데이터 수집 및 가공

○ NCBI SRA에서 배추(*Brassica rapa*), 무(*Raphanus sativus*), 고추(*Capsicum annuum*), 그리고 수박(*Citrullus lanatus*) 4개 작물에 대한 transcriptome 정보를 resequencing 데이터에 대한 검색식에서 DNA 대신 RNA를 적용하여 1차적으로 확보한 결과, 총 1,152개 BioProject로부터 3,443개의 SRA파일이 식별되었다. Resequencing 데이터의 경우 BioProject의 수가 이에 비해 적고 각 BioProject ID의 목적을 수동으로 검색하여 정리 이후, 불필요한 BioProject를 선별하고 이에 대한 데이터의 활용 여부를 결정할 수 있었으나 이와 같은 방식의 적용은 transcriptome의 경우에 적용할 수 없었다. 추후 Bioproject에 대한 정보를 선별하는 방식의 확인 혹은 4대 작물의 주요 연구 사례 및 문헌을 검색하고 이에 대한 Bioproject ID 및 이에 포함된 파일을 다운로드 받아 활용하는 방식으로 데이터 다운로드 및 가공을 효율성을 높일 수 있는 방식으로 자료의 검색 방식을 결정할 것으로 생각된다.

라. 채소종자 데이터베이스 이용의 효율성 재고 방안의 검토

○ 현재 4개 작물의 resequencing 데이터(총 1,019개 SRA 파일)의 raw data가 반영하는 염기서열의 길이는 6.67 Tbp이고 파일의 저장에 필요한 용량은 3.29 Tbyte로 추산되었으나 각 raw 파일을 fastq 포맷으로 전환하게 되면 그 이상의 저장 용량이 필요할 것으로 예상된다. 이러한 대량의 생물데이터와 이를 가공하기 위해서는 높은 수준의 전산 인프라를 갖춘 기관의 지원이 필요하다. 2021년부터 시작될 바이오 연구 데이터스테이션 사업에서 각 사업단에서 구축한 데이터베이스 연계가 계획되어있으므로, 과제 종료 이후 연구 데이터스테이션인 한국생명공학연구원 국가생명연구자원정보센터(Korean bioinformation center) 또는 부처 데이터 센터인 NABIC으로 데이터 연계를 추진할 계획이다. 부처 데이터 센터인 NABIC으로 연계된 데이터도 바이오 연구 데이터스테이션에 연계되고, 데이터스테이션에서 국내 육종 회사 및 관련 연구자들이 활용할 수 있도록 데이터를 서비스 할 것으로 기대한다.

6. NCBI SRA 변이 정보의 생산 및 채소종자 데이터베이스 업데이트

가. NCBI SRA 변이 데이터의 생산

○ 현재 4개 채소작물의 resequencing 데이터를 sra 파일 형식으로 NCBI로부터 다운로드를 수행하였으며, 이는 sratoolkit의 fastqdump 기능을 통하여 fastq 파일로 전환하였다. 또한 저장용량의 축소를 위하여 gzip을 통하여 최종적으로 fastq.gz 포맷으로 압축 후 변이 정보의 생산에 활용할 수 있도록 하였다. 또한 작물의 다양한 계통별로 구성된 fastq.gz 파일로부터 본 연구에서 구축한 파이프라인을 활용하여 fastq 파일을 각 작물의 표준유전체에 clean read를 mapping하여 bam 파일을 생산한 이후 계통별 변이 데이터를 산출 중에 있으며 이는 추후 채

소중자데이터베이스의 이관 및 리뉴얼 과정에서 업데이트가 될 예정이다.

나. 작물별 거대 집단의 변이 데이터 생산의 효율성 재고

○ 현재 4개 NCBI SRA로부터 확보한 4개 작물의 활용가능한 resequencing 데이터는 배추 (*Brassica rapa*) 443개, 수박(*Citrullus lanatus*) 372개, 고추(*Capsicum annuum*) 158개, 무 (*Raphanus sativus*) 46개로 나타났다. 전체 집단의 크기가 증가함에 따라 집단 SNP를 구성하는 단계에 기존보다 많은 시스템 리소스 소모 및 분석시간의 연장이 수반되었다. 추후 데이터의 업데이트 마다 집단 변이 분석을 반복해야 하기 때문에 해당 절차에 대한 작업 효율화가 필요하여 이를 위한 별도의 파이프라인을 구축이 필요하다.

○ 작물 표준유전체는 목본성 및 야생종의 식물 유전체에 비해 대부분 이배체이며 크로모솜의 개수 또한 20개 이하로 작은 경향이 있다. 또한 작물의 경제적 중요성으로 인하여 대부분 작물 유전체는 최초 버전의 공개 이후 지속적인 업데이트가 이루어져 왔다. 최근 대두된 long read sequencing, Chromosome conformation capture, optical mapping과 같은 신규 염기서열 재분석 및 조립 기술의 발달이 적용되면서 최종 유전체 산물의 N50은 지속적으로 향상되었다. 그로인해 현재 작물의 표준유전체는 크로모솜 단위로 신뢰성 있게 조립이 되어 변이 데이터 산출에 활용이 가능하다.

○ 집단 변이데이터의 산출은 계통별로 산출한 변이를 하나로 합치는 과정으로 이 단계에서 계통별로 구성된 bam 파일이 활용된다. 이 bam 파일의 경우 비록 binary 형식으로 압축이 되었으나 파일 개당 수백 Mb에서 크기는 수 Gb의 용량을 차지하는 파일이다. 해당 파일들은 각 작물의 첫 번째부터 마지막 크로모솜까지의 표준유전체 상의 mapping 결과를 가지고 있다. 이를 이용하여 집단을 이루는 계통별로 각 크로모솜에 대한 mapping 결과를 분할하여 크로모솜 별로 집단 변이를 구성한 후 최종단계에서 하나의 파일로 통합하는 파이프라인의 설계가 가능하다.

다. 작물의 집단의 크로모솜별 변이 산출 파이프라인의 개발

○ samtools를 활용하여 작물 집단을 이루는 계통별 bam 파일을 각 크로모솜 별로 분할한 이후 집단의 크로모솜 별 변이를 산출하는 방식으로 배추 156개 계통 기준으로 1주일이 소요되는 작업시간을 2일 수준으로 단축할 수 있었다.

라. 채소작물데이터베이스의 pan-genome 측면에서의 확장 준비

○ 과거 작물의 표준유전체는 해당 작물에서 학술 및 시장에서의 중요성이 큰 순계화된 계통을 대상으로 구축이 되었다. 그러나 sequencing 비용의 지속적인 절감 및 거대유전체 재현을 위한 다양한 long read sequencing 기반의 유전체 재조립 기술이 공개되면서 각 작물 별로 다양한 아종의 표준유전체가 공개되기 시작하였다. 양배추의 경우 현재 농업적으로 주요하게 활용되고 있는 다양한 아종들에 대한 표준유전체가 조립되어 공개가 되었으며 일부는 현재 유전자 예측(gene annotation) 또한 수행되고 있다.

제4절 신규 분자마커 개발을 위한 유전연구

1. 전사체 정보를 활용하여 특정 표현형의 유전분석을 위한 Indel 마커 생산 방식의 개발

현재 작물 뿐만이 아닌 다양한 계(system)에 속하는 종(species)들을 대상으로 종이 나타내는 특정한 표현형의 기반에 있는 유전자의 발현 양상을 확인하기 위해 RNA-seq이 널리 사용되고 있다. 이는 조직 및 세포내에서 발현하는 mRNA를 특정 조건하에서 추출하고 이를 sequencing을 위한 library로 구성하여 다양한 기종의 sequencer를 통해 해당 mRNA의 서열을 읽어나가는 과정이다. 이러한 방식으로 구성된 염기서열은 de novo assembly를 통해 만들어진 transcript sequence 혹은 특정 종에 대해 이미 알려진 reference genome 상의 gene coding 영역에 alignment되어 유전자별 발현량을 산출하고 후속 분석을 진행하는데 주로 이용된다.

RNA-seq의 기술 자체는 특정 처리를 가한 샘플에서의 처리별 유전자 발현량의 산출뿐만이 아니라 RIL(Recombinant Inbred Line) 혹은 NIL과 같이 유전적 배경이 같지만 특정 형질에서 차이를 보이는 계통간의 유전자 발현 양상을 비교하기 위해 사용되는 경우가 있다. 이와 같이 서로 상당 부분의 유전적 배경(genetic background)을 공유하지만 표현형이 상이한 두 계통을 상정하였을 때 1차적으로 두 계통 간에 일부 유전자를 coding 하는 영역(exon)에 변이가 있어 서로 다른 표현형을 유도하는 것을 가정할 수 있다. 그렇다면 RNA-seq으로 얻어진 염기서열의 de novo assembly를 통해 각 계통에서 발현되는 transcript의 염기서열을 산출할 수 있으며 이를 서로 비교하게 되면 전사체로부터 유래한 sequence 데이터로도 두 계통을 구분할 수 있는 exon 상의 변이를 식별하고 이를 분자마커로 만들 수 있다는 방법을 도출할 수 있게 된다.

가. Raw sequence data의 전처리

본 작업에서 서로 구분되는 두 형질에 대해 구성된 read sequence는 de novo assembly 과정을 거쳐서 형질 별 mRNA transcript를 재현해야 하므로 read sequence 상에 세포내에서 발현하는 유전물질 유래의 sequence 외에 다른 요소를 허용할 수 없다. 따라서 sequence 라이브러리 구성에 사용되는 adapter 서열 및 vector 서열의 제거 및 sequence error rate가 큰 서열의 제거가 필수적이다. raw sequence 데이터의 quality 확인은 fastqc를 통해 수행하였으며 overrepresented sequence(contaminant sequence)의 제거 및 1차 phred score 30기준 sequence trimming은 trimmomatic을 통해 수행하였다. 이후 2차 quality check에서 read sequence의 좌우 말단에서 GC contents에 요동을 보이는 경우 해당 길이의 sequence를 seqtk를 통해 trimming하였다. 또한 작업의 최소 read 길이인 30bp를 만족시키지 못하는 paired end sequence를 식별하고 pair를 이루는 sequence만 남기기 위해 2차 trimmomatic을 sequence quality filtering 기준 없이 처리하여 pair를 이루는 read sequence만을 후속 작업인 de novo assembly에 이용하였다.

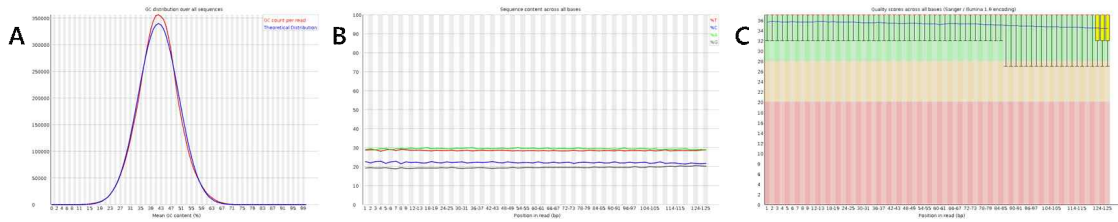


그림 1. QC 이후의 de novo assembly의 활용 가능 sequence quality의 예시. A) read sequence의 GC가 이론치와 유사, B) read sequence 전체의 GC가 안정적으로 일치, C) phred score가 낮은 read data의 제거

나. 표현형 별 read data의 통합 및 de novo assembly

전처리 과정을 거쳐서 확보한 신뢰성이 높은 read sequence를 그룹 별로 통합하고 이를 활용하여 de novo assembly를 수행한다. Read sequence의 de novo assembly는 Trinity를 통해 수행할 수 있으며 이는 reference 기반의 alignment를 통한 contig 구성이 아닌 read sequence내의 공통적인 kmer를 찾아 이를 기반으로 contig를 구성한다. 따라서 de novo assembly 결과를 최적화 시킬 수 있는 kmer의 탐색을 jellyfish라는 소프트웨어를 통해 수행한 후 de novo assembly를 진행하여 서로 다른 두 표현형을 보이는 그룹 별로 발현하는 transcript sequence의 구조를 재구성 할 수 있다.

다. Indel marker 생산 파이프라인의 구축

Read 데이터의 전처리 이후에 형질별로 통합되어 de novo assembly가 완료된 transcript sequence 중 하나를 기준으로 하여 1차적으로 표준 유전체의 coding sequence에 BLASTN을 수행한다. BLAST 수행 결과에서 de novo transcript가 표준 유전자의 coding sequence에 대해 80%이상의 서열 유사도를 보이는 것을 filtering 한 이후 transcript와 표준 유전자와의 관계에 대한 정보를 확인할 수 있다. 그리고 de novo assembly transcript의 방향을 표준 유전자의 방향(5' - 3')과 일치시키는 작업을 수행한다. 다음 단계로는 두 번째 표현형의 transcript를 첫 번째 transcript의 리스트에 BLASTN을 수행하여 첫 번째와 두 번째 transcript sequence간의 같은 유전자를 coding 하는 transcript들을 연관 짓고 이를 초기 단계에서 얻은 표준 유전자와의 관계와 합쳐 표준 유전자 - 첫 번째 transcript - 두 번째 transcript 간의 관계에 대한 데이터를 확보할 수 있다. 이후 표준 유전자를 coding하는 전체 서열을 표준 유전자의 표준 유전체 상의 위치정보를 참조하여 samtools의 명령어를 통해 생산하여 한 표준 유전자에 대한 세 염기서열을 clustalW를 통해 대조한다. 이를 통해 표준 유전자의 exon영역에서 두 표현형으로부터 유래한 transcript sequence에서 5bp 이상의 차이를 보이는 영역을 target으로 하여 해당 위치를 marking하고 이를 primer3를 통해 indel marker를 구성하게 된다.

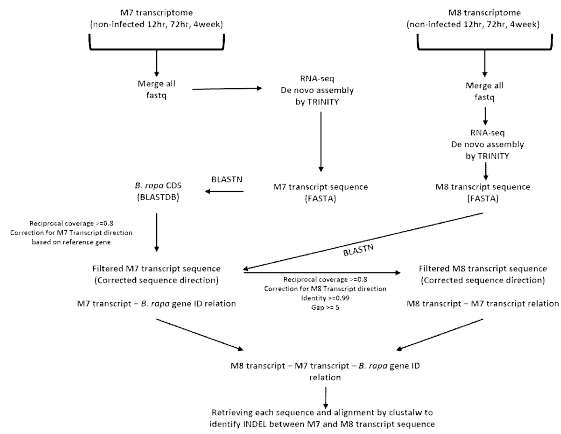


그림 2. 전사체 데이터 기반의 Indel marker 개발 플랫폼의 개요도 (배추 뿌리혹병에 대한 저항성 차이를 보이는 두 계통 M7과 M8의 RNA-seq 데이터로부터의 indel 마커 생산 사례)

2. GWAS(Genome-wide association study)의 관점에서 접근한 배추 노화 형질에 대한 유전분석
 배추의 GWAS(Genome-wide association study)실현을 위해서는 배추의 다양한 형질을 대상으로 조사한 표현형 데이터(Phenotype data)와 GWAS 집단을 이루는 각 계통별 유전체 상에 분포하는 변이 데이터(Genotype data)가 필요하다. 본 실험실에서 연구 중인 다양한 표현형들에 의해 대변되는 형질들 중에서 배추의 수확 후 저장과정에서의 노화 양상(Shelf-life)을 장기간 조사한 결과, GWAS 집단 내에서 수확 후 잎의 노화에서 현저한 차이를 보이는 두 계통(27142: 느린 노화, 27160: 빠른 노화)을 선별할 수 있었다 (그림 3).

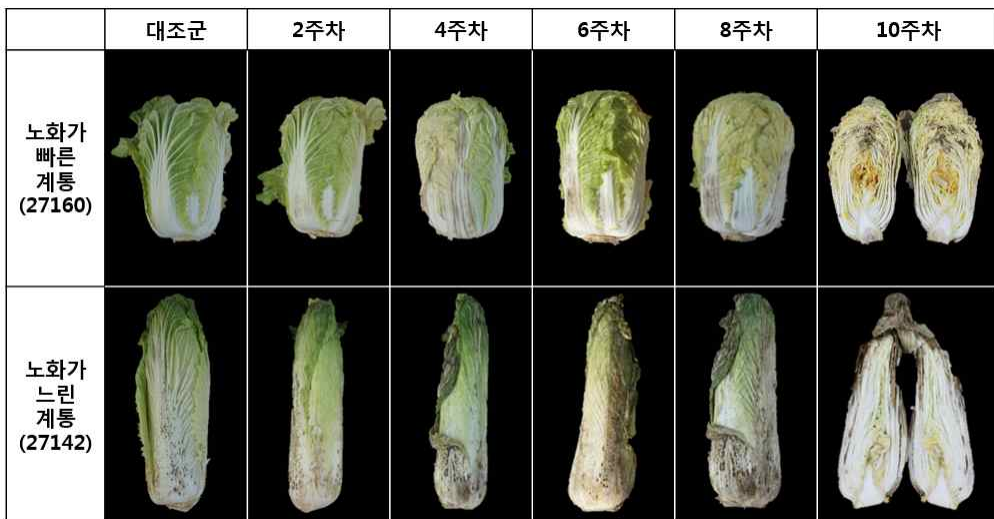


그림 3. GWAS 집단내에서 수확 후 저장이후의 노화 과정에서 현저한 차이를 보이는 두 계통 27142와 27160의 저장 기간에 따른 변화 양상

배추의 저장과정에서의 노화 양상을 지배하는 생리현상의 규명을 위해 수확한 배추 잎조직의 생리 실험을 진행함과 동시에 해당 생리현상의 기저에 깔려있는 유전적 요인에 대한 분석을 수행하였다.

가. GWAS 분석을 위한 input 데이터의 생산

연구대상인 형질에 연관(association)되어 있는 유전체 상의 영역을 찾기 위해 표현형 데이터와 유전자형 데이터를 요구하는 것은 GWAS 분석이 기존의 QTL 분석과 공유하는 요인이다. 그러나 QTL 분석이 유전적 영역의 표시를 F₁ 이후의 분리집단에서 준비된 마커들에 대한 재조합(recombination)에 의해 생산되는 연관지도로 표현하는 반면, GWAS 분석은 대개 자연집단을 이루는 개체들에 대해 수행되며 변이의 위치는 연관지도 상의 마커의 상대적 거리가 아닌 공개된 reference genome의 물리적 위치로 표현된다. 그 결과 GWAS 분석은 계통의 re-sequencing이 reference genome에 대해 충분한 coverage로 수행되었을 경우, 기존의 연관지도 보다 더 높은 resolution을 갖게 된다. 비록 GWAS 분석이 특정 계통만이 갖는 특이적 유전적 영역을 다룰 수 없다는 단점을 갖고 있지만 수집한 모든 계통이 reference genome에 대하여 공유하는 유전적 영역을 다루고 그 영역 상에서 나타나는 변이 데이터를 생산할 수 있다는 점은 큰 이점이다. 본 작업에서는 GWAS 분석에 필요한 기본 데이터를 생산하였다.

(1) 배추 156개 계통의 수확 후 노화 과정의 phenotyping data 생산

노화에 따른 상태 변화는 여러 양상을 보이고, 이에 따라 각기 다른 기준으로 평가가 가능하다. 또한 노화의 양상에 영향을 주는 요인은 기상 환경, 영양 상태, 생육 상태, 생육 기간 등 다양하다. 이에 따라 기존에 별도로 실행하였던 실험들을 기준으로 표현형 평가의 최적화를 이루고자 하였다. 이에 따라 본 연구실에서 보유한 배추 264 계통을 대상으로 ㉠ 과종 후 4주 개체의 ㉡ 4, 5번 잎을 채취하여 ㉢ 25°C ㉣ 암조건으로 ㉤ 80mm 페트리디쉬에 filter paper 1매와 잎을 넣어 ㉥ 3M Air-Pore tape으로 밀봉하고 ㉦ 저장 0일, 5일, 12일, 20일에 ㉧ 각각의 시료를 스캔하여 사진을 확보하였다 (그림 4). 배추 264개 계통 중 re-sequencing 정보가 있는 계통을 선별하였으며 표현형에 대한 index를 구성하고 잎의 황화 정도에 따라 황화 정도가 높을수록 높은 값의 indexing을 주었다. indexing은 최소 0점에서 최대 6점까지 분포하였으나 전체 156개 계통 중 극소수의 계통들만이 0에 가까운 표현형 값이 부여되었다 (그림 5). yellowing indexing의 히스토그램을 관찰한 결과, yellowing에 association된 allele이 조사 집단 내에서 rare variant로 취급되어야 하는 것을 알 수 있었다.

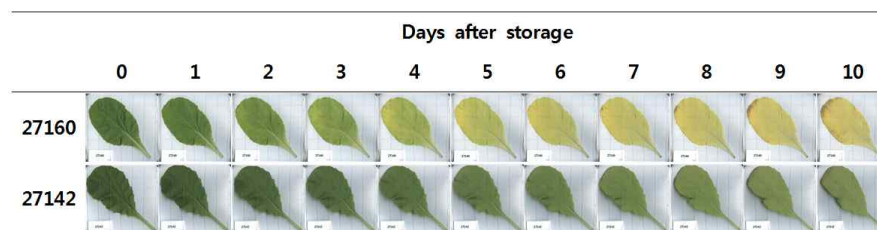


그림 4. 노화 elite line으로 선정된 27160과 27142의 저장 기간별로 획득한 저장 시료의 사진

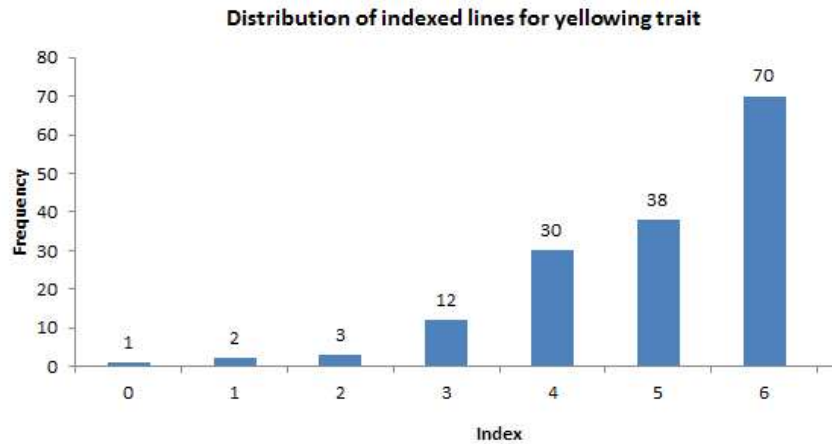


그림 5. 표현형 조사 대상 계통 배추 156개 line의 yellowing indexing distribution

(2) 배추 156개 계통의 수확 후 노화 과정의 genotyping data 생산

수확 후 잎의 노화에 대한 GWAS 분석을 위한 배추 genotyping의 절차는 본 보고서의 3절 1.의 개선된 계통별 SNP calling 방식을 적용하였다. 배추 표준유전체 2.0 version을 기준으로 삼아 156개 계통에 대하여 homozygous alternative allele만을 집단 내에서 추출하였을 때 총 5,541,255 개의 잠재적인 SNP position로 작업을 시작하였으며 이후 절차에 따른 filtering 방식을 적용하였다. 그러나 현저한 느린 노화를 보이는 계통의 수가 전체 집단 내에서 극소수 였기 때문에 MAF(minimum allele frequency) 0.05가 아닌 0.01까지 허용하여 filtering을 수행하였다. 최종적으로 1,107,951개의 SNP position을 산출하였고 이를 hapmap 포맷으로 전환하여 genotyping data로 활용할 수 있었다.

표 1. Filtering 과정에 따른 156개 배추 집단의 가용 SNP position의 수량 변화

| Filtering treatment steps | Number of filtered SNPs |
|--|-------------------------|
| Primary putative Snps | 5,541,255 |
| Bi-allelic only by vcftools | 5,361,078 |
| Missing rate under 10%, MAF threshold 0.01, hardy-weinberg thres hold 0.001 by PLINK | 1,148,339 |
| Chromosome variants only by sed command | 1,107,951 |

나. GAPIT을 통한 표현형과 유전자형과의 association의 분석

연구자에 의해 준비된 표현형 데이터와 유전자형 데이터의 연관 분석에는 주로 GAPIT과 Tassel5.0이 널리 쓰이나 R package 기반의 GAPIT이 100만 여개 이상의 변이데이터를 처리할 수 있기 때문에 GAPIT을 통한 분석을 채택하였다. 이 과정을 통해 노화의 진행 양상에 연관성을 가질 것이라 기대할 수 있는 배추 유전체 상의 SNP의 위치 정보를 생산할 수 있었다.

(1) Manhattan plot의 작성

Manhattan plot은 기본적으로 제공되는 GAPIT의 분석 결과 중 하나이며 유전체 전체 영역에서 산출된 SNP 발생 위치 각각에 대한 표현형에 대한 연관 여부를 log 수식으로 전환된 p-value에 대한 값으로 표현한 도식이다. 각 크로모좀 영역별로 peak를 보이는 영역이

나 기준선 위에 위치하는 점에 해당하는 SNP의 위치가 조사된 표현형과 강한 연관성을 갖고 있는 것으로 간주된다.

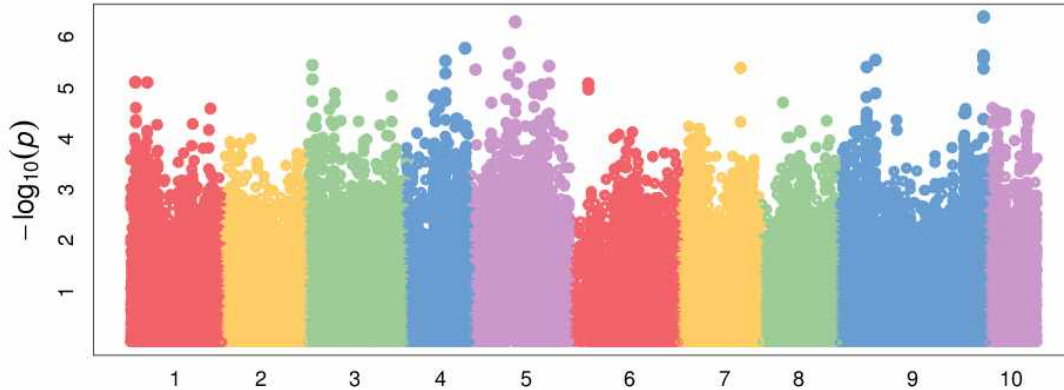


그림 6. 노화 표현형에 대한 조사 집단의 배추 156개 계통에 대한 manhattan plot

GAPIT의 자체적으로 보유한 처리 절차에 따르면 분석과정에 반영된 마커의 수에 따라 유의한 SNP로 간주할 수 있는 기준선을 제시한다. 본 작업에서 입력된 1,107,951개의 마커는 GAPIT의 bonfferoni 수식에 의해 manhattan plot에서 약 8 이상의 값을 갖는 SNP를 표현형과 유의한 연관성을 갖는 SNP로 주장할 수 있다. 그러나 값으로 표현한 도식이다. 각 크로모솜 영역별로 peak를 보이는 영역이나 기준선 위에 위치하는 점에 해당하는 SNP의 위치가 조사된 표현형과 강한 연관성을 갖고 있는 것으로 간주된다. 그러나 본 작업 결과에서는 1,3,4,5,9번 크로모솜에서 peak를 이루는 구간은 나타났으나 기준선 8을 넘는 SNP는 나타나지 않았다. 이는 GAPIT의 default로 설정된 통계분석 모델이 본 작업과 같은 rare variant를 다루는데 적합한지에 대한 검증의 필요성을 제기한다. 유체에 관련된 연구 결과에서는 통계모델의 변경하였을 때 같은 dataset에서 나타는 유의한 SNP의 수가 변화하는 것을 보인 사례가 존재한다. 추후 rare variant를 다루는데 적합한 모델을 찾아 GWAS 분석을 추가 수행할 필요가 있다고 사료된다.

(2) SNP position의 snpEff 결과를 활용한 associated SNP position의 annotation

GAPIT은 분석 이후 조사 표현형과 연관된 SNP의 position별 다양한 통계량을 제공하지만 표현형과 높은 연관성을 보이는 SNP position이 유전자를 이루는 구역에 위치하는지 혹은 유전자간 영역에 위치하는 등과 같은 정보를 제공하지 않는다. 표현형 연관 SNP의 위치정보를 조사 집단에 대하여 수행한 snpEff 결과의 SNP별 위치정보와 연동하여 GAPIT의 결과에 snpEff에 의한 SNP위치의 발생이 transcript의 전사에 미치는 영향 및 SNP와 유전자와의 위치관계 또한 식별할 수 있다. 이와 같은 전제를 기반으로 관련 pipeline을 구축하였으며 이를 이용하여 manhattan plot에서 나타난 상위 100개의 SNP 위치에 대한 annotation을 진행하였으며 표 9는 그 중 상위 20개에 대한 결과이다.

표 2. 노화 형질 associated SNP 상위 20개의 annotation 결과

| Rank | SNP position | Annotation | Associated gene ID | <i>A. thaliana</i> orthologue | Gene symbol |
|------|--------------|-------------------------|---------------------------|-------------------------------|--------------|
| 1 | A09:53269044 | intron_variant | BraA09006811 | AT2G13680.1 | <i>CALS5</i> |
| 2 | A05:16221208 | upstream_gene_variant | BraA05001490 | AT3G21610.1 | N/A |
| 3 | A04:21381458 | upstream_gene_variant | BraA04002954 | AT2G41560.1 | <i>ACA4</i> |
| 4 | A05:13916918 | downstream_gene_variant | BraA05001386 | AT4G28550.1 | N/A |
| 5 | A05:13916836 | missense_variant | BraA05001385 | AT3G20720.2 | N/A |
| 6 | A09:53269034 | intron_variant | BraA09006811 | AT2G13680.1 | <i>CALS5</i> |
| 7 | A09:53271202 | intron_variant | BraA09006811 | AT2G13680.1 | <i>CALS5</i> |
| 8 | A09:13954466 | upstream_gene_variant | BraA09002628 | N/A | N/A |
| 9 | A04:14229817 | intron_variant | BraA04001818 | AT2G24645.1 | N/A |
| 10 | A03:2190804 | upstream_gene_variant | BraA03000495 | N/A | N/A |
| 11 | A05:28502285 | upstream_gene_variant | BraA05003098 | N/A | N/A |
| 12 | A09:10724467 | upstream_gene_variant | BraA09002038 | N/A | N/A |
| 13 | A05:17597267 | intergenic_region | BraA05001633-BraA05001634 | N/A | N/A |
| 14 | A07:22267065 | downstream_gene_variant | BraA07003059 | AT1G70740.1 | N/A |
| 15 | A09:53269060 | intron_variant | BraA09006811 | AT2G13680.1 | <i>CALS5</i> |
| 16 | A05:1733386 | intergenic_region | BraA05000054-BraA05000055 | N/A | N/A |
| 17 | A04:14235476 | synonymous_variant | BraA04001818 | AT2G24645.1 | N/A |
| 18 | A05:13916834 | missense_variant | BraA05001385 | AT3G20720.2 | N/A |
| 19 | A03:2190800 | upstream_gene_variant | BraA03000495 | N/A | N/A |
| 20 | A01:2099779 | missense_variant | BraA01000413 | AT4G34390.1 | <i>XLG2</i> |

3. 작물 표현형 및 생리현상을 지배하는 후보 유전자 식별을 위한 전사체 기반의 PPI 네트워크 연구
 배추는 결구를 형성하는 작물로서, 내엽과 외엽에 여러 가지 차이가 있다 (for nutritional value, appearance, taste, color and texture). 눈으로 식별 가능한 이런 차이 외에 어떤 기능적 차이가 존재한다. 그동안 *B. rapa*의 leaf에 관한 연구는 특정 목적에 맞는 differentially expressed 유전자들의 집합을 찾고자 하는 연구는 많이 진행되어 왔다. 그러나 세포의 기능은 발현된 유전자 자체의 기능뿐만 아니라, 유전자들 사이에 존재하는 복잡한 조절 관계에 의해 결정되기 때문에, 두 조직의 차이에 대한 유전자 집합을 발굴하는 것보다, common or specific 하게 조절되는 핵심 네트워크를 발견하는 것이 더욱 중요하다.

본 작업에서는 RAN-seq를 통해 배추 표준 유전체에서 나타난 유전자별 발현량을 산출했다. RNA 추출을 위한 시료는 배추(*Brassica rapa* ssp. *pekinensis*) Chiifu 401-42계통을 생육시켜 조직별로 각각 채취하여 -70°C에 보관하여 이용하였다. 온도 21°C, 상대습도 70%, 장일(광 16시간, 암 8시간) 조건의 배양실에서 4주간 생육시킨 개체를 노지에 정식하고 10주간 생육시켜 결구된 개체에서 결구 외엽 및 내엽 시료를 선정하였다. 전체 결구 외엽 중 최대 크기의 잎인 5번째 잎을 결구 외엽으로 선정하였고, 결구 내에서 외엽과 확연히 다른 색을 보이는 4번째 잎을 결구 내엽으로 선정하였다. Illumina Hiseq 2000 플랫폼을 사용하여 *B. rapa*의 세 가지 단계 (seeding, rosette and inner leaves)를 나타내는 RNA 샘플을 준비하고 시퀀싱했다. Illumina 시퀀싱 실행 결과 총 142M 개의 read sequence가 포함되어 있으며 평균 길이는 100bp이다. 고품질 트리밍 이후 13,242 만개의 읽기 (84.60 %)가 데이터 세트에 남아있었다. 그런 다음 참조된 사본에 매핑 된 필터링 된 읽기가 추가되어 105,190,000 개의 매핑되었다 (표 3).

표 3. 시퀀싱 결과

| File name | Sample | Raw reads | Trimmed reads | Mapped reads | Unmapped reads | Mapping rate |
|------------------|--------------|--------------------|--------------------|--------------------|-------------------|----------------|
| TN1312R4974_1.fq | Old leaf | 22,928,264 | 21,328,689 | 15,994,012 | 5,334,677 | 74.99 % |
| TN1312R4974_2.fq | inner leaf | 22,928,264 | 21,328,689 | 16,064,209 | 5,264,480 | 75.32 % |
| TN1312R4977_1.fq | Old leaf | 23,995,399 | 22,243,406 | 17,964,791 | 4,278,615 | 80.76 % |
| TN1312R4977_2.fq | rosette leaf | 23,995,399 | 22,243,406 | 18,040,484 | 4,202,922 | 81.10 % |
| TN1312R4978_1.fq | Young leaf | 24,478,869 | 22,638,293 | 18,527,449 | 4,110,844 | 81.84 % |
| TN1312R4978_2.fq | Young leaf | 24,478,869 | 22,638,293 | 18,598,360 | 4,039,933 | 82.15 % |
| Total | | 142,805,064 | 132,420,776 | 105,189,305 | 27,231,471 | 79.36 % |

이를 통해 얻어진 조직 특이적 DEG를 찾았고, 잘 알려진 유전자네트워크를 통해 조직 특이적 protein-protein interaction 네트워크를 동정하였다 (그림 7). 어떤 유전자가 상대적으로 많은 인터랙션을 갖는지 찾기 위해, 구조적 분석을 수행했고, 노드별 degree를 살펴보았다. functional partners 가 많을수록 허브 노드일 확률이 크다. 이것은 네트워크 상에서 중요한 위치에 놓여 있음을 의미한다. inner leaf sub network에서는 Bra021202, Bra022179, Bra031134, Bra028865, Bra005755, Bra009569, Bra000150, Bra005012, Bra017051 이런 노드가 degree 80 이상을 갖고 있었다. 이중에서 The P5CS1(Bra000150, Bra005012, Bra017051) gene is conserved in human, chimpanzee, Rhesus monkey, dog, cow, mouse, rat, chicken, zebrafish, fruit fly, mosquito, C. elegans, rice, and frog. 많은 종에 보존되어 있는 유전자일수록 중요한 유전자임을 의미하기 때문에, degree가 높은 유전자들이 네트워크 상에서 중요한 위치에 있음을 확인할 수 있었다. 또한 outer leaf sub network에서는 Bra020569, Bra021202, Bra022179, Bra022988, Bra040227, Bra018167, Bra003055, Bra038281, Bra000150, Bra005012, Bra017051, Bra006947, Bra009742, Bra012402, Bra019161, Bra030621이 110 이상의 degree를 갖는 것을 확인했다. 이중에서 ORG1(structural molecule activity, protein kinase activity, kinase activity), MPK3(다양한 스트레스의 반응과 관련이 있음), P5CS1, BT3등이 포함되어 있었다. 결국 각 조직별 degree가 높은 유전자들이 매우 중요한 역할을 하는 것임을 확인할 수 있었다.

이 네트워크의 특성들을 알아보기 위해, 구조적, 기능적 특성을 비교분석했다. 그 결과, 안쪽 잎(inner leaf)에서는 이차대사물질의 생합성과 관련된 프로세스, 생식 및 개화 관련 프로세스를, 바깥쪽의 잎(outer leaf)에서는 빛과 관련된 개화자극 및 화기유도 (floral organ identity), 다양한 외부 자극에 대한 반응 프로세스가 관련되어 있음을 확인했다. 또한, 이들과 관련된 유전자의 네트워크를 구축함으로써 어떤 유전자가 잠재적 바이오마커인지를 찾았다. 결론적으로, 제안된 연구 방법을 통하여 배추 결구의 내·외엽에서 유전자 발현양상의 변화를 네트워크 측면에서 관측할 수 있었고, 이런 분석 방법이 서로 다른 생물학적 이벤트를 설명하는데 이용될 수 있음을 보였다.

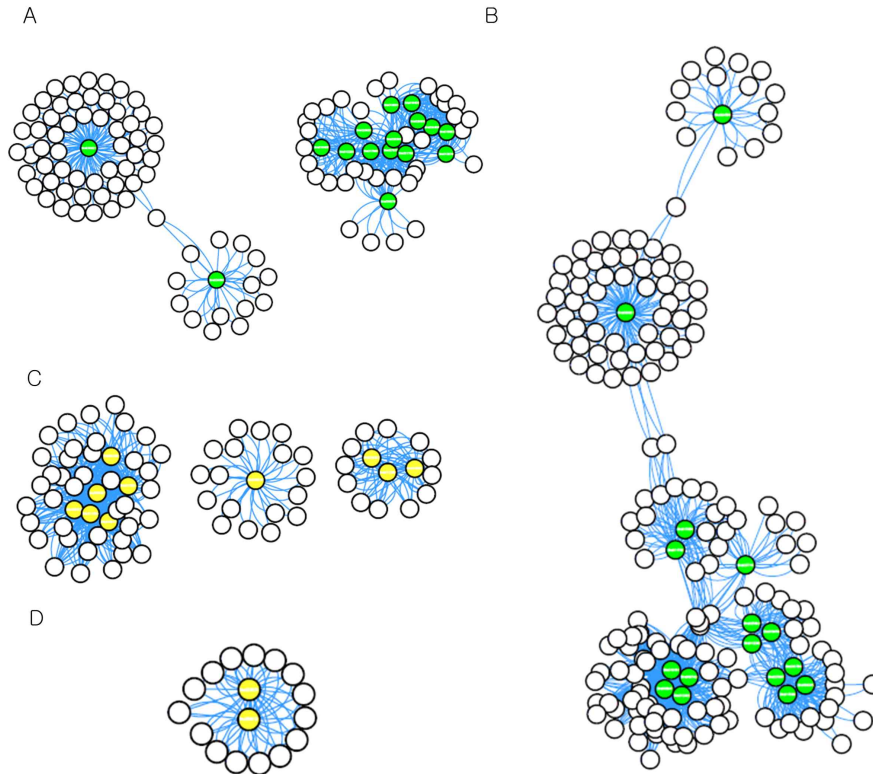


그림 7. 내외엽(ILN, RLN)의 대표적인 서브네트워크. (A) 항상성과 관련된 RLN의 서브 네트워크 (Bra036335, Bra018502, Bra017898, Bra027837, Bra031769, Bra013579, Bra018501, Bra019406, Bra029505, Bra034466, Bra038378, Bra030621, Bra012238, Bra016459, Bra025832, Bra024536). (B) 다양한 외부 자극에 대한 응답과 관련된 RLN의 서브 네트워크 (Bra006947, Bra009742, Bra012402, Bra027837, Bra031769, Bra011792 및 Bra017856) (C) 서브 네트워크 내인성 호르몬에 의한 발달 과정과 관련된 ILN (Bra000150, Bra005012, Bra017051, Bra003919, Bra010263, Bra011236, Bra024032, Bra000637, Bra010611, Bra011784). (D) 글루타민 생합성 과정과 관련된 ILN의 서브 네트워크 (Bra008612, Bra023573).

표 4. 내외엽 서브네트워크의 노드에 대한 GO term 분석

| Tissue | GO_ID | Descriptions | Gene Ratio | Bg Ratio | P-value |
|--------------|---|---|------------|-------------|-------------|
| Rosette leaf | GO:0000103 | sulfate assimilation | 10/192 | 25/30914 | 1.19E-14 |
| | GO:0019752 | carboxylic acid metabolic process | 8/192 | 14/30914 | 1.96E-13 |
| | GO:0019344 | cysteine biosynthetic process | 9/192 | 37/30914 | 2.68E-11 |
| | GO:0016126 | sterol biosynthetic process | 9/192 | 61/30914 | 2.39E-09 |
| | GO:0045454 | cell redox homeostasis | 11/192 | 241/30914 | 3.57E-06 |
| | GO:1902074 | response to salt | 4/192 | 20/30914 | 4.74E-05 |
| | GO:0009970 | cellular response to sulfate starvation | 4/192 | 21/30914 | 5.38E-05 |
| | GO:0042545 | cell wall modification | 7/192 | 123/30914 | 7.89E-05 |
| | GO:0045490 | pectin catabolic process | 7/192 | 159/30914 | 0.000343059 |
| | GO:0048232 | male gamete generation | 3/192 | 15/30914 | 0.000525767 |
| | GO:0006662 | glycerol ether metabolic process | 4/192 | 40/30914 | 0.000556336 |
| | GO:0006096 | glycolytic process | 6/192 | 124/30914 | 0.000598019 |
| | GO:0048443 | stamen development | 4/192 | 42/30914 | 0.000605592 |
| | GO:0010439 | regulation of glucosinolate biosynthetic process | 3/192 | 17/30914 | 0.000661808 |
| | GO:0042542 | process | 5/192 | 100/30914 | 0.001676895 |
| | GO:0006298 | response to hydrogen peroxide | 3/192 | 24/30914 | 0.001695803 |
| | GO:0010093 | mismatch repair | 3/192 | 24/30914 | 0.001695803 |
| | GO:0006995 | specification of floral organ identity | 3/192 | 28/30914 | 0.002636515 |
| | GO:0034599 | cellular response to nitrogen starvation | 4/192 | 69/30914 | 0.003335515 |
| | GO:0016036 | cellular response to oxidative stress | 5/192 | 123/30914 | 0.003617992 |
| GO:0010188 | cellular response to phosphate starvation | 2/192 | 12/30914 | 0.007923337 | |
| GO:0048497 | response to microbial phytotoxin | 2/192 | 18/30914 | 0.016404537 | |
| GO:0048441 | maintenance of floral organ identity | 2/192 | 23/30914 | 0.023760983 | |
| GO:0048507 | petal development | 2/192 | 25/30914 | 0.027529975 | |
| GO:0010218 | meristem development | 3/192 | 81/30914 | 0.036010043 | |
| | | response to far red light | | | |
| Inner leaf | GO:0019752 | carboxylic acid metabolic process | 8/103 | 14/30914 | 1.19E-15 |
| | GO:0035556 | intracellular signal transduction | 13/103 | 313/30914 | 3.80E-10 |
| | GO:0018105 | peptidyl-serine phosphorylation | 7/103 | 132/30914 | 1.81E-06 |
| | GO:0009835 | fruit ripening | 3/103 | 24/30914 | 0.000315415 |
| | GO:1900056 | negative regulation of leaf senescence | 3/103 | 28/30914 | 0.000484535 |
| | GO:0010540 | basipetal auxin transport | 3/103 | 32/30914 | 0.000624344 |
| | GO:0009739 | response to gibberellin | 5/103 | 175/30914 | 0.001074765 |
| | GO:0009693 | ethylene biosynthetic process | 3/103 | 40/30914 | 0.001098958 |
| | GO:0006542 | glutamine biosynthetic process | 2/103 | 11/30914 | 0.001682583 |
| | GO:0006970 | response to osmotic stress | 5/103 | 207/30914 | 0.00182307 |
| | GO:0050665 | hydrogen peroxide biosynthetic process | 2/103 | 19/30914 | 0.004639545 |
| | GO:0009611 | response to wounding | 6/103 | 399/30914 | 0.005141078 |
| | GO:0045893 | positive regulation of transcription,DNA-templated | 6/103 | 400/30914 | 0.005141078 |
| | GO:0010468 | transcription,DNA-templated | 4/103 | 171/30914 | 0.005834067 |
| | GO:0009854 | regulation of gene expression | 2/103 | 25/30914 | 0.0068626 |
| | GO:0010582 | oxidative photosynthetic carbon pathway | 2/103 | 35/30914 | 0.012294685 |
| | GO:0051260 | floral meristem determinacy | 2/103 | 35/30914 | 0.012294685 |
| | GO:0045892 | protein homooligomerization | 4/103 | 260/30914 | 0.022241248 |
| | GO:0048364 | negative regulation of transcription, | 4/103 | 296/30914 | 0.032593417 |
| | GO:0010228 | DNA-templated root development | 3/103 | 170/30914 | 0.035648215 |
| | | vegetative to reproductive phase transition of meristem | | | |

4. 배추 뿌리혹병(Clubroot disease) 형질의 RNA-seq 데이터를 활용한 InDel marker의 개발

2010년 이후로 NGS(Next Generation Sequencing)의 보편화 및 비용의 절감으로 인해 현재 다양한 작물을 대상으로 내병성과 같은 유용 표현형을 조절하는 유전자의 발현 양상의 확인을 위해 RNA-seq이 널리 사용되고 있다. 이를 통해 특정 형질 한 가지에 차이를 보이는 두 근연 계통의 RNA 발현 양상을 확인할 수 있다. 뿐만 아니라 RNA-seq을 통해 얻은 fastq file을 가공하여 각 계통에서 발현하는 transcript의 염기서열을 재구성하고 이를 서로 비교할 수 있다. RNA-seq의 기술 자체는 특정 처리를 가한 샘플에서의 처리별 유전자 발현량의 산출뿐만이 아니라 RIL(Recombinant Inbred Line) 혹은 NIL과 같이 유전적 배경이 같지만 특정 형질에서 차이를 보이는 계통간의 유전자 발현 양상을 비교하기 위해 사용되는 경우가 있다. 이와 같이 서로 상당 부분의 유전적 배경(genetic background)을 공유하지만 표현형이 상이한 두 계통을 상정하였을 때 1차적으로 두 계통간에 일부 유전자를 coding 하는 영역(exon)에 변이가 있어 서로 다른 표현형을 유도하는 것을 가정할 수 있다. 따라서 de novo assembly를 통해 만들어진 두 계통의 transcript sequence를 연구 대상인 중의 reference genome 상에 예측된 유전자와 대조하여 상이한 표현형을 보이는 두 근연 계통의 transcript 서열에서 서로 차이를 보이는 영역을 찾을 수 있으며 이를 coding 영역 상의 변이를 식별할 수 있는 마커로 활용할 수 있다. 본 작업에서는 제 3절 1에서 소개한 자체적으로 구축한 파이프라인을 통해 뿌리혹병에 저항성에 차이를 보이는 두 계통(M8, M7)의 transcriptome을 대상으로 de novo assembly를 수행하였다. 각 계통별로 발현하는 transcript의 서열을 재구성하고 이를 reference genome 상에 예측된 유전자와 alignment 후 두 계통에서 5bp 이상의 염기서열 차이를 보이는 영역을 찾고, primer3 소프트웨어를 활용하여 해당 영역에 대한 프라이머를 제작하였다 (표 5).

표 5. 배추 뿌리혹병의 transcriptome 기반으로 산출된 indel 마커의 기본 정보

| Marker ID | Forward primer | Reverse primer | Product size(bp) | Target gene ID |
|------------------------------|-----------------------|-----------------------|------------------|----------------|
| M7_sequence_44717_in_2_DEL8 | GTGCTTTCTCTGTTCGTCGG | GTTTCCCCACGACTTTCAT | 218 | Bra009753 |
| M7_sequence_45488_in_2_INS6 | TTGATGTTTTACGGGCTTCC | TTTGACGACCGTCTTTTGC | 160 | Bra002255 |
| M7_sequence_49137_in_1_INS6 | ATCAACCACCTCCTGGCTCT | CACCATCACCTTCTTGCAT | 176 | Bra005404 |
| M7_sequence_49743_in_2_INS3 | TATGGGAAGCCAGTTGCTGA | CTGACTTGATTTTCGCCTTCG | 118 | Bra002464 |
| M7_sequence_51428_in_2_INS5 | GAACCGCATTTAAGGCCATC | TAGTGGCAATCCCTCACAA | 184 | Bra006122 |
| M7_sequence_51502_in_2_DEL10 | GGAGAGAGTGAGCGAGAGCA | ACGGGGATGATAAGCTGGAC | 117 | Bra011941 |
| M7_sequence_51889_in_2_INS8 | ATGAGAGACACCTGAGCGA | AGTAACATGCAACCAAGCGG | 102 | Bra020603 |
| M7_sequence_51896_in_2_INS8 | GTCTGCGTTTGTGTTGAGGAC | TGACGCAAAAACCTGTTGTA | 169 | Bra020603 |
| M7_sequence_51928_in_2_DEL15 | AGCCACTTCGTGTGTTGCTC | GTCCTTGTCCCAGCCAGATT | 134 | Bra032281 |
| M7_sequence_52517_in_2_DEL3 | TGAGGTTTGAGATGGGAGGA | GTCATCTGCTGCCTTGGAAA | 115 | Bra028676 |
| M7_sequence_55863_in_2_INS6 | GACCTGGTGAAGCCTTTA | CATCGGTTGAGTACCATCA | 176 | Bra030388 |
| M7_sequence_56526_in_1_DEL3 | GCTGTGAAGGTTGCTGATCC | TTGTGCTTACTTCCGGTTG | 186 | Bra003493 |
| M7_sequence_57356_in_1_DEL3 | AGAGGGTGAGGCCAGTGTCT | GACATTGATTCGGCTGGCTA | 122 | Bra038025 |
| M7_sequence_57406_in_3_DEL6 | CCTGGAAGAGAAAACGGAGG | TTCATCTCAATGGCTCCACA | 161 | Bra011773 |
| M7_sequence_58286_in_2_DEL5 | CGCCATCGTCAACAAGAACT | ATACCCGAGCAAAGCCACAT | 103 | Bra008967 |
| M7_sequence_58397_in_1_DEL15 | TCGGACATAGTAGCGGGATG | CAAAGCAGGCATGAATCTCC | 153 | Bra012672 |
| M7_sequence_58873_in_1_INS3 | CGGAGTCACGAAGATAGCCA | GAAATCCATCACTGACCCCA | 175 | Bra001129 |
| M7_sequence_59551_in_1_INS3 | AAACCGTCGCAAGCTACTCA | CACAGGTTTCGGCACATTTT | 129 | Bra028373 |
| M7_sequence_60008_in_2_DEL13 | ACGGTGATGGTCACTATGGG | TTGGTTTTGTGGCCGTTAAG | 187 | Bra016584 |
| M7_sequence_60038_in_2_DEL7 | GCAGACACCTCCCAAGAACA | TCCAACGTGAATAGATCCCA | 102 | Bra012617 |
| M7_sequence_60038_in_3_INS5 | CCTGAGAGTGCCAAAGGGAT | GAACGTATGCCACAGACCCA | 140 | Bra012617 |
| M7_sequence_60147_in_1_DEL6 | CGTGGGATTTTCTCTCGTGT | TGCTGAAATTTGTTGGGTGTG | 123 | Bra035212 |
| M7_sequence_60682_in_3_DEL10 | GCATCTCCAACACCGATCAC | CCACCATATCACCAGAACCG | 251 | Bra037237 |
| M7_sequence_61825_in_1_DEL23 | CGAGCAAAACCTAGCTTCA | GCTTGGCGTCTGACCTCTT | 140 | Bra030754 |
| M7_sequence_68988_in_2_INS12 | CCGGAACCACTTTGAGAAGA | GAGCTTGGTCCGTTGATACA | 211 | Bra028702 |
| M7_sequence_69797_in_8_INS12 | TGTCACAGATTGGTGAAGG | GACGCTCTCTCTCGTTTCCC | 126 | Bra011912 |
| M7_sequence_7191_in_1_DEL6 | ACGATCCTACCCCTCCTCCT | TGTTACGGTTAGCCAATGCG | 117 | Bra021840 |
| M7_sequence_69836_in_1_INS6 | CGTGGATCTCCTCCTCTCT | GCTTCAACATCATTGCCCTC | 178 | Bra024005 |

| | | | | |
|-------------------------------|------------------------|--------------------------|-----|-----------|
| M7_sequence_69925_in_1_INS12 | GTTGTGTGTTTCAAATGGCTGA | GACAGGTTGGATGGCGTTTA | 179 | Bra004576 |
| M7_sequence_69931_in_1_INS12 | GGATGTCACCTCCTCATCTTCA | GACAGGTTGGATGGCGTTTA | 146 | Bra004576 |
| M7_sequence_70008_in_2_INS3 | TGCGTCAGTCACAGGAATTG | TCCAAGGTAAGCTCATAACGCA | 237 | Bra009342 |
| M7_sequence_70066_in_1_DEL3 | GATCAAGTGGAAACACAGGGG | CAGGTTTCATGGGTTTCTTCTG | 162 | Bra006635 |
| M7_sequence_70325_in_4_DEL3 | TCTCAATCGAAAGGGGATCA | TCGAATCAGCTACGCTCCTC | 176 | Bra032908 |
| M7_sequence_8707_in_1_DEL9 | TCTGAAAAGCATAGCCGTGG | TGTATGGGAGTGGAAAGGGG | 246 | Bra037145 |
| M7_sequence_71421_in_3_DEL5 | CCCATTAACATGGCTCTTCG | CATAGCATCATAGGCCCGGT | 101 | Bra013533 |
| M7_sequence_72396_in_1_DEL6 | GTGCTTGCTGAGTCTGGAG | CCTCCTTACAATCTTCTTCCCC | 155 | Bra021733 |
| M7_sequence_73096_in_1_INS3 | GATGACCTCGCCGATAACAA | ACTCGGTTTCGATAGCTTCA | 193 | Bra021148 |
| M7_sequence_73649_in_3_DEL13 | AGACGGTGTGAGGTGCAAG | TTTTGACAGCGAAGGAGGAA | 171 | Bra000679 |
| M7_sequence_75223_in_2_INS6 | TTCAGAGATCCCTCCTCGT | TTATCCTCCGACGAACCCCTT | 153 | Bra010390 |
| M7_sequence_76005_in_1_INS3 | CAGTAACAACGGAGGGAGCA | CCAATGGCTCTTTAACCCTG | 176 | Bra013255 |
| M7_sequence_76223_in_2_DEL18 | GATGGTGGAGATGGGTGTGT | TGGCTGGCTCTTCTCTCTG | 151 | Bra022384 |
| M7_sequence_78702_in_1_INS3 | AACTTTGCGGAGCTTGTA | AAAGCGTTCACCGTTTGGTT | 156 | Bra019432 |
| M7_sequence_79658_in_1_DEL6 | TGCTAACGTGGTGGTGAAGTC | AACGACCTTTGAGCCGTACA | 190 | Bra004498 |
| M7_sequence_10413_in_1_DEL3 | TAACATCCTCCTCGGACGC | TCGCATTTGGAAGTTGGTCT | 159 | Bra012141 |
| M7_sequence_80742_in_2_DEL9 | TTCAGATGCAGGAAGCCCTA | TACTCCTCCTGAACGCCAAC | 172 | Bra038285 |
| M7_sequence_82857_in_3_DEL5 | TGCTCTGTTTCATATCTTGGC | TACTCAGTCCAAGCCAGAGA | 120 | Bra011168 |
| M7_sequence_83406_in_1_DEL27 | GGATTCCACGAATTCAGTTCT | TGTTGTGACTTGTGAGGCAA | 180 | Bra040549 |
| M7_sequence_83406_in_2_INS10 | GGAGTTGCATAATTTCTGGGG | AGGTGAAGAAGCACACATTGG | 186 | Bra040549 |
| M7_sequence_83817_in_1_INS6 | TCAGATTGATGTCAGCGTGC | CCAACGGATTCACCATCATC | 181 | Bra036718 |
| M7_sequence_83824_in_2_INS14 | TGTTTCCCAATTCATCCGAG | AGGCTCCCAAGCAAGAAA | 197 | Bra003929 |
| M7_sequence_84590_in_3_INS10 | TGCAGAAGGGAATGAAGAGG | TTACTCAGAAGCAGCGAGGC | 146 | Bra030581 |
| M7_sequence_86041_in_4_DEL4 | CGGGAAGAAGACCCGAGAAAC | GCAAACAGAAAATGAACTTGCC | 235 | Bra016536 |
| M7_sequence_86418_in_5_INS6 | GCAGGGGTGTTGATGTCAC | AGTAGGCTCAGGCTTGGGTG | 218 | Bra032913 |
| M7_sequence_87265_in_7_INS7 | GTGCAAGGGAGGCTTTGTTT | AGCAAACCTTAAATGCTCCACCA | 168 | Bra027158 |
| M7_sequence_87494_in_1_INS12 | CATCCTCCTCAACAACCTCGC | AACTTGATCCACAGCCCTCCC | 165 | Bra038956 |
| M7_sequence_90126_in_1_DEL3 | TTGCTGGATAGGAAACGTGG | TTGCCTTTGTCCTTCCTCTG | 175 | Bra004946 |
| M7_sequence_2756_in_1_INS6 | GTGGAAGCTGAGATGAAGCG | TCCTCCCCAGAGATATTCAGC | 143 | Bra032393 |
| M7_sequence_91177_in_1_DEL5 | TCTGCGATTGCTTCTTCGAT | GTACTTGGCCCCACCATTTT | 115 | Bra014964 |
| M7_sequence_91331_in_2_INS6 | GTTGTTGTGCTCAGATCGGG | CTCCAAAATGCTCTCGGGTT | 218 | Bra009444 |
| M7_sequence_91489_in_1_INS15 | CGGCAATGTGCTAATGTCTT | TGCATAACTTTCCCTTAGCTCC | 248 | Bra008816 |
| M7_sequence_93461_in_2_DEL9 | TCAGCAAACTACTGGAGCAA | CAGCAGGAAGAGCAGTGGAT | 200 | Bra006677 |
| M7_sequence_94122_in_2_INS3 | TTCAGGAAGTGGCGAAGAAG | GCCTCATCATCAGAAACCC | 222 | Bra016505 |
| M7_sequence_96454_in_3_INS9 | CCAAGGAGATTAGGCGCTTC | CATGGCCCTTTGTTGTTTAC | 134 | Bra012753 |
| M7_sequence_96950_in_2_INS5 | ACTCGAATCACCTCTCCCT | CTTCCACGGCTTGTCTCTG | 227 | Bra016358 |
| M7_sequence_11862_in_2_INS12 | TAGAAATGAGCGAGCTTCGG | GGGCTTGTTCCTGCTCTTCT | 105 | Bra022749 |
| M7_sequence_98562_in_1_INS6 | GAAGGAAAAGCAACCAAGG | GAACCAGTTCACCGTCTTA | 172 | Bra000549 |
| M7_sequence_98562_in_2_DEL3 | CTTTGCTGGTATCATGCGCT | CGGTTCTTCCATCTCTTCCC | 135 | Bra000549 |
| M7_sequence_99275_in_2_DEL16 | TCAAAACCTCAAGCTCGCAT | GGCTGCCTCGTCTCTTAT | 113 | Bra002286 |
| M7_sequence_100065_in_2_DEL7 | GAAGAACTTGTGGGTGTGCC | TCCAAAGCGTGTAACAATCTCA | 154 | Bra007342 |
| M7_sequence_100068_in_2_DEL7 | GATCTTCAAAGGCTGCACA | TCCAAAGCGTGTAACAATCTCA | 193 | Bra007342 |
| M7_sequence_101025_in_2_DEL9 | AGGTGATGGTGGATTACAG | CATTGTTTCATCAGACCAGC | 111 | Bra025123 |
| M7_sequence_12161_in_3_INS16 | GAAGCTGGAACTTGGCTCA | CGATCTGCAACTTCGCATAAA | 148 | Bra006922 |
| M7_sequence_101309_in_2_DEL5 | TTGTTGCTCTTGGCTCAGT | CCACTTTGACATCTTTCCGTG | 188 | Bra005794 |
| M7_sequence_101309_in_3_DEL3 | AGGCTTGACTTGATTTGTCG | TGATGTGTCACAATTTCCCTG | 103 | Bra005794 |
| M7_sequence_110222_in_1_DEL6 | CGTGCTTCAAGGATTCACAC | ATGCTGCTCTTTCGGTGGTG | 109 | Bra024277 |
| M7_sequence_110422_in_2_DEL3 | AGACGTTTCGTCGCAGTTA | AGACAGAGGAGGAGGAGGCTT | 140 | Bra039436 |
| M7_sequence_110998_in_2_INS3 | TGCTCCATTGTTGATTGCTG | CTGAAAACGCCATGACAAAT | 160 | Bra001163 |
| M7_sequence_111086_in_3_INS12 | CCACTCTCCGAGTACCACCA | TGCAGAGGCAGAAGCAGAAT | 176 | Bra015875 |
| M7_sequence_111086_in_4_INS4 | TTCTCTCCGCTTCTGGAGT | ACAGCAACAGCCAAGCTACC | 233 | Bra015875 |
| M7_sequence_12877_in_2_DEL7 | TCCCTCATCCGTCATTCAAC | TCACTCGCTTACCCCTCAA | 150 | Bra037502 |
| M7_sequence_111423_in_2_DEL3 | AGGCAATGGCTCTTCTGTG | CTGCCAAAACCTCCATTCTCA | 189 | Bra039439 |
| M7_sequence_111423_in_3_DEL8 | CGCATGGTTCTTCTTCCAA | CAGAGAACTAACCGAATCATAAGT | 107 | Bra039439 |
| M7_sequence_111907_in_2_INS9 | CAATCCAGTCTTCAACCGA | GAAGATTTCCCTCCGCAATA | 137 | Bra040446 |
| M7_sequence_111916_in_1_DEL5 | CTCAAAGCAGGGTGGGATTT | GAGGCTGGAAATAAGGGTGC | 200 | Bra040446 |
| M7_sequence_112164_in_1_INS3 | TTGGTGAAAATCGTGAAGGATG | ATTTCGGCTAGTTCTCTCCA | 139 | Bra020309 |
| M7_sequence_114519_in_2_INS8 | GTTCCGACAGTGTGTTGGGA | GATCGCCATGTATGAATCCG | 191 | Bra004135 |
| M7_sequence_116274_in_1_INS3 | GGCAGCAAAAAGATCAAACCA | TGATGATCCCGCTAGAAGA | 101 | Bra000304 |
| M7_sequence_13763_in_2_DEL15 | AAGGAACCCAAACAGCATT | GCGAAGATTGTTGTTGCTGA | 158 | Bra007061 |
| M7_sequence_2869_in_1_INS6 | ACGGAGAAGAGGTACGGCTC | CACGTATGACCAAAATGCC | 201 | Bra019593 |
| M7_sequence_14819_in_1_DEL7 | AGGAAGACTCTGGTCAGGGC | GGCTAATCGTCTGTACGCCA | 120 | Bra003663 |
| M7_sequence_16420_in_1_DEL12 | GGAGCATAACCTCCTCGTCC | GGGGAGAAAGGTGACTCGAT | 114 | Bra040667 |
| M7_sequence_16500_in_2_DEL3 | GCGAGTCGAGCAAAAGCATC | CAGCTCTGATTGGATGGGA | 180 | Bra018735 |

| | | | | |
|------------------------------|-----------------------|-------------------------|-----|-----------|
| M7_sequence_16501_in_2_INS6 | AATGAAGAGGAGGCAGCACA | TTTGGGCAGCAAGAGTGGT | 133 | Bra018735 |
| M7_sequence_22990_in_1_DEL3 | CAGGAGCTGAGGAGGTGTG | CCCTCCCTGAAACTGAGTCC | 188 | Bra029250 |
| M7_sequence_24528_in_4_INS12 | TAAGGACTGCCCTCTATGCG | CATGGAACATCAGCGTAAATC | 184 | Bra003180 |
| M7_sequence_25445_in_1_DEL15 | CACCTGATGGAGACGATGT | TGTCTTGAGTTCAGCTTATTGGC | 119 | Bra015141 |
| M7_sequence_25512_in_3_INS8 | CAAACCCCTGGCTCAGAAGA | TGAATGCAACCATCCTCACA | 117 | Bra026721 |
| M7_sequence_26948_in_13_INS3 | CTTTTGCTCGGCATGAAGAA | TGACCTCTTGACAACCTGCCC | 100 | Bra038282 |
| M7_sequence_28969_in_1_INS3 | ATGGCTGTGATTGGTATCGC | CACCACCAACTGAGCAAAT | 160 | Bra008386 |
| M7_sequence_29054_in_1_DEL6 | GCAAACAGGCGAATGAAGAG | CAAAACAACCTTCTCCAGCG | 185 | Bra014790 |
| M7_sequence_29785_in_3_DEL3 | AGGCTTCGAGGAGTTCATGC | CTGGATGATTCCGCTCTCCT | 123 | Bra009452 |
| M7_sequence_29785_in_4_INS9 | CGTTGCCCTCGGAGAATAATG | TCCATCCATCCACGAACAAT | 186 | Bra009452 |
| M7_sequence_29785_in_5_INS9 | CCTGATCTTGAGACGGTGA | GCCACTCCCTCTACGGTTA | 142 | Bra009452 |
| M7_sequence_30116_in_23_DEL3 | CAGTGATGTGCAGTTGTGGG | GAATCGCAAGGGTTGAAGTG | 243 | Bra010248 |
| M7_sequence_33700_in_2_DEL9 | TGTTCTGAATACCAAGCGG | AATCGCAATATCGCTGAAG | 152 | Bra033646 |
| M7_sequence_33882_in_22_INS6 | TGCAAGTTTGACCATCTCC | AGTCGTGTCAGTCGCTCCAC | 102 | Bra034744 |
| M7_sequence_35393_in_1_DEL3 | GAAAATGAAACTCACGGCCA | CGGAATCAGAGAGACCAACG | 173 | Bra003229 |
| M7_sequence_40787_in_1_INS17 | TTGCCTTGTTATGGAGGC | GGTTGTCCCACTGTCCAAAA | 235 | Bra029989 |
| M7_sequence_43418_in_1_INS9 | CCTCCAACGACGAATCTCAA | AGGGAAGAAAACAGACCCGA | 197 | Bra031783 |
| M7_sequence_43897_in_1_INS8 | ACAAGACTGGCTTCCGATGA | TGTAATTTGGCGGAAACTG | 173 | Bra006057 |
| M7_sequence_43899_in_1_DEL3 | GAAGCCAAATCAATGAGGA | GCCACGTTTTCAGCAAGAGA | 177 | Bra006057 |

5. 배추의 결구 내의 위치에 따른 DEG 분석 및 조직 특이적, 공통적 서브 네트워크 재구성
 배추는 결구 내의 위치에 따라서 영양가, 맛, 색상 및 질감이 다르다. 예를 들어 내부의 잎은 작고 노란색을 띠고, 부드러우며 파삭파삭한 식감이 특징인 것에 반해 결구 겉의 잎은 초록색에 거친 질감과 쓴맛이 느껴진다. 우리는 후술할 일련의 연구를 통해 이런 위치에 따른 조직별 차이는 기능적 차이로부터 기인한다는 결과를 얻었다. 또한 각 조직의 공통적인 혹은 특이적인 유전자, 서브네트워크를 동정했다.

가. 배추 조직 공통적, 특이적 유전자 동정

여러 조직 간의 DEG(differentially expressed gene) 발굴 연구는 참조 조직(reference tissue)을 어떻게 선택하는가에 따라 그 결과가 다양하였다. 가령, Bra038535 유전자의 발현량은 5개 조직에서 매우 저조하나, 어린잎을 대조군으로 하면 꽃 조직에서 매우 유의한 수치로 DEG로 간주된다. 반대로, Bra002316은 5개 조직에서 발현이 가장 높으나 어린잎을 대조군으로 하면 꽃 조직에서 DEG로 선별되지 못했다. 이렇게 조직 특이적으로 매우 주요한 유전자가 대조군에 따라서 DEG로 선별되지 못하는 경우가 발생한다. 따라서 본 연구에서 조직 특이적 유전자를 찾는 방법으로 두 조직 간의 상대적 발현량 차이를 분석하는 기존 연구 방법과는 달리, RNA-seq¹⁾결과의 절대적 발현량을 기반으로 다음과 같이 선별하였다.

(1) 절대 발현량 기반의 유전자 동정

우선, 모든 조직에 공통적으로 발현되는 유전자(tissue commonly expressed gene)을 찾기 위하여, 우리는 모든 조직에 높은 발현량을 가지며 동시에 조직별 차이(variance)가 적은 값일수록 공통된 고유의 속성을 유지할 것으로 판단하였고, 이를 기준으로 동정하였다. 그

1) 샘플로부터 추출한 RNA를 Illumina TruSeq RNA Sample Preparation Kit v2 (Illumina, San Diego, CA)을 통해 cDNA library를 구성하였다. KAPA library quantification kit (Kapa biosystems KK4854)로 library quantification을 수행하였고, Hiseq2000 (Illumina, San Diego, CA)을 통해 RNA-seq 수행 (142.81 million total raw reads)했다. SolexaQA를 통해 염기서열 정보 필터링했고, Bowtie를 활용하여 reference assembly를 완료했다.

결과 발현량을 log2로 변환한 값이 8 이상이며, variance가 0.5일때 functional trait(BP: Biological Process, MP: Molecular Function, CC: Cellular Component)가 가장 잘 유지되는 것을 확인했다(표 6). 이들 값의 기준을 높이면 MP나 CC가 거의 나타나지 않았다. 결과적으로, 우리는 8,581개의 공통적으로 발현되는 유전자를 찾았다. 이들에서 single-organism process에 관련된 유전자들이 유의하게 많음이 확인되었다(GO:0044699, GO:0044763). 또한, 이를 기반으로 Tissue specific expressed gene에서도 발현 임계값 8을 기준으로 선별한 결과, 조직별로 각각 1268(flower), 1233(root), 314(inner leaf), 453(outer leaf), 522(young leaf)개의 유전자를 동정하였다.

(2) 동정된 유전자의 특성 평가

본 연구에서는 유전적 특성(genetic properties)을 알아보기 위하여, BRAD에서 제공하는 genetic marker를 기반으로 중복 여부를 확인했다. 그 결과 root 조직 특이적 유전자들 중에서 글루코시놀레이트(glucosinolate) 관련 21개, 안토시아닌(anthocyanin) 관련 4개, 옥신(auxin) 관련 14개 유전자를 확인했고(표 7), outer leaf 특이적 유전자들에서 resistance와 관련된 5개 유전자(Bra014241, Bra013134, Bra037139, Bra019273, Bra027779)를 확인했다. 또한 각 조직 특이적 유전자들이 10개의 배추 염색체(chromosome)에 다양하게 분포되어 있음을 확인했다(표 8).

표 6. GO Term 분석 결과

| GO term | | description | FDR |
|------------|----|---|----------|
| GO:0015979 | BP | photosynthesis | 5.51E-23 |
| GO:0046686 | BP | response to cadmium ion | 1.49E-15 |
| GO:0009735 | BP | response to cytokinin | 1.96E-13 |
| GO:0009409 | BP | response to cold | 3.52E-13 |
| GO:0006096 | BP | glycolytic process | 5.17E-10 |
| GO:0009416 | BP | response to light stimulus | 2.86E-09 |
| GO:0042742 | BP | defense response to bacterium | 4.51E-08 |
| GO:0009658 | BP | chloroplast organization | 1.11E-07 |
| GO:0009651 | BP | response to salt stress | 5.09E-07 |
| GO:0006412 | BP | translation | 1.21E-05 |
| GO:0009624 | BP | response to nematode | 5.05E-05 |
| GO:0006357 | BP | regulation of transcription from RNA polymerase II promoter | 8.72E-05 |
| GO:0009408 | BP | response to heat | 0.000137 |
| GO:0009845 | BP | seed germination | 0.000154 |
| GO:0009414 | BP | response to water deprivation | 0.000188 |
| GO:0016567 | BP | protein ubiquitination | 0.000177 |
| GO:0030154 | BP | cell differentiation | 0.000498 |
| GO:0040008 | BP | regulation of growth | 0.001264 |
| GO:0007623 | BP | circadian rhythm | 0.001457 |
| GO:0006633 | BP | fatty acid biosynthetic process | 0.001642 |
| GO:0009611 | BP | response to wounding | 0.003182 |
| GO:0009617 | BP | response to bacterium | 0.003527 |

| | | | |
|------------|----|--|----------|
| GO:0009737 | BP | response to abscisic acid | 0.00343 |
| GO:0005507 | MF | copper ion binding | 3.85E-08 |
| GO:0003729 | MF | mRNA binding | 1.28E-07 |
| GO:0016887 | MF | ATPase activity | 5.72E-07 |
| GO:0003924 | MF | GTPase activity | 3.71E-06 |
| GO:0000981 | MF | RNA polymerase II transcription factor activity, sequence-specific DNA binding | 3.83E-05 |
| GO:0001135 | MF | transcription factor activity, RNA polymerase II transcription factor recruiting | 4.26E-05 |
| GO:0003735 | MF | structural constituent of ribosome | 3.06E-05 |
| GO:0004674 | MF | protein serine/threonine kinase activity | 0.000285 |
| GO:0051082 | MF | unfolded protein binding | 0.00059 |
| GO:0042802 | MF | identical protein binding | 0.003527 |
| GO:0004842 | MF | ubiquitin-protein transferase activity | 0.005193 |

표 7. root 조직의 genetic marker 중복 여부

| Reference of gene information | Glucosinolate related group | Anthocyanin related group | Auxin related group |
|--|-----------------------------|---------------------------|---------------------|
| <i>Brassica rapa</i> geneID (From reference ver 1.2) | Bra009245 | Bra001820 | Bra015374 |
| | Bra005949 | Bra029211 | Bra000160 |
| | Bra013000 | Bra029212 | Bra032876 |
| | Bra035954 | Bra036040 | Bra019060 |
| | Bra015939 | | Bra036997 |
| | Bra029434 | | Bra026365 |
| | Bra000760 | | Bra030074 |
| | Bra018524 | | Bra040812 |
| | Bra021947 | | Bra010338 |
| | Bra030246 | | Bra015694 |
| | Bra032010 | | Bra004515 |
| | Bra015935 | | Bra003044 |
| | Bra008132 | | Bra039264 |
| | Bra015938 | | Bra009856 |
| | Bra015936 | | |
| | Bra003817 | | |
| | Bra027623 | | |
| | Bra025668 | | |
| | Bra011822 | | |
| | Bra010645 | | |
| Bra017872 | | | |

표 8. 조직 특이적 유전자들의 배추 염색체(chromosome)분포

| Tissue | A01 | A02 | A03 | A04 | A05 | A06 | A07 | A08 | A09 | A10 | Total | Average | S.D |
|------------|-----|-----|------|-----|-----|-----|-----|-----|------|-----|-------|---------|--------|
| Old leaf | 36 | 44 | 67 | 32 | 38 | 43 | 51 | 25 | 77 | 27 | 440 | 44 | 16.87 |
| Inner leaf | 22 | 20 | 38 | 23 | 30 | 35 | 31 | 31 | 46 | 30 | 306 | 30.6 | 7.86 |
| Young leaf | 46 | 47 | 79 | 32 | 61 | 57 | 61 | 29 | 63 | 35 | 510 | 51 | 16.01 |
| Flower | 129 | 124 | 189 | 94 | 113 | 98 | 134 | 97 | 179 | 82 | 1239 | 123.9 | 35.85 |
| Root | 121 | 116 | 167 | 88 | 105 | 116 | 135 | 94 | 180 | 78 | 1200 | 120 | 32.92 |
| Common | 798 | 775 | 1239 | 567 | 825 | 887 | 837 | 639 | 1210 | 593 | 8370 | 837 | 231.03 |

나. 배추 조직 공통적, 특이적 부분 네트워크 동정

본 작업은 PPIN(Protein - protein interaction network)와 유전체 발현 데이터를 통합하여, 배추의 조직 특이적 부분 네트워크(c)를 발굴하는 과정으로 이는 통합된 네트워크(a)와 유전자 발현 데이터(b)를 활용해 재구성할 수 있다. 그림 1에 작업에 대한 절차를 정리하였다.

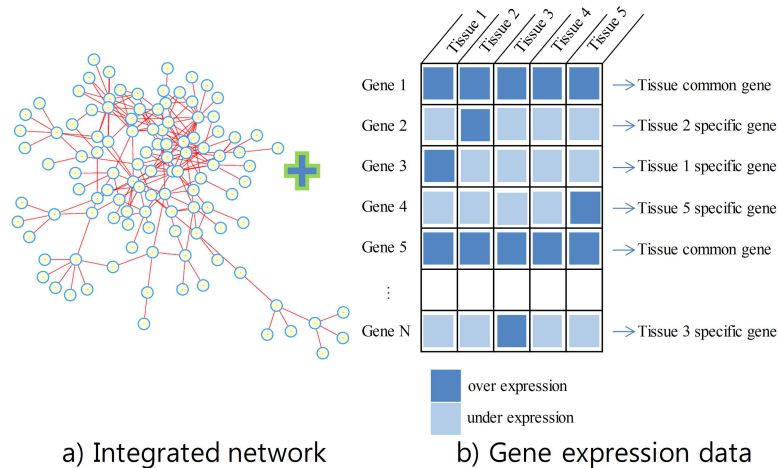


그림 8. 부분 네트워크의 재구성 방법

이 유전자를 기반으로 STRING²⁾ 데이터베이스를 참조하여 네트워크를 동정했다. 네트워크의 사이즈는 링크의 신뢰도(confidence) 값에 따라서 low(0.15), medium(0.4), high(0.7), highest(0.9)로 나뉘는데, 1에 가까울수록 매우 신뢰한다는 의미를 갖는다. 우리는 이전 연구에서 사용된 기준에 따라 0.6으로 산정하였다. 그 결과 5개의 조직에서 공통적으로 발현되는 서브 네트워크는 280,766개의 링크로 재구성되었고, 각 조직별 서브네트워크는 4128, 2290, 276, 380, 354개의 링크로 재구성되었다.

그림 8는 배추의 다섯 개 조직의 특이적 부분 네트워크를 보여준다. 이 그림은 Cytoscape 3.3.0에서 작성되었다. 여기에서, 연결정도(degree)는 해당 노드에 직접 연결되어 있는 노드들의 개수 (또는 연결선의 개수)를 말한다. degree가 가장 높은 4개의 유전자의 링크는 빨간색 라인으로 표시되었다. degree가 매우 높은 노드는 허브 노드라 할 수 있는데 이 유전자를 기반으로 다양한 기능이 여러 링크로 연결되어 있을 것으로 해석할 수 있다. 4개의 유전자들은 LRR(Leucine rich repeat)-RLK와 관련된 것임을 확인했다(PFAM ID: PF08263, PF13855). 이들은 뿌리 분열 조직 내 대부분의 세포에서 발견된다고 잘 알려져 있다.

2) 단백질 단백질 상호 작용 네트워크를 얻기 위해, 우리는 STRING 10.5 데이터베이스를 이용했다. 모든 링크는 0에서 1까지의 점수를 가지며, 1에 가까운 링크일수록 높은 신뢰도를 갖는 링크로 간주된다.

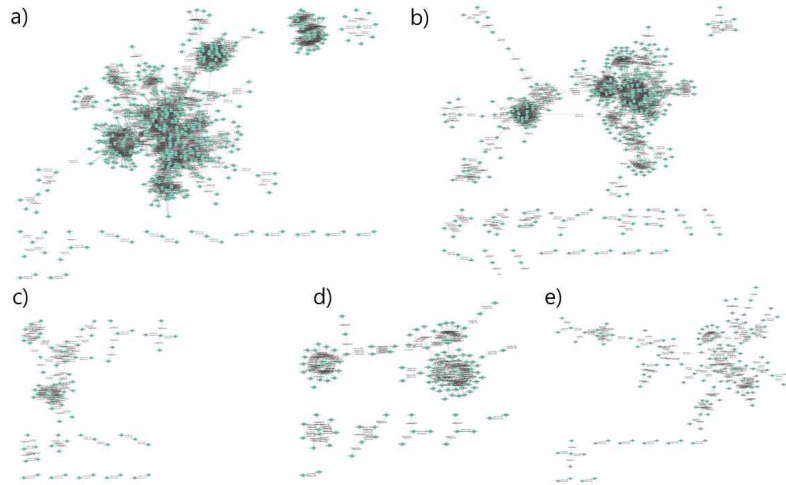


그림 9. 배추의 다섯 개 조직의 특이적 부분 네트워크

다. 서브 네트워크의 기능적 특성 분석

확인된 네트워크의 기능적 특성을 분석하기 위하여 GO분석³⁾을 수행한 결과를 [표 5]에 정리하였다. 결과적으로, 우리가 찾은 조직 특이적으로 발현된 유전자(tissue specific expressed gene)들이 조직의 특성을 잘 대변하는 것으로 확인되었다. 꽃(flower) 조직 특이적 네트워크는 화분 및 생식관련 기능, 뿌리(root) 조직 특이적 네트워크는 세포벽 관련된 기능과 관련이 있음을 확인했다. 또한 내엽(inner leaf) 조직 특이적 네트워크는 식물 대사 관련 기능, 외엽(outer leaf) 조직 특이적 네트워크는 외부 환경 및 그에 대한 방어(defense)에 관련된 기능, 어린 잎(young leaf) 조직 특이적 네트워크는 엽록체(chloroplast)와 관련이 있음을 확인했다.

3) GO 주석(GO Annotation)은 유전자 또는 단백질의 특성을 나타내는데 사용된다(<http://geneontology.org/>). GO 분석은 소규모 및 대규모 분석 모두를 위한 유용한 도구이다. GO의 기능성(functional) 주석은 GO 데이터베이스에서 얻었다. P값은 hypergeometric test를 통해 획득하였다.

표 9. 조직 특이적인 서브 네트워크의 유전자 기능 분석

| Tissue | GO ID | Description | Genes | FDR |
|----------------------|-----------------------------|--|--------|----------|
| 꽃 flower | 0022414 | reproductive process | 121 | 1.29E-11 |
| | 0000003 | reproduction | 121 | 1.64E-11 |
| | 0048437 | floral organ development | 31 | 5.83E-10 |
| | 0048438 | floral whorl development | 24 | 1.52E-7 |
| | 0048444 | pollen tube development | 22 | 1.11E-6 |
| | 0048868 | pollen tube growth | 17 | 5.82E-6 |
| 뿌리 root | 0044036 | cell wall macromolecule metabolic process | 31 | 1.69E-12 |
| | 0071554 | cell wall organization or biogenesis | 74 | 2.26E-12 |
| | 0042546 | cell wall biogenesis | 34 | 4.74E-12 |
| | 0071555 | cell wall organization | 49 | 8.90E-12 |
| | 0009733 | carbohydrate derivative transport | 9 | 0.0012 |
| | 0071365 | response to auxin | 32 | 0.0013 |
| | 0009734 | nucleotide-sugar transport | 6 | 0.0024 |
| 1901264 | nitrogen compound transport | 25 | 0.0097 | |
| 내엽 inner leaf | 0090304 | nucleic acid metabolic process | 72 | 1.32E-5 |
| | 0046483 | heterocycle metabolic process | 79 | 5.42E-5 |
| | 0006725 | cellular aromatic compound metabolic process | 79 | 1.90E-4 |
| | 1901360 | organic cyclic compound metabolic process | 79 | 3.95E-4 |
| | 0034641 | cellular nitrogen compound metabolic process | 92 | 4.81E-4 |
| 외엽 outer leaf | 1901700 | response to oxygen-containing compound | 64 | 5.87E-12 |
| | 0010200 | response to chitin | 18 | 7.28E-11 |
| | 0010033 | response to organic substance | 71 | 6.28E-10 |
| | 0050896 | response to stimulus | 157 | 9.30E-10 |
| | 0009719 | response to endogenous stimulus | 64 | 1.04E-9 |
| | 0010243 | response to organonitrogen compound | 19 | 1.28E-9 |
| 어린잎 young leaf | 0051188 | cofactor biosynthetic process | 12 | 0.0023 |
| | 0009411 | response to UV | 8 | 0.0039 |
| | 0044281 | small molecule metabolic process | 49 | 0.0042 |
| | 0010224 | response to UV-B | 6 | 0.0063 |
| | 0009658 | chloroplast organization | 16 | 0.0097 |
| | 0052696 | flavonoid glucuronidation | 7 | 0.0130 |

라. 서브 네트워크의 위상적 특성 분석

발굴된 유전자들의 위상적 특성을 분석하기 위하여 네트워크 다양한 지표 (average distance, diameter, degree, density, clustering coefficient)를 이용할 수 있다. 우리는 네트워크 분석 도구인 Cytoscape에 내재되어 있는 NetworkAnalyzer를 이용하여 생물학적 네트워크의 분석(<http://med.bioinf.mpi-inf.mpg.de/netanalyzer/help/2.7/>)을 수행하고, 네트워크의 직경, 이웃의 평균 개수 및 노드의 연결된 쌍의 수, 평균 클러스터링 계수, 위상 계수, 및 최단 경로의 길이와 같은 다양한 네트워크 파라미터의 분포를 계산하였다. 우리는 개별 유전자의 네트워크에 의한 특성을 계산하였고, 이에 대한 공통 유전자와 조직 특이적 유전자 그룹간의 차이를 비교하였다. 공통 유전자는 조직 특이적 유전자에 비해 네트워크의 중심부에 위치하고 있다는 사실을 확인하였고, 동시에 다른 유전자들과 더 많은 연결 관계를 맺고 있다는 사실을 확인하였다 (그림 10, 11, 12).

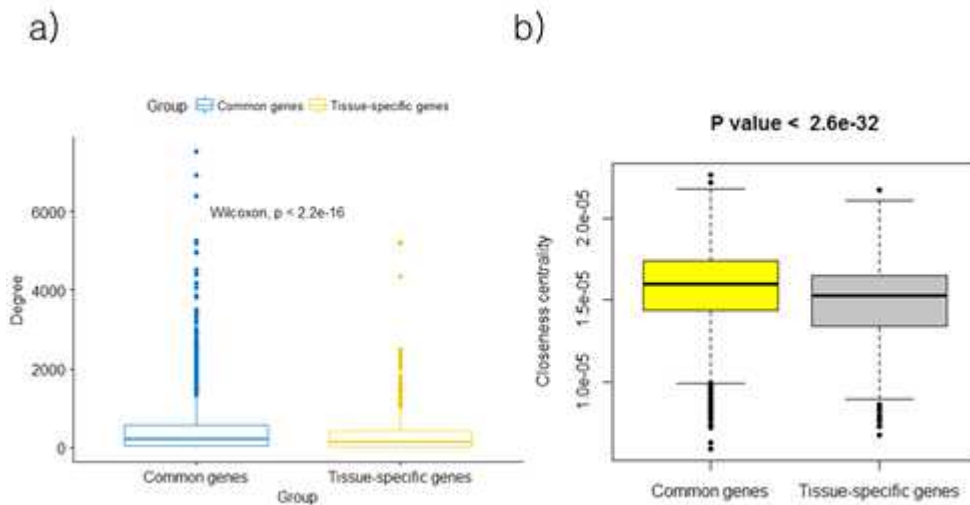


그림 10. 부분 네트워크의 위상적 특성 분석 결과 (1)

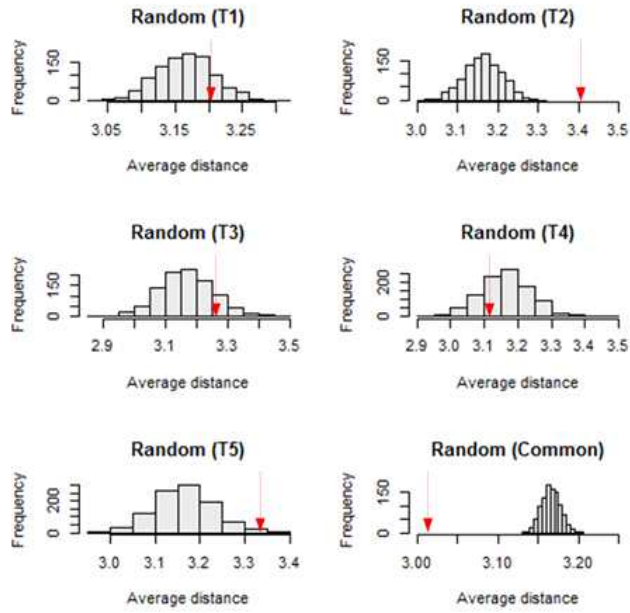


그림 11. 부분 네트워크의 위상적 특성 분석 결과 (2)

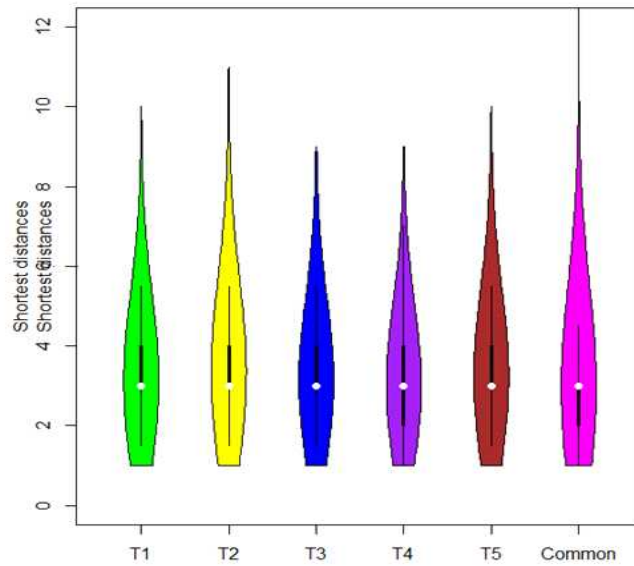


그림 12. 5개 조직 특이적, 공통적 네트워크의 shortest distance에 대한 결과

본 연구에서는 조직 특이적, 공통적 서브네트워크의 평균 거리 분포를 정리하였다. 여기에서 T1(꽃), T2(뿌리), T3(내엽), T4(외엽), T5(어린잎)은 각 조직별 특성을, Common은 공통적 조직의 특성을 나타낸다. T1에 해당되는 유전자 군이 100개라면 100개 유전자 간 평균 네트워크 거리는 3.22(빨간 선)이고, 임의의 100개 유전자 간 평균 네트워크 거리를 1000번 재추출하여 계산한(막대그래프) 분포도이다. 그 결과 공통적 노드들 사이의 거리가 임의의 두 노드 사이의 거리보다 매우 가깝다는 것을 확인했다. 이것은 노드 간 연결성이 높다는 것을 의미하고, 또한 서로 긴밀하게 연결되어 상호작용(interaction)이 많다는 것을 의미한다. 또한 공통 네트워크는 노드들 간 매우 치밀하게 연결되어 있다는 것을 뜻한다. 그림 5는 조직 특이적 네트워크에서 유전자들이 얼마나 가까운지 characteristic path length 값으로 분포를 확인하기 위한 것이다. 가운데 하얀점은 메디안 수치를 의미하며, 조직 특이적 5개 결과와 common의 결과가 상이함을 쉽게 알 수 있다. 또 다른 지표로 평균 이웃의 수(ave. number of neighbors)를 보면 8.5(flower), 6.6(root), 3.6(inner), 3.2(outer), 2.7(young tissue)로서 네트워크의 크기가 클수록 더 많은 이웃의 수가 있음을 확인했다.

6. 배추 유용형질 관련 데이터의 내재된 특성을 분석

○ 모든 조직에 공통적으로 발현되는 유전자(tissue commonly expressed gene)를 찾기 위하여, 우리는 모든 조직에 높은 발현량을 가지면서 동시에 조직별 차이(variance)가 적은 값일수록 공통된 특성을 유지할 것으로 판단하였고, 이를 기준으로 동정하였다. 유전자 발현량을 log2로 변환한 값이 8 이상이며, 조직별 발현량 편차가 0.5 이하를 갖는 유전자 그룹에서 functional trait (BP: Biological Process, MP: Molecular Function, CC: Cellular Component)가 가장 잘 유지되는 것을 확인했다. 발현량 편차 기준을 0.5 이상으로 높이면 획득된 유전자 군에 유의한 MP나 CC가 거의 나타나지 않았다. 조직 특이적 유전자들은 특정 조직에서 발현량 log2값이 8 이상이며 다른 조직에서는 그 이하인 유전자들로 정의하였다. 발견된 공통 및 조직 특이적 유전자를 기반으로 STRING 데이터베이스를 참조하여 조직-특이적 네트워크를 동정했다. 네트워크의 사이즈는 링크의 신뢰도(confidence) 값에 따라서 low(0.15), medium(0.4), high(0.7), highest(0.9)로 나뉘는데, 1에 가까울수록 매우 신뢰한다는 의미를 갖는다. 우리는 이전 연구에서 적용된 기준에 따라 0.6으로 산정하였다.

○ 네트워크 분석 도구인 Cytoscape에 내재되어 있는 NetworkAnalyzer를 이용하여 5개의 조직 특이적 네트워크 및 조직 공통 네트워크의 위상적 특성을 분석하였다. 여러 위상적 특성 중, 우리는 개별 단백질과 상호작용하는 이웃의 수 (degree), 개별 노드가 전체 네트워크 중심에 위치하는 정도 (closeness centrality) 그리고 네트워크를 구성하는 단백질쌍 간의 평균 거리 (average distance between nodes)를 계산하였다. 아래의 그림 1은 부분 네트워크를 재구성 하는 방법에 대한 설명이다(그림13).

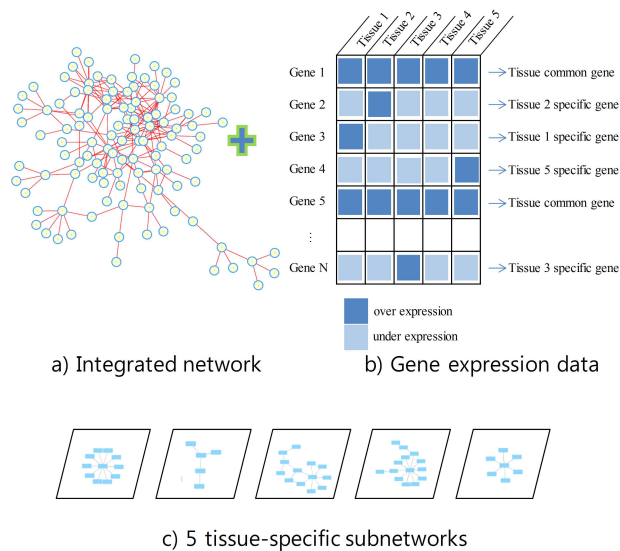


그림 13 부분 네트워크의 재구성 방법

7. 신호전달경로/유전자 네트워크를 효율적으로 동정할 수 있는 기법을 개발

○ 배추 조직 공통적, 특이적 유전자 동정

기존 연구에서 여러 조직 간의 DEG (differentially expressed gene) 발굴 연구는 참조 조직 (reference tissue)을 어떻게 선택하는가에 따라 그 결과가 다양하였다. 가령, Bra038535 유전자의 발현량은 5개 조직에서 매우 저조하나, 어린잎을 대조군으로 하면 꽃 조직에서 매우 유의한 수치로 DEG로 간주된다. 반대로, Bra002316은 5개 조직에서 발현이 가장 높으나 어린잎을 대조군으로 하면 꽃 조직에서 DEG로 선별되지 못했다. 이렇게 조직 특이적으로 매우 주요한 유전자가 대조군에 따라서 DEG로 선별되지 못하는 경우가 발생한다.

○ 본 연구에서 조직 특이적 유전자를 찾는 방법으로 두 조직 간의 상대적 발현량 차이를 분석하는 기존 연구 방법과는 달리, RNA-seq의 절대적 발현량과 조직별 발현량 편차를 기준으로 선별하였다 (방법2.3). 결과적으로, 우리는 8,581개의 공통적으로 발현되는 유전자를 찾았다. 이들에서 single-organism process에 관련된 유전자들이 유의하게 많음이 확인되었다(GO:0044699, GO:0044763). 또한, 동일한 기준을 통해 조직별로 각각 1268(flower), 1233(root), 314(inner leaf), 453(outer leaf), 522(young leaf)개의 조직 특이적 유전자를 동정하였다.

○ 우리는 발굴된 조직 특이적 유전자들의 유전적 특성(genetic properties)을 알아보기 위하여, BRAD에서 제공하는 genetic marker를 기반으로 중복 여부를 확인했다. 그 결과 root 조직 특이적 유전자들 중에서 글루코시놀레이트(glucosinolate) 관련 21개, 안토시아닌 (anthocyanin) 관련 4개, 옥신(auxin) 관련 14개 유전자를 확인했고, outer leaf 특이적 유전자들에서 resistance와 관련된 5개 유전자를 확인했다. 또한 각 조직 특이적 유전자들이 10개의 배추 염색체(chromosome)에 다양하게 분포되어 있음을 확인했다.

○ 배추 조직 공통적, 특이적 부분 네트워크 동정

우리는 PPIN에서 가장 잘 알려진 STRING 데이터베이스와 앞서 발굴한 유전자 리스트를 이용하여 공통 및 조직 특이적 부분 네트워크를 동정하였다 (방법 2.3) 다양한 조직에서 공통적으로 발현되는 부분 네트워크는 280,766개의 링크로 재구성되었고, 각 조직별 부분 네트워크는 4128, 2290, 276, 380, 354개의 링크로 재구성되었다 (그림 12).

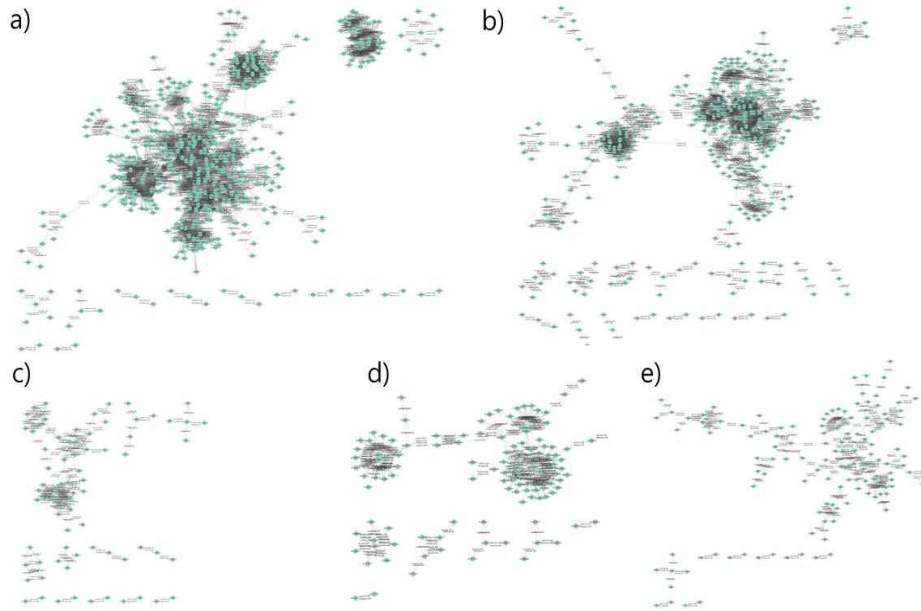


그림 14. 배추의 다섯 개 조직의 특이적 부분 네트워크

○ 그림 14는 조직 특이적으로 발굴된 다섯 가지 부분 네트워크를 보여준다(그림 20. a)꽃 b)뿌리, c) 내엽, d)외엽, e)어린 잎 조직). 이 그림은 Cytoscape 3.3.0에서 작성되었고, default layout으로, style은 simple 1로 그려졌다. 여기에서, 연결정도(degree)는 해당 노드에 직접 연결되어 있는 노드들의 개수 (또는 연결선의 개수)를 말한다. 이 결과 중 root 네트워크에서 degree가 가장 높은 4개의 유전자들은 LRR(Leucine rich repeat)-RLK와 관련된 것임을 확인했다(PFAM ID: PF08263, PF13855). 이들은 뿌리 분열 조직 내 대부분의 세포에서 발견된다고 잘 알려져 있다.

○ 부분 네트워크의 기능적 특성 분석

확인된 네트워크의 기능적 특성을 분석하기 위하여 GO분석을 수행한 결과를 [표 9]에 정리하였다. 결과적으로, 우리가 찾은 조직 특이적으로 발현된 유전자(tissue specific expressed gene)들이 조직의 특성을 잘 대변하는 것으로 확인되었다. 꽃(flower) 조직 특이적 네트워크는 화분 및 생식관련 기능, 뿌리(root) 조직 특이적 네트워크는 세포벽 관련된 기능과 관련이 있음을 확인했다. 또한 내엽(inner leaf) 조직 특이적 네트워크는 식물 대사 관련 기능, 외엽(outer leaf) 조직 특이적 네트워크는 외부 환경 및 그에 대한 방어(defense)에 관련된 기능, 어린 잎(young leaf) 조직 특이적 네트워크는 엽록체(chloroplast)와 관련이 있음을 확인했다.

표 10. 조직 특이적인 부분 네트워크의 유전자 기능 분석

| Tissue | GO ID | Description | Genes | FDR |
|----------------------|--|--|---------|----------|
| 꽃 flower | 0022414 | reproductive process | 121 | 1.29E-11 |
| | 0000003 | reproduction | 121 | 1.64E-11 |
| | 0048437 | floral organ development | 31 | 5.83E-10 |
| | 0048438 | floral whorl development | 24 | 1.52E-7 |
| | 0048444 | pollen tube development | 22 | 1.11E-6 |
| | 0048868 | pollen tube growth | 17 | 5.82E-6 |
| 뿌리 root | 0044036 | cell wall macromolecule metabolic process | 31 | 1.69E-12 |
| | 0071554 | cell wall organization or biogenesis | 74 | 2.26E-12 |
| | 0042546 | cell wall biogenesis | 34 | 4.74E-12 |
| | 0071555 | cell wall organization | 49 | 8.90E-12 |
| | 0009733 | carbohydrate derivative transport | 9 | 0.0012 |
| | 0071365 | response to auxin | 32 | 0.0013 |
| | 0009734 | nucleotide-sugar transport | 6 | 0.0024 |
| 1901264 | nitrogen compound transport | 25 | 0.0097 | |
| 내엽 inner leaf | 0090304 | nucleic acid metabolic process | 72 | 1.32E-5 |
| | 0046483 | heterocycle metabolic process | 79 | 5.42E-5 |
| | 0006725 | cellular aromatic compound metabolic process | 79 | 1.90E-4 |
| | 1901360 | organic cyclic compound metabolic process | 79 | 3.95E-4 |
| 0034641 | cellular nitrogen compound metabolic process | 92 | 4.81E-4 | |
| 외엽 outer leaf | 1901700 | response to oxygen-containing compound | 64 | 5.87E-12 |
| | 0010200 | response to chitin | 18 | 7.28E-11 |
| | 0010033 | response to organic substance | 71 | 6.28E-10 |
| | 0050896 | response to stimulus | 157 | 9.30E-10 |
| | 0009719 | response to endogenous stimulus | 64 | 1.04E-9 |
| 0010243 | response to organonitrogen compound | 19 | 1.28E-9 | |
| 어린잎 young leaf | 0051188 | cofactor biosynthetic process | 12 | 0.0023 |
| | 0009411 | response to UV | 8 | 0.0039 |
| | 0044281 | small molecule metabolic process | 49 | 0.0042 |
| | 0010224 | response to UV-B | 6 | 0.0063 |
| | 0009658 | chloroplast organization | 16 | 0.0097 |
| 0052696 | flavonoid glucuronidation | 7 | 0.0130 | |

○ 조직 공통 및 조직 특이적 네트워크의 위상적 특성 분석

네트워크 위상적 특성을 분석하면 네트워크를 구성하는 단백질들의 전반적 연결 관계를 한 눈에 파악할 수 있다. 우리는 조직 공통 및 조직 특이적 네트워크의 위상적 특성을 계산하였고, 이를 조직 공통 네트워크와 조직 특이적 유전자 그룹간의 차이를 비교하였다. 공통 유전자는 조직 특이적 유전자에 비해 네트워크의 중심부에 위치하고 있다는 사실을 확인하였고 (그림 13a), 동시에 다른 유전자들과 더 많은 연결 관계를 맺고 있다는 사실 ($P=2.2E-16$)을 확인하였다 (그림 20b).

○ 본 연구에서는 조직 특이적, 공통적 부분 네트워크의 평균 거리 분포를 [그림 15c]에 정리하였다. 여기에서 T1(꽃), T2(뿌리), T3(내엽), T4(외엽), T5(어린잎)은 각 조직별 특성을, Common은 공통적 조직의 특성을 나타낸다. [그림 15c]에서 T1에 해당되는 유전자 군이 100개라면 100개 유전자 간 평균 네트워크 거리는 3.22(빨간 선)이고, 임의의 100개 유전자 간 평균 네트워크 거리를 1000번 재추출하여 계산한(막대그래프) 분포도이다. 그 결과 공통적 노드들 사이의 거리가 임의 두 노드 사이의 거리보다 매우 가깝다는 것을 확인했다. 이것은 노드 간 연결성이 높다는 것을 의미하고, 또한 서로 긴밀하게 연결되어 상호작용(interaction)이 많다는 것을 의미한다. 또한 공통 네트워크는 노드들 간 매우 치밀하게 연결되어 있다는 것을 뜻한다. 또 다른 지표로 평균 이웃의 수(ave. number of neighbors)를 보면 8.5(꽃), 6.6(뿌리), 3.6(내엽), 3.2(외엽), 2.7(어린 잎 조직)로서 네트워크의 크기가 클수록 더 많은 이웃의 수가 있음을 확인했다.

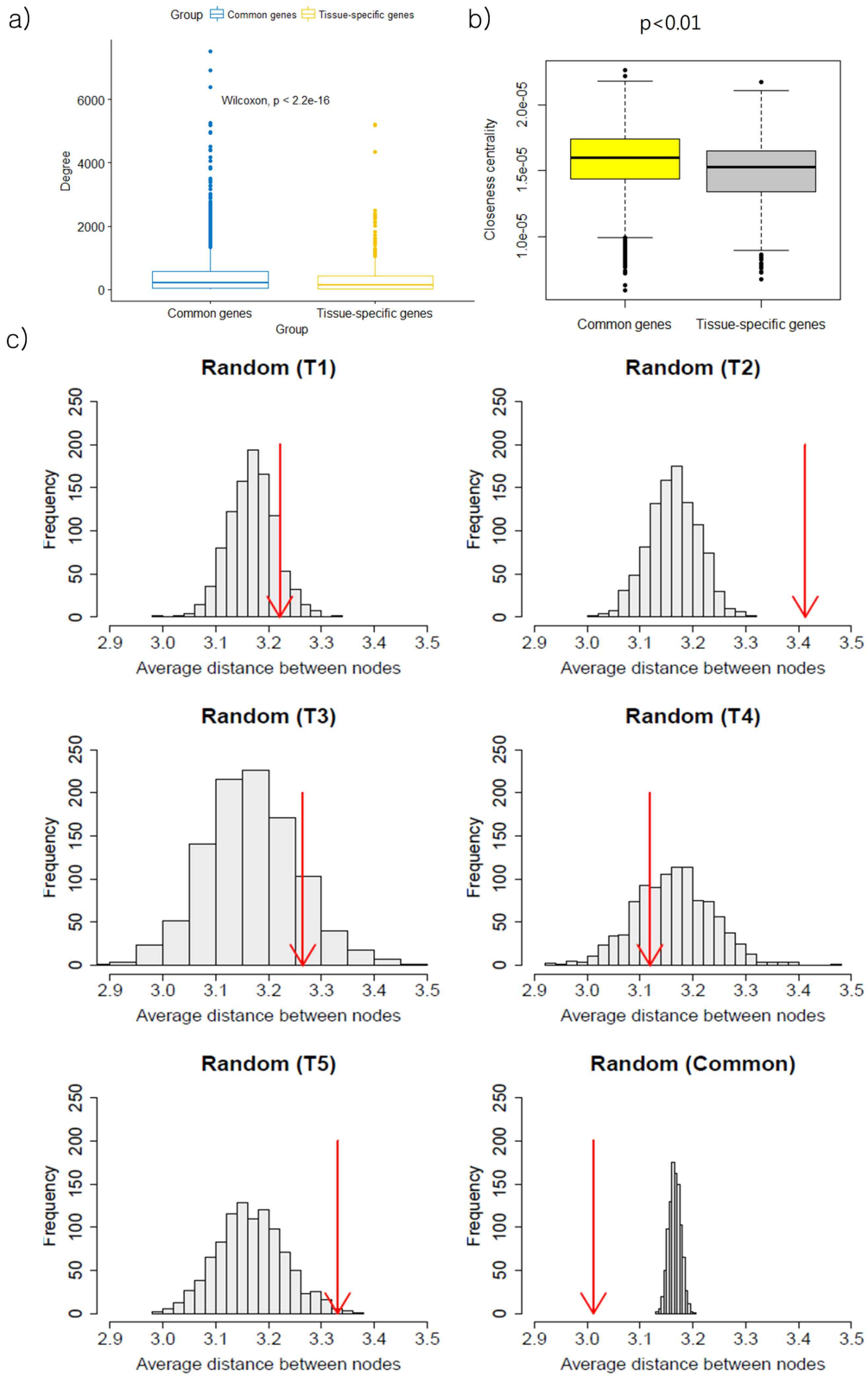


그림 15. 부분 네트워크의 위상적 특성 분석 결과

8. 컴퓨터 시뮬레이션을 통한 신호전달경로의 동역학적 특성을 분석하고 신호전달경로에 의한 반응을 결정짓는 핵심 분자를 규명

○ GRN에서 특정 조직의 활성 하위 네트워크를 재구성하기 위해서는 두 가지 종류의 정보가 필요하다. 즉, 물리적으로 가능한 모든 유전자 규제 상호작용을 포함하는 통합 GRN과 특정 조직에서 활성 상태인 분자 목록이다. 첫째, 통합 GRN 데이터는 ARM(Abradopsis Transcriptional Regulatory Map)에서 다운로드 되었다. 현재 *Brassica* 네트워크에는 알려진 GRN이 없지만, 애기장대, 쌀, 보리, 옥수수, 토마토, 감자, 담배와 같은 일부 모델 유기체의 경로 정보를 이용할 수 있다. 이 중 *A. thaliana*는 배추와 무를 포함한 십자학과에 속하는 식물이다. 게다가, ATRM이 높은 신뢰도의 전사적 규제 지도를 가진 최첨단 데이터 세트이다. 따라서, 우리는 시퀀스 유사성(E-값 E1E-10 및 적용 범위 0.0.75)을 기반으로 BLASTP 프로세스에 의해 모든 *B. rapa* 유전자 ID를 *A. thaliana* 유전자 ID로 변환했다. 다음으로, 우리는 가장 큰 연결 구성요소를 *B. rapa*의 GRN으로 확인하였고, GRN에는 725개의 노드와 1379개의 링크가 포함되어 있다.

○ 네트워크 분석에 사용될 식물 재료로 배추의 레퍼런스 식물인 Chiifu 401-42를 사용하였다. 생육 조건은 21°C, 65% 습도, 낮 16시간, 밤8시간의 조건으로 2개의 그룹으로 실험하였다. 첫 번째 그룹은 4주간 생육하였고 다른 한 그룹은 14주 동안 생육하여 결구형성된 상태까지 키웠다. Inner은 결구의 최외부 잎에서부터 4번째의 잎을 사용하였다. rosette 잎은 식물의 최외부에서 5번째 되는 잎을 사용하였다.

○ RNA sequencing은 Illumina HiSeq 2000 platform을 이용하였으며, seedling, rosette 잎, inner 잎을 사용하였다. Chiifu를 기준으로 mapping rate은 79.36%을 달성하였다.

○ FANMOD를 이용한 네트워크 motifs 식별

네트워크 모티프의 식별은 GRN의 설계 원칙을 밝히는 데 유용한 단계이다. 네트워크 모티프를 식별하기 위해 FANMOD (FAst Network MOTif Detection)가 통합 GRN에 사용된다. p 값 <0.05 및 Z 점수 > 2가 선택 기준으로 선택되었고 100,000 개의 임의 네트워크가 생성된다. 무작위 네트워크 생성을 위해 각 노드에 대한 활성화 및 금지 링크 수를 보존하면서 에지를 다시 배선하였다.

○ Functional annotation of DEGs

R 패키지 'clusterProfiler'를 사용하여 유전자 온톨로지 (GO) 범주가 유전자 집합에서 통계적으로 과대 표현되는지 여부를 계산하였다. p- 값은 초기 hypergeometric 테스트를 통해 얻어졌으며 FDR 보정이 적용되었다.

○ ODE modeling and simulation analysis

MATLAB (2013a 버전; TheMathWorks, Inc, Natick, MA, USA)의 함수 ode45를 사용하여 미분 방정식을 통합하였다. 여기서 k_{dx} , k_{dy} , k_{dz} 는 X, Y, Z (출력)의 열화율 상수를 나타낸다. 반응 매개 변수 k_{xy} (k_{xz} 및 k_{yz})는 X (X 및 Y)의 운동 반응을 나타내며 Y (Z 및 Z)

의 중요한 반응을 유도하며 기사에서는 경쟁 OR- 게이트 논리에 중점을 둔다. N은 Hill 계수를 나타낸다

○ Identification of active subnetworks for the inner and rosette leaves

기본 GRN에서 특정 조직의 활성 하위 네트워크를 재구성하려면 물리적으로 가능한 모든 유전자 조절 상호 작용을 포함하는 통합 GRN과 특정 조직에서 활성 목록의 두 가지 정보가 필요하다. 첫번째는, 통합 GRN 데이터에서의 Arabidopsis Transcriptional Regulatory Map (ATRM)의 정보이다. 두번째로는 차별적으로 발현된 유전자 전사체 분석을 기반으로 inner 및 rosette 잎에서 활성 유전자 정보이다.

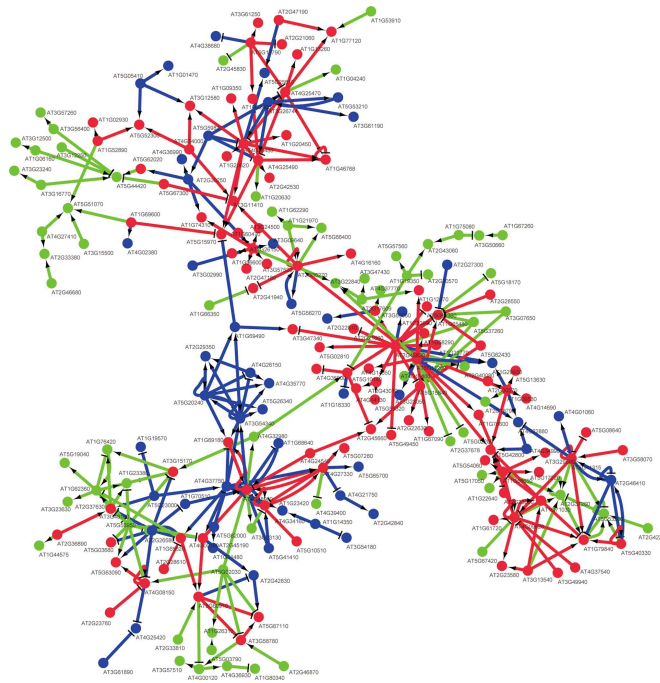


그림 16. 내부 및 외부 잎의 통합된 활성 하위 네트워크

○ 활성 하위 네트워크에서 Brassica ID는 TAIR ID로 변환되었다. ILRN과 RLRN이 통합되었다. 노드(링크)의 경우, 적색, 녹색 및 청색은 각각 ILRN에서만, 그리고 RLRN에서만 양쪽 활성 하위 네트워크에서 식별된 유전자(규제)를 나타낸다. 218개 유전자(335개 규정) 중 102개 유전자(124개 규정)가 양쪽 활성 하위 네트워크에서 확인됐다. 조직별 링크의 비율이 조직별 노드보다 훨씬 높다는 점에 주목했다 (그림16).

○ 둘째로, 서로 다른 유전자 표현된 대화 상자 분석을 바탕으로 내부와 로제트 잎에서 활동적인 유전자를 확인했다. 다르게 표현된 성적표는 내부(로제트)와 어린 잎의 유전자 발현 프로파일 사이에서 absolute log2 fold-change 11.0(fold-change ≥ 2 와 -2)의 기준으로 결과 데이터를 필터링하여 결정하였다. 여기서는 기준이 높을수록 DEG가 더 적게 선택되었다. 따라서, 결과적으로 발생하는 활성 하위 네트워크는 점점 더 작아질 것이다. 다양한 DEG 선택 기준(fold-change > 1.0(1.2배, 1.4배)을 적용했다. 그 결과, 내측과 외측 잎의 DEG는 각각 6460개(5,043개, 4,151)와 6,729개(5,304개, 4,284)가 확인되었다.

○ 마지막으로, 각 잎에 대한 활성 하위 네트워크를 식별하기 위해 다음과 같은 규칙을 사용했는데, 기본 네트워크의 링크에서 나온 두 개의 유전자가 모두 DEG 세트에 포함되어 있다면, 그 두 개의 유전자와 링크는 활성 하위 네트워크에 포함된다. 그 후, 네트워크의 활성 유전자가 함께 작용해야 하는 것이 생물학적으로 합리적이기 때문에 가장 큰 연결 구성요소를 확인했다. 그 결과, 169개(119개 및 107) 노드와 231개(150개 및 133) 가장자리로 구성된 ILRN과 RLRN을 각각 재구성할 수 있었고, 151개(119개 및 102개)와 228개(169개 및 146개) 가장자리로 구성되었다.

○ 생물학적으로 의미 있는 하위 네트워크를 획득했는지를 더 테스트하기 위해, 각각 내부와 로제트 잎에서 DEG의 수를 보존하면서 10,000개의 무작위 DEG 세트를 샘플링하고, 가장 큰 연결 구성요소의 크기를 계산했다. 우리는 식별된 ILRN과 RLRN이 랜덤 DEG 세트의 구성품에 비해 훨씬 큰 연결 구성요소를 형성한다는 것을 발견했다(그림 2). 구체적으로는, 기본 네트워크의 204개 DEG와 208개 DEG 중에서 각각 169개 노드와 151개 노드가 ILRN으로 식별되었다. 이 결과는 활성 하위 네트워크 구축에 단순한 규칙을 적용했지만, 대부분의 DEG가 활성 하위 네트워크의 노드로 포함되었음을 나타낸다. 이것은 또한 우리가 발견한 DEG가 GRN에서 서로 밀접하게 상호 연관되어 있다는 것을 의미한다. 따라서, 우리는 식별된 ILRN과 RLRN이 내부와 로제트 잎의 서로 다른 발달 과정을 나타낼 수 있다고 결론 내릴 수 있다 (그림17).

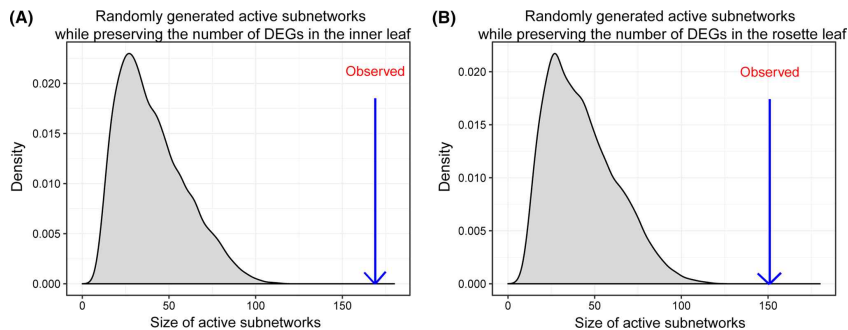


그림 17. 임의로 선택한 DEG 세트를 기준으로 활성 하위 네트워크의 크기 분포. 각각 내측(A) 및 외측(B) 잎의 DEG 수를 보존하면서 무작위로 선택된 10,000개의 DEG 세트로부터 얻은 활성 네트워크의 크기 분포이다. "Observed"는 내부(A) 및 외부(B) 잎의 활성 하위 네트워크의 크기를 나타낸다.

○ 네트워크 모티프는 네트워크에서 빈번하고 고유한 하위그래프 패턴으로 정의된다. 네트워크 모티프를 분석하면 진화 설계 원리(13)를 찾는 데 도움이 된다. 네트워크 모티프 중, FFL(Feed Forward Loop) 구조는 두 개의 입력 전사 인자로 구성되며, 이 인자는 다른 인자를 조절하며, 둘 다 대상 유전자를 공동으로 조절한다. FFL은 세 개의 링크(13개)의 규제 유형(활성화 또는 억제)의 조합에 따라 더욱 일관성 있는 FFL과 일관성 없는 FFL로 나눌 수 있다. 각각의 일관성 있는 FFL에서 두 노드 사이의 간접 경로는 직접 경로와 전체 기호가 동일하다. 반대로, 각 불변의 FFL에서 간접 경로의 기호는 직접 경로의 기호와 반대된다.

○ RLRN과 ILRN의 3노드 네트워크 모티프 분석에서, 두 가지 유형의 일관성 있는 FFL이 ILRN에서 농축된 반면, 세 가지 유형의 일관성 없는 FFL이 RLRN에서 농축된 것을 확인했다(그림 18). 각 조직별 활성 하위 네트워크에 대한 유의한 FFLs 중에서 각각 가장 대표적인 일관성(C1)과 불일치(I1) FFLs 두 개를 선택한 다음(그림 18 참조), 일반 차등 방정식에 기초한 수학적 모델의 구성을 통해 동적 특성을 조사했다.






| | | Identified network motifs (Z-score (p value)) | | | | |
|--------------------------|--------------------|---|---|---|---|---|
| Network motif types | | Coherent FFLs | | Incoherent FFLs | | |
| Network motif IDs | | C1 | C2 | I1 | I2 | I3 |
| Network motif topologies | |  |  |  |  |  |
| Active subnetwork | DEG cut-off (log2) | | | | | |
| ILRN | 1.0 | 8.08 (p<0.0001) | 8.46 (p<0.0001) | ns | 3.46 (p=0.003) | ns |
| | 1.2 | 6.46 (p<0.0001) | 3.70 (p=0.001) | ns | 3.07 (p=0.004) | ns |
| | 1.4 | 5.75 (p=0.0001) | 3.24 (p=0.003) | ns | ns | ns |
| RLRN | 1.0 | ns | 3.31 (p=0.003) | 4.00 (p=0.0009) | 9.09 (p<0.0001) | 3.94 (p=0.003) |
| | 1.2 | ns | 2.85 (p=0.006) | 4.50 (p=0.0004) | 4.81 (p=0.0001) | 5.40 (p=0.0001) |
| | 1.4 | ns | 2.94 (p=0.0053) | 4.49 (p=0.0005) | 5.14 (p=0.0001) | 5.22 (p=0.001) |

그림 18. Network motif analysis of ILRN and RLRN. 각 네트워크 모티브에 대해, Z 점수는 $(N_{real} - N_{rand}) / \text{standard deviation}$ 에 의해 얻었는데, 여기서 N_{real} 은 주어진 네트워크에서 관측된 네트워크 모티브의 수를 나타내며, N_{rand} 와 S_{rand} 는 각각 랜덤화된 네트워크에서 나타나는 평균 및 표준 편차를 나타낸다. 각 P-값은 실제 네트워크에 비해 랜덤화 네트워크에서 네트워크 모티브가 더 많이 관측된 횟수를 10만 회 이상 나타낸다. 'ns'는 'not marent'를 나타낸다.

○ 모델을 구성한 후, 두 가지 다른 유형의 자극(noise-free on-step stimulus, noisy on- and off-step stimulus)을 고려했다. 우리는 coherent FFL이 소음 신호에 대해 느리지만 상당히 강력한 반응(그림 17A)을 발생시킬 수 있다는 것을 발견했다. 반면에, 우리는 incoherent FFL이 빠른 반응을 보인 다음 일관된 입력(그림 19A)과 잡음 신호에 적용할 수 있다는 것을 알게 되었다(그림 19B).

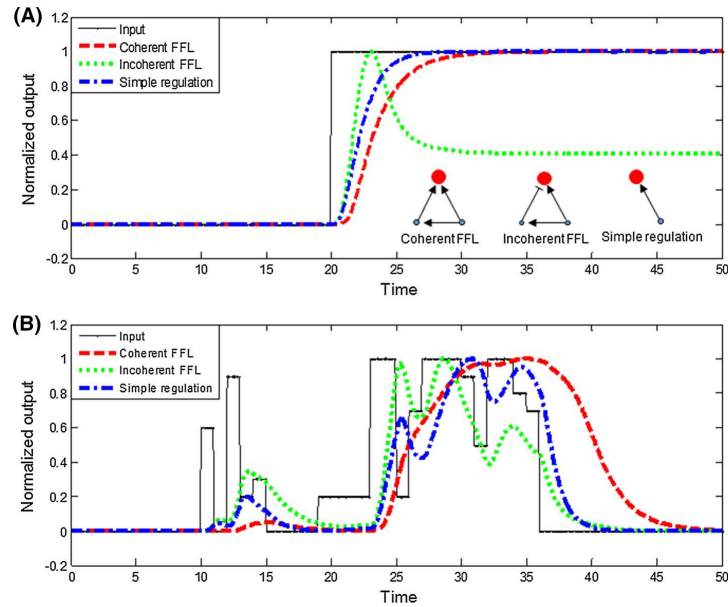


그림 19. coherent FFL과 incoherent FFL에 대한 수학 모델의 시뮬레이션 결과. noise-free on-step(A) 및 noisy on- and off-step(B) 자극에 대한 출력 노드의 반응 곡선. Coherent FFL, incoherent FFL 및 간단한 규제 모델에 대한 시뮬레이션을 수행했다. (A)에서 출력 노드(빨간색 원)가 있는 3개 모델의 아키텍처를 설명했다. 출력 응답은 최대값이 1로 표준화되었다.

○ 왜 ILRN과 RLRN은 각각 Coherent FFL과 incoherent FFL을 독점적으로 가지고 있는가? 유전자 조절 네트워크의 숨겨진 메커니즘을 밝히는 것은 생물학에서 중요한 문제이다. 세포는 단순히 유전자 발현 수준에 의해 결정될 수 있는 정적 시스템이 아니다. 대신 유전자 간의 조절 관계에 의해 지배되는 동적 시스템이기 때문에 단순히 일반적으로 발현되는 유전자를 발견하는 것보다 다른 조직에서 활동하는 네트워크를 식별하는 것이 더 중요하다. 따라서 우리는 ATRM에서 얻은 배추 두 조직의 구조적 및 기능적 차이를 참조 네트워크로 발견했다. coherent feedforward loops는 inner 잎에서 상당히 풍부 해졌지만 incoherent feedforward loops는 resette 잎에서 상당히 과도하게 표현되었다는 결론은 두 활성 하위 네트워크의 토폴로지 특성을 분석하여 달성되었다. 이 결론은 특정 유전자 세트의 분석으로 추론 할 수 없다.

○ 반면, 로제트 잎은 여러 가지 환경적 스트레스를 직접적으로 받는다. 따라서 로제트 잎은 환경의 변화에서 살아남기 위해 생리적, 분자적, 세포적 수준의 적응적 과정을 촉발할 수 있어야 한다(표 9). 흥미롭게도 표 11과 같이 표피세포 분화의 빈도는 식물의 물과 가스 교환을 극대화하기 위한 적응과정이다. 우리는 또한 로제트 잎의 다양한 대응 프로그램(Cold acclimation, 표 10)이 적응 프로세스여야 한다는 점에 주목한다. 따라서, 일관성 없는 FFL은 입력 신호(19)에 대한 적응형 응답을 가질 수 있는 최적의 규제 구조 중 하나라고 설명할 수 있다. 전체적으로 FFL 네트워크 모티브의 편향된 관찰 패턴이 내부와 로제트 잎의 세포 기능과 밀접한 관련이 있다고 결론지을 수 있다.

표 10. GO term enrichment analyses for the genes in the FFL network motifs.

| Active subnetwork | GO ID | Description | Gene ratio | Background ratio | FDR | |
|-------------------|------------|--|----------------------------------|------------------|----------|----------|
| ILRN | GO:0009628 | Response to abiotic stimulus | 97/167 | 312/709 | 0.009457 | |
| | GO:0009416 | Response to light stimulus | 58/167 | 161/709 | 0.009457 | |
| | GO:0009314 | Response to radiation | 58/167 | 162/709 | 0.009457 | |
| | GO:0009718 | Anthocyanin-containing biosynthetic process | compound | 14/167 | 24/709 | 0.038841 |
| | GO:0009409 | Response to cold | | 39/167 | 103/709 | 0.044157 |
| | GO:0046283 | Anthocyanin-containing metabolic process | compound | 16/167 | 31/709 | 0.05651 |
| | GO:0009266 | Response to temperature stimulus | | 47/167 | 134/709 | 0.05651 |
| | GO:2000026 | Regulation of multicellular organismal development | | 39/167 | 106/709 | 0.05651 |
| | GO:0051239 | Regulation of multicellular organismal process | | 39/167 | 107/709 | 0.063053 |
| | GO:0009813 | Flavonoid biosynthetic process | | 20/167 | 45/709 | 0.080424 |
| | GO:0046148 | Pigment biosynthetic process | | 19/167 | 42/709 | 0.080424 |
| | RLRN | GO:0009266 | Response to temperature stimulus | 47/150 | 134/709 | 0.01354 |
| GO:0009628 | | Response to abiotic stimulus | 88/150 | 312/709 | 0.01354 | |
| GO:0010035 | | Response to inorganic substance | 46/150 | 141/709 | 0.052684 | |
| GO:0009409 | | Response to cold | 36/150 | 103/709 | 0.052684 | |
| GO:0009414 | | Response to water deprivation | 30/150 | 81/709 | 0.052684 | |
| GO:0009415 | | Response to water | 30/150 | 82/709 | 0.056709 | |

○ 각 활성 하위 네트워크에서 각각의 네트워크 모티브를 구성하고 있는 유전자를 선택하여 GO 용어 농축 분석을 수행했다. 풍부한 분석은 생물학적 공정(BP) 온톨로지(Ontology)를 기반으로 수행되었다. 유전자 비율과 배경 비율은 각 네트워크 모티브의 모든 유전자 중에서 특정한 GO ID를 가진 유전자의 비율과 기저 GRN의 모든 유전자 중에서 각각 특정한 GO ID를 가진 유전자의 비율을 나타낸다. 가장 중요한 다섯 개의 GO 아이디가 제시되었다(표10).

제5절 채소 작물의 유용 형질 관련 분자마커의 범용화

1. 기 보고된 배추의 다양한 형질 관련 분자마커의 SNP 마커 전환을 위한 배추 수집단(Core collection)의 변이 정보 생산 및 방식의 개선

본 실험실에서는 골든씨드프로젝트의 설립 이전부터 배추의 유전분석 및 다양한 형질을 대상으로한 GWAS(Genome-wide Association Study) 실험을 위한 기반을 구축하기 위해 한중일 삼국에서 수집한 배추(*Brassica rapa*) 유전자원 197점, 무(*Raphanus sativus*) 2점, 흑겨자(*Brassica nigra*) 2점에 대해 Whole genome resequencing(WGS)을 수행하였으며 2011년에 공개된 배추의 표준 유전체 1.2 version과 당시 사용가능했던 분석 플랫폼을 활용하여 각 수집 계통에 대한 SNP 정보를 생산하였다. 그러나 2016년 배추의 표준 유전체 2.0 version이 공개되었으며 지난 5년간 WGS data로부터 SNP를 생산하는 분석 프로그램 및 파이프라인의 발전이 괄목한 만한 성과를 거두었다. 따라서 과거, 본 실험실에서 구성한 배추의 SNP 생산방식을 현재 사용가능한 플랫폼 및 표준 유전체를 활용하여 재분석함으로써 SNP의 정밀도 및 양이 향상된 새로운 배추 집단의 SNP matrix를 구성의 필요성이 제기되었다.

가. 배추 표준 유전체 2.0 version 개요

2011년 국제 공동연구를 통해 배추의 표준 유전체가 공개되었으며 genome assembly에 illumina의 paired end type의 라이브러리가 활용되었다. 그러나 이는 현재 genome assembly에 널리 쓰이고 있는 pacbio read에 비해 상당히 짧은 길이를 가지며 당시 예측된 배추(cv. chiifu) 유전체 크기의 약 60%만을 반영하였다. 대부분의 assembled contig 및 scaffold가 유전자를 coding하는 영역에 주로 집중이 되었다. 반면 2016년 웹상에 공개된 배추 표준 유전체 2.0 ver [6]은 pacbio read를 활용하여 기존보다 길이가 향상된 contig 및 scaffold를 구성하고 이전 version의 genome assembly에 활용된 illumina read로 base correction을 수행하였다. 또한 assembled scaffold 상에서의 gene prediction 또한 다시 수행되어 기존의 표준 유전체의 크기에 약 100 Mb 정도 증가된 유전체에서 7,053개의 유전자가 추가로 예측되었다. (표 1)

표 1. 배추 표준 유전체의 version에 따른 비교

| Version | Total assembled length | Number of scaffolds | Scaffolds containing genes | Total predicted genes |
|---------|------------------------|---------------------|----------------------------|-----------------------|
| ver 1.2 | 283.8 (Mb) | 40,367 | 284 | 41,173 |
| ver 2.0 | 389.2 (Mb) | 86,986 | 398 | 48,826 |

이와 같이 예측된 유전자의 개수와 유전체 크기가 증가된 표준 유전체 정보를 GWAS와 같은 유전분석이나 분자마커의 개발 등 다양한 목적에 활용하게 되면 기존에 사용되어왔던 표준 유전체 정보를 사용했을 때보다 얻을 수 있는 SNP 및 indel과 같은 변이 데이터의 양이 향상될 것이라 예상이 된다.

표 2. 배추 표준 유전체의 version upgrade에 따른 chromosome별 assemble chromosome의 길이 및 예측된 유전자 수의 변화

| Chromosome | Total assembled length (ver 1.2) | Total assembled length (ver2.0) | Length difference | Gene number difference |
|------------|----------------------------------|---------------------------------|-------------------|------------------------|
| A09 | 37,123,981 | 54,546,898 | +17.42 Mb | +1480 |
| A06 | 26,273,242 | 39,861,403 | +13.59 Mb | +1008 |
| A03 | 31,716,688 | 36,455,009 | +4.74 Mb | +918 |
| A05 | 23,941,934 | 36,115,060 | +12.17 Mb | +832 |
| A01 | 28,608,137 | 33,885,992 | +5.28 Mb | +666 |
| A02 | 27,848,129 | 30,435,970 | +2.59 Mb | +366 |
| A07 | 22,587,824 | 29,764,480 | +7.18 Mb | +931 |
| A08 | 21,596,550 | 27,726,665 | +6.13 Mb | +784 |
| A04 | 18,967,243 | 23,467,635 | +4.5 Mb | +524 |
| A10 | 17,595,035 | 18,561,454 | +0.97 Mb | +254 |

나. Re-sequenced genome fastq 파일의 전처리

계통별로 genome sequencing을 수행한 결과인 fastq 파일에 대하여 적절한 quality check(QC)와 trimming이 수행되지 않으면 이후 분석에서 잘못된 변이 데이터를 산출하게 된다. 따라서 fastq 파일내의 염기서열 상에서 adapter 혹은 vector 서열같은 비유전체 염기서열 및 phred score가 낮아 calling error 확률이 높은 염기서열을 제거해야 한다. fastq 파일의 QC는 fastqc를 통해 진행되며 분석 결과인 html 파일을 확인했을 때 크게 GC contents, sequence error rate, 그리고 overrepresented sequence와 같은 세 가지 사항에 대하여 filtering의 초점을 맞추게 된다 (그림 1).

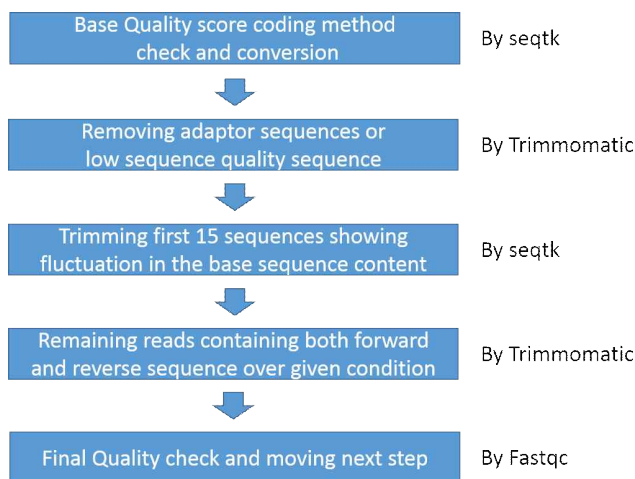


그림 1. Genome sequence fastq 파일의 quality trimming 작업 절차 개요

NGS 기술 도입 초기에 구성된 fastq 파일의 경우 sequence quality가 Phred score가 Q64로 작성되어 있는 경향이 있다. sequence quality filtering 이후의 과정인 BWA - mem에 의한 read alignment의 경우 Q33으로 코딩된 sequence만을 대상으로 하기 때문에 Q64를 Q33으로 전환하는 작업이 필요하다. sequence quality coding의 변환이 완료되면 각 계통별로 overrepresented sequence를 정리하고 이를 contaminant sequence로 간주하여 fastq 파일내에서 제거하는 작업을 Trimmomatic을 통해 Phred score 30 기준의 sequence quality trimming과 함께 수행하였다. Trimmomatic에 의한 filtering 이후 read sequence

말단의 일부의 라이브러리 구성 서열이 남아 GC contents 상에서 요동을 보일 수 있는데 이는 seqtk를 통해 제거한다. 말단의 제거 이후 일부 read sequence는 read 자체가 삭제되거나 30 bp 이하로 그 길이가 줄어들 수 있기 때문에 Trimmomatic을 통해 추가 실행하여 최소 길이가 30bp를 만족하고 forward와 reverse read가 전부 남아 있는 paired read만을 남긴다. 이와 같은 전 과정을 거치고 fastqc를 통해 필터링된 결과를 재확인하여 문제가 없을 경우 다음 단계인 reference에 대한 read alignment를 수행하게 된다 (그림 2).

| ASCII_BASE=33 Illumina, Ion Torrent, PacBio and Sanger | | | | | | | | | | | |
|--|---------|-------|----|---------|-------|----|---------|-------|----|---------|-------|
| Q | P_error | ASCII | Q | P_error | ASCII | Q | P_error | ASCII | Q | P_error | ASCII |
| 0 | 1.00000 | 33 ! | 11 | 0.07943 | 44 , | 22 | 0.00631 | 55 7 | 33 | 0.00050 | 66 B |
| 1 | 0.79433 | 34 " | 12 | 0.06310 | 45 - | 23 | 0.00501 | 56 8 | 34 | 0.00040 | 67 C |
| 2 | 0.63096 | 35 # | 13 | 0.05012 | 46 . | 24 | 0.00398 | 57 9 | 35 | 0.00032 | 68 D |
| 3 | 0.50119 | 36 \$ | 14 | 0.03981 | 47 / | 25 | 0.00316 | 58 : | 36 | 0.00025 | 69 E |
| 4 | 0.39811 | 37 % | 15 | 0.03162 | 48 0 | 26 | 0.00251 | 59 ; | 37 | 0.00020 | 70 F |
| 5 | 0.31623 | 38 & | 16 | 0.02512 | 49 1 | 27 | 0.00200 | 60 < | 38 | 0.00016 | 71 G |
| 6 | 0.25119 | 39 ' | 17 | 0.01995 | 50 2 | 28 | 0.00158 | 61 = | 39 | 0.00013 | 72 H |
| 7 | 0.19953 | 40 (| 18 | 0.01585 | 51 3 | 29 | 0.00126 | 62 > | 40 | 0.00010 | 73 I |
| 8 | 0.15849 | 41) | 19 | 0.01259 | 52 4 | 30 | 0.00100 | 63 ? | 41 | 0.00008 | 74 J |
| 9 | 0.12589 | 42 * | 20 | 0.01000 | 53 5 | 31 | 0.00079 | 64 @ | 42 | 0.00006 | 75 K |
| 10 | 0.10000 | 43 + | 21 | 0.00794 | 54 6 | 32 | 0.00063 | 65 A | | | |

| ASCII_BASE=64 Old Illumina | | | | | | | | | | | |
|----------------------------|---------|-------|----|---------|-------|----|---------|-------|----|---------|-------|
| Q | P_error | ASCII | Q | P_error | ASCII | Q | P_error | ASCII | Q | P_error | ASCII |
| 0 | 1.00000 | 64 @ | 11 | 0.07943 | 75 K | 22 | 0.00631 | 86 V | 33 | 0.00050 | 97 a |
| 1 | 0.79433 | 65 A | 12 | 0.06310 | 76 L | 23 | 0.00501 | 87 W | 34 | 0.00040 | 98 b |
| 2 | 0.63096 | 66 B | 13 | 0.05012 | 77 M | 24 | 0.00398 | 88 X | 35 | 0.00032 | 99 c |
| 3 | 0.50119 | 67 C | 14 | 0.03981 | 78 N | 25 | 0.00316 | 89 Y | 36 | 0.00025 | 100 d |
| 4 | 0.39811 | 68 D | 15 | 0.03162 | 79 O | 26 | 0.00251 | 90 Z | 37 | 0.00020 | 101 e |
| 5 | 0.31623 | 69 E | 16 | 0.02512 | 80 P | 27 | 0.00200 | 91 [| 38 | 0.00016 | 102 f |
| 6 | 0.25119 | 70 F | 17 | 0.01995 | 81 Q | 28 | 0.00158 | 92 \ | 39 | 0.00013 | 103 g |
| 7 | 0.19953 | 71 G | 18 | 0.01585 | 82 R | 29 | 0.00126 | 93] | 40 | 0.00010 | 104 h |
| 8 | 0.15849 | 72 H | 19 | 0.01259 | 83 S | 30 | 0.00100 | 94 ^ | 41 | 0.00008 | 105 i |
| 9 | 0.12589 | 73 I | 20 | 0.01000 | 84 T | 31 | 0.00079 | 95 _ | 42 | 0.00006 | 106 j |
| 10 | 0.10000 | 74 J | 21 | 0.00794 | 85 U | 32 | 0.00063 | 96 ` | | | |

그림 2. Sequencing platform에 따른 phred score coding 방식의 차이(http://www.drive5.com/usearch/manual/quality_score.html)에서 발췌

다. 배추 표준 유전체 2.0 ver에 대한 read alignment

각 계통에 대한 정제된 read sequence는 BWA를 통해 reference genome에 mapping 되었 으며 mapping 결과로 sequence alignment map(SAM) 형식의 파일을 얻었다. 이는 picard-tool에 의해 후속 연산에 불필요한 duplicated read들을 식별하고 이를 data pool에 서 제거한다. 남은 read sequence는 GATK에 의해 계통별 read sequence상의 nucleotide 각각을 수박 표준 배추의 신규 표준 서열에 대조하여 SNP와 Indel에 대한 데이터를 생산 되어 bcftool에 의해 variant calling format(VCF) 형식으로 전환된다. 얻어진 VCF 파일은 하나의 파일 내에 특정 계통이 보이는 SNP와 Indel에 대한 표준 유전체 상의 위치를 전부 갖기 때문에 이를 vcftool을 사용하여 SNP에 대한 VCF 와 Indel에 대한 VCF로 분리하였 다.

라. 기존 작물 집단의 SNP matrix 구성의 문제점의 확인

이전 단계에서의 배추 및 수박의 계통들에 대한 수집 및 생산한 genome sequence 데이터들에 대한 SNP 생산 방식은 이전부터 범용적으로 variant calling에 사용되고 있는 GATK 파이프라인을 따르고 있다. 단, 본 연차에서 부터 생산하는 SNP 생산 방식은 과거에 사용하였던 read alignment 알고리즘인 BWA - aln 을 BWA - mem으로 전환한 것과 배추의 update된 표준 유전체를 reference로서 도입한 것이 가장 큰 개선점이다. 그 이후의 단계인 mark duplicate, indel realignment 그리고 variant calling으로 이루어지는 절차는 이전의 방식과 동일하다.

이전의 집단의 SNP matrix를 구성할 때, 각 계통별 SNP 정보를 포함하고 있는 vcf 파일들을 GATK의 CombineVariant 기능을 이용하여 집단을 구성하는 계통들의 변이정보를 하나로 합친 vcf 파일을 만들었다. 그러나 이 경우 한 계통에서 특이하게 보유하는 SNP position이 있을 때 CombineVariant의 결과로 다른 계통들의 SNP 정보가 이 position에 대해서 missing value를 갖게 된다. 그 결과, 이러한 missing value를 갖는 position이 나타난 이유는 크게 두 가지로 들 수 있다. 첫째로 실제 특이한 SNP를 가지고 있지만 해당 위치에 align된 read sequence가 없어서 missing value로 나타나게 된 것을 들 수 있다. 두 번째로는 해당 위치에 align된 read sequence가 reference와 동일하여 이 계통의 variant calling에서 variant로서 나타나지 않은 것을 생각할 수 있다. 이전 과정에서는 모든 missing value를 reference의 allele로 치환하는 것으로 처리하여 SNP의 missing value가 없는 matrix를 만들었지만 후속 분석과정인 GWAS에서 false positive의 가능성이 높은 SNP를 다수 생산한다는 문제점을 내포하였다 (그림 3).



그림 3. GATK CombineVariant 방식을 통한 통합 VCF 파일 구성의 문제점. 통합된 VCF 파일상의 ./(missing value)가 나타나는 실질적인 원인 파악이 불가능하다.

상술한 기존의 SNP 생산방식의 맹점을 해결하기 위해서는 SNP로 추정되는 각 유전체 상의 position에 대해서 수집된 계통 전체에 대하여 동시에 SNP를 calling하는 방식의 도입 필요성이 제기되었다.

마. Joint variant calling의 도입

하나의 작물 집단을 이루는 계통들에 대하여 re-sequencing을 수행하고 이에 대한 reference alignment 결과를 가지고 있을 때 유전체 상의 각 위치에 대한 변이의 여부를 모든 계통으로부터 산출하는 방식을 확인하였다 (그림 4).

Variant Calling

To convert your BAM file into genomic positions we first use mpileup to produce a BCF file that contains all of the locations in the genome. We use this information to call genotypes and reduce our list of sites to those found to be variant by passing this file into bcftools call.

You can do this using a pipe as shown here

```
samtools mpileup -ugf <ref.fa> <sample1.bam> <sample2.bam> <sample3.bam> | bcftools call -vm0 z -o <study.vcf.gz>
```

Alternatively if you need to see why a specific site was not called by examining the BCF, or wish to spread the load slightly you can break it down into two steps as follows:

```
samtools mpileup -go <study.bcf> -f <ref.fa> <sample1.bam> <sample2.bam> <sample3.bam>  
bcftools call -vm0 z -o <study.vcf.gz> <study.bcf>
```

To prepare our VCF for querying we next index it using tabix:

```
tabix -p vcf <study.vcf.gz>
```

Additionally you may find it helpful to prepare graphs and statistics to assist you in filtering your variants:

```
bcftools stats -F <ref.fa> -s - <study.vcf.gz> > <study.vcf.gz.stats>  
mkdir plots  
plot-vcfstats -p plots/ <study.vcf.gz.stats>
```

Finally you will probably need to filter your data using commands such as:

```
bcftools filter -O z -o <study_filtered.vcf.gz> -s LOWQUAL -i 'QUAL>10' <study.vcf.gz>
```

Variant filtration is a subject worthy of an article in itself and the exact filters you will need to use will depend on the purpose of your study and quality and depth of the data used to call the variants.

그림 4. Samtools manual webpage(http://www.htslib.org/workflow/#mapping_to_variant)

상에 기술된 multiple sample로부터의 variant calling 방식 및 절차

samtools의 mpileup 기능과 bcftools의 call 기능을 리눅스 분석 서버의 셸 프롬프트 상에서 파이프를 연결하면 복수의 bam 파일(계통별 read data가 reference alignment 후 정제된 결과)로부터 유전체 상의 잠재적인 변이 위치에 대하여 계통별 변이 정보를 동시에 얻을 수 있다. 작업의 효율화를 위하여 배추 각 계통에서 얻은 SNP position의 리스트를 취합하고 중복된 포지션의 경우 하나만 남긴 리스트를 만들었다. 그리고 samtools와 bcftools의 piped process에 반영하여 이 배추 집단에서 unique한 포지션에 대해서만 joint variant calling을 수행하도록 하였으며 리스트를 Chromosome(A01 ~ A10) 번호 별로 분할하여 작업을 분석 서버 내에서 병렬 처리함으로써 효율화할 수 있었다.

바. SNP matrix의 신뢰성 향상을 위한 filtering 처리

배추 집단에 대하여 joint variant calling을 통해 SNP matrix를 구성한 이후 분석에 사용가능한 신뢰성 있는 SNP만을 남기기 위한 작업 절차가 필요하다. filtering의 우선 추후 유전 분석의 편이를 위해 multiple allelic position을 1차 SNP matrix에서 제거한 bi-allelic position만을 갖는 vcf file을 만든다. 이어서 SNP position의 최소 quality를 보장하기 위해 SNP position의 QUAL 30 이상 그리고 한 SNP position에 대한 모든 계통별 SNP의 genotyping quality(GQ)의 평균이 30 기준으로 filtering을 수행한다. 이후 PLINK software를 통해 MAF(minimum allele frequency) 기준이 0.05, missing rate 10%이내 Hardy-weinberg threshold 0.001 조건으로 최종 filtering을 수행하였다 (그림 5).

Workflow of SNP calling from collected germplasms

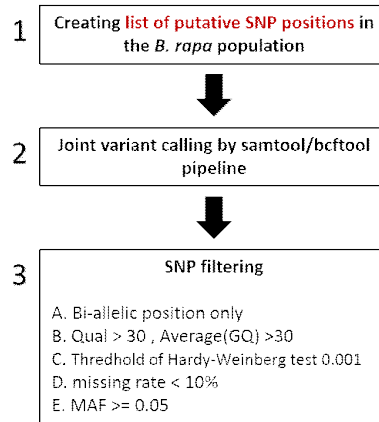


그림 5. Multiple sample의 joint variant calling 및 집단 1차 SNP matrix의 최종 필터링 과정 개요

2. 배추 수집단의 SNP 위치에 대한 annotation 방식의 확립

snpEff는 자바(java) 기반으로 작성된 프로그램으로 연구 대상인 생물의 표준 유전체 및 유전체 상의 예측된 유전자의 위치 정보를 기반으로 입력된 변이 위치에 대하여 변이의 발생에 따른 전사체의 구조 변경에 대한 영향력 및 변이 위치의 생물학적 분류(exon, intron, intergenic region)를 수행하는 프로그램이다. snpEff에 의한 input data로 사용된 작물 집단의 계통별 변이 양상을 기록한 vcf file의 변이 발생 위치에 따른 영향력의 평가는 html 형식으로 출력되고 이를 리눅스 분석 서버에서 윈도우 운영체제로 전송 후 파일 내의 정보를 시각적으로 확인할 수 있다. 그리고 텍스트 형식으로 제공되는 파일을 통해 SNP가 발생한 유전자와 그 구조상의 위치, 유전자 구조상에 SNP 발생에 의한 codon이 지정하는 아미노산의 변경 효과를 표와 그림을 통해 시각적으로 확인할 수 있다. 현재 골든씨드프로젝트에서 다루는 5대 채소 작물 중 배추와 수박의 표준 유전체 정보를 확보하고 이를 snpEff에서 변이 위치의 annotation을 수행하기 위한 체계를 구성한 상태이다. 추후 고추(pepper)와 무(radish)를 대상으로 현재 두 작물에 대해 공개된 모든 표준 유전체 정보를 수집하고 SNP annotation을 위한 체계를 구성할 계획이다.

표 3. 주요 작물의 표준 유전체의 공개 일시 및 수행기관

| Crop | Scientific name | Predicted genome size | Journal and released date | Institute |
|-----------------|--------------------------|-----------------------|---|-----------|
| Chinese cabbage | <i>Brassica rapa</i> | 529 Mb | Nature Genetics; 2011.09. | MBGP |
| Radish | <i>Raphanus sativus</i> | 573 Mb | Theoretical and Applied Genetics; 2015.02 | RDA |
| Watermelon | <i>Citrullus lanatus</i> | 425 Mb | Nature Genetics; 2012.12. | IWGI |
| Pepper | <i>Capsicum annuum</i> | 2700 Mb | Nature Genetics; 2014.01 | SNU |

3. 배추 GWAS 분석체계의 정교화를 위한 배추 표준 유전체 기반 정보의 생산

현재, 배추 표준 유전체의 염기서열 뿐만이 아닌 유전체 상에 분포하는 다양한 속성에 대한 정보가 상당 부분 update가 되었지만, 여전히 read sequence의 genome assembly에서 구성된 scaffold가 10개의 chromosome으로 통합되지 않은 상황이다. 추후 ordering이 완료된 scaffold들의 사이에 존재하는 gap을 그 위치에 해당하는 염기서열 정보를 update하고, 통합되지 않은 다양한 길이의 scaffold의 위치를 확정해야만 배추의 실제 유전체 정보를 최대한으로 반영하는 표준 유전체 정보를 생산할 수 있다. 이를 위한 기반 연구로서 Bionano genomics의 Irys platform을 활용하여 chiifu 계통에 대하여 optical mapping을 수행하였다 (그림 6).

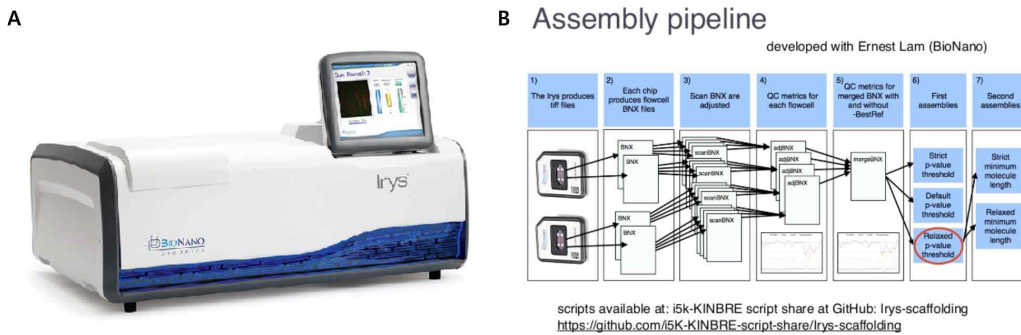


그림 6. Bionano Irys platform; A) Irys Instrument (IN-011-01), B) Irys output data의 분석 파이프라인

Irys는 NGS(Next generation sequencing)의 직접적인 염기서열의 생산이 아닌 특정 제한효소의 작용부위를 유전체상에 시각적으로 표시하는 방식의 sequencing 결과를 생산한다. Irys의 read size는 현재 NGS 방식에서 가장 긴 길이의 read를 생산하는 PacBio의 read size를 크게 상회하며 생산한 read상의 제한효소 위치의 패턴을 기준으로 생산된 read의 assembly를 수행하여 기존의 PacBio 방식에 비해 assembled contig size가 더욱 향상된 scaffold를 구성할 수 있다. 그러나 이 방식은 제한 효소의 위치만을 알려주는 optical map만을 생산하기 때문에 별도의 NGS 방식으로 구성된 scaffold의 sequence 상의 제한 효소 위치의 패턴을 optical map 상의 pattern과 맞추어 NGS scaffold의 ordering과 scaffold의 assembly를 수행해야만 한다. 본 작업에서는 향후 배추 표준 유전체의 update를 목표로 배추의 표준 계통 chiifu에 대한 optical mapping을 수행하였다.

가. Optical molecule의 구성

Optical mapping을 위해 두 개의 서로 다른 제한 효소(BspQI, BssSI)를 병행적으로 처리하고 각 제한 효소에 따른 optical molecule을 구성하였다. 그 결과, BspQI와 BssSI에 대하여 각각 177,437, 133,475개의 read를 구성하였으며 이들은 각각 44,064Mb, 27,936Mb의 genome size에 상응한다. 또한 각 제한효소 별 read에 대한 label density는 7.4/100kb, 10.5/100kb로 BspQI가 BssSI에 비해 더 높은 밀도로 제한효소의 위치가 labeling된 것으로 나타났다(표 4).

표 4. 서로 다른 두 제한효소를 기반으로 구성된 Optical read의 기본 정보

| Enzyme | Quantity (Mb) | Mol. N50 (Kb) | Number of molecules | Avg. Label Density (per 100 Kb) |
|--------|---------------|---------------|---------------------|---------------------------------|
| BspQI | 44,064.0 | 244.9 | 177,437 | 7.4 |
| BssSI | 27,936.6 | 198.4 | 133,475 | 10.5 |

나. Optical read의 assembly 결과

서로 다른 두 제한효소(BspQI, BssSI)로부터 얻어진 optical read를 read 자체 내의 제한 효소 위치 정보를 기반으로 하여 read의 assembly를 수행하였다. 수행 결과는 표 5과 같다.

표 5. 서로 다른 두 제한효소를 기반으로 구성된 Optical read의 기본 정보

| Enzyme | Number of Consensus Genome Maps | Consensus Genome Maps Size (Mb) | Consensus Genome Maps N50 (Mb) | Effective coverage of assembly |
|--------|---------------------------------|---------------------------------|--------------------------------|--------------------------------|
| BspQI | 723 | 410.575 | 0.905 | 93.691 |
| BssSI | 994 | 326.061 | 0.348 | 78.574 |

Read assembly 결과, BspQI 기반 read는 410.575 Mb의 genome size를 갖는 723개의 scaffold를 구성하였으며 BssSI 기반 read의 경우 상대적으로 작은 326.061 Mb 크기의 994개의 scaffold를 구성하였다. 또한 N50 수치를 비교한 경우 BspQI의 N50이 BssSI의 그것에 비해 약 3배 높은 수치를 보였다. 그리고 배추 표준 유전체 2.0 version이 assemble genome size가 389.2 Mb로 알려진 배추 유전체 크기의 약 80%를 다루지만 총 86,986개의 scaffold로 이루어진 것을 감안했을 때 본 작업에서 얻은 optical scaffold는 배추 표준 유전체의 길이보다 약 20Mb 길 뿐만 아니라 이를 723개의 scaffold로 통합하였다. 이는 추후 optical scaffold를 다른 NGS sequence와 결합하여 새로운 표준 유전체를 구성하였을 때 이상적인 경우 723개의 scaffold로 이루어진 배추 표준 염기서열을 얻을 수 있다. 또한 기존의 chiifu의 linkage map 정보를 활용 혹은 추가 생산함으로써 신규 배추 표준 유전체를 산출된 723개 보다 더 적은 수의 scaffold로 구성할 수 있음을 기대할 수 있다.

4. 배추 표준 유전체의 업데이트에 다른 집단 SNP 산출 방식의 재고 필요성

일반적인 채소작물의 경우, 그 지놈의 구조가 2배체이며 여러 순계들의 교배조합 상에서 F₁이 우수한 표현형을 나타나게 하는 양친을 선발하여 종자를 생산 및 판매한다. 이와 같은 양친을 육성하기 위해서는 반수체 생식세포의 double haplodization이나 수세대에 걸친 inbreeding을 통해 양친 계통의 유전적 순도를 높여 F₁ 생산 시 모든 F₁세대 유전적으로 동일하게 만들어 상품으로서의 품질 균일성을 도모한다. 그러나 채소작물은 그 대다수가 자가불화합성(self-compatibility), 웅성불임성(male sterility), 이형예(heterostylism) 그리고 자식열세와 같은 자가수정(selfing)의 효율을 저해하는 생물학적 체계를 지니고 있어 폐화수분(cleistogamy)을 하는 화본과 작물(Poaceae)과 같이 계통의 유전적 순도를 예측하고 유지시키는 것이 어렵다. 또한 유전체 자체가 진화과정 상에서의 polyploidy event와 retrotransposon 등의 활동으로 그 구조내의 상당한 부분이 중복서열로 이루어져 있어 sequencing 이후 de novo assembly의 결과를 통합하여 scaffold 수준의 일차적 결과를 chromosome 수준으로 높이는 것이 기술적으로 어렵다. 이를 극복하기 위해 기존의 illumina의 paired end sequencing보다 더 긴 길이의 read를 구성하는 Pacbio 및 유전체의 제한효소 인식 위치를 매우 긴 수준으로 patterning하는 Bionano의

optical mapping(Irys) 기술이 도입되어 기존의 de novo assembly의 한계를 넘어 더욱 정확한 chromosome sequence를 만들어 나가고 있는 것이 현 유전체 sequencing 프로젝트의 추세이다. 배추 또한 2011년 MBGP에 의해 첫 번째 reference genome이 공개된 이래 2016년에는 Pacbio 기술이 도입된 2.0 ver이 공개되었고 2018년에는 Bionano Irys가 도입된 3.0 ver이 공개되어 그동안 크로모솜에 통합되지 못했던 scaffold를 상당부분 통합한 개선된 reference genome이 공개되었다. 이와 같이 reference genome이 지속적으로 업데이트 됨으로써 더욱 정확한 SNP 정보를 얻기 위해서는 SNP 구성 작업도 업데이트되는 reference genome의 버전에 맞추어 갱신될 필요성이 있다. 따라서 배추를 기준으로 현재까지 공개된 reference genome의 특성을 서로 비교하고 새롭게 개발된 reference genome으로부터 현재 보유 중인 배추 수집단의 sequencing 결과를 적용하여 SNP 개발을 재수행하였다. 또한 작업의 진행과정에서 최종 SNP loci의 수에 영향을 미치는 주요인을 선정하고 요인들의 조합에 따라 SNP matrix를 구성하고 각 matrix의 특성을 평가하여 유전분석에 유리한 조건을 찾고자 하였다.

가. 현행 배추 표준유전체의 개발 이력

배추는 2011년 이후로 현재까지 총 세 번에 걸쳐 공개 reference genome의 버전의 업데이트가 진행되어 왔다. 어떤 sequencing platform을 이용하였는가에 따라 reference genome의 버전이 구분되며 버전이 올라갈수록 Illumina 이후에 공개된 기술이 하나씩 도입되어왔다.



그림 7. 공개되어있는 현행 배추 reference genome 관련 출판 논문

이와 같은 새로운 기술은 기존의 방식보다 더 길이가 증가된 read sequence를 확보하는 것에 초점이 맞추어져 있으며 이를 위한 Pacbio 기술의 경우 kilo base pair 단위의 sequence read를 얻을 수 있었지만 염기서열의 정확도가 Illumina의 그것에 비해 낮아 기존에 생산된 Illumina sequence data를 통해 sequence correction을 하는 방식으로 Illumina 방식이 달성하지 못한 contig의 길이 증가와 Pacbio가 취약한 형태를 보이는 sequence의 신뢰성을 상호보완하는 방향으로 reference genome을 만들 수 있었다. 그러나 Pacbio sequencing 만으로도 채소작물의 지놈 전반적으로 나타나는 중복영역의 문제를 전부 해결하기는 어려워 Bionano Irys를 통해 지놈 전체의 제한효소 위치를 patterning하고 여기에 Pacbio 기반의 contig를 대응시켜 과거보다 더욱 개선된 N50의 scaffold 및 pseudomolecule을 얻을 수 있게 되었다. 가장 최근에 공개된 HI-C sequencing은 chromosome 구성을 위한 scaffold의 ordering 방식으로 제안되어 구성이 완료된 scaffold를 방향 및 위치를 조정하는데 유용하게 쓰이고 있다. 배추의 경우 2011년 첫 번째 reference genome이 공개된 이래, 시퀀싱 기술의 발달과 공개에 힘입어 지속적으로 버전을 업데이트 시켜왔고 과거에 비해 상당한 수준의 N50 수치의 개선 및 chromosome 상에 통합되지 못한 scaffold의 수 역시 크게 줄이는 발전을 거듭하였다.

표 5. 배추 reference genome의 버전에 따른 genome property의 비교

| Genome status | Ver 1.5 | Ver 2.0 | Ver 3.0 |
|-----------------------------------|-----------|-----------|-----------|
| A01 | 26.79Mbp | 33.89Mbp | 29.6Mbp |
| A02 | 26.94Mbp | 30.44Mbp | 31.44Mbp |
| A03 | 31.77Mbp | 36.46Mbp | 38.15Mbp |
| A04 | 19.27Mbp | 23.47Mbp | 21.93Mbp |
| A05 | 25.3Mbp | 36.12Mbp | 28.49Mbp |
| A06 | 25.21Mbp | 39.86Mbp | 29.17Mbp |
| A07 | 25.88Mbp | 29.76Mbp | 28.93Mbp |
| A08 | 20.83Mbp | 27.73Mbp | 22.98Mbp |
| A09 | 38.88Mbp | 54.55Mbp | 45.16Mbp |
| A10 | 16.41Mbp | 18.56Mbp | 20.73Mbp |
| Scaffolds | 27.58Mbp | 60.59Mbp | 56.56Mbp |
| Total genome size | 284.86Mbp | 391.41Mbp | 353.14Mbp |
| No. of Scaffolds | 40,357 | 86,842 | 1,093 |
| Scaffold proportion in the genome | 9.68% | 15.48% | 16.02% |
| No. of Identified genes | 41020 | 48826 | 45985 |
| Published year | 2011 | 2016 | 2018 |

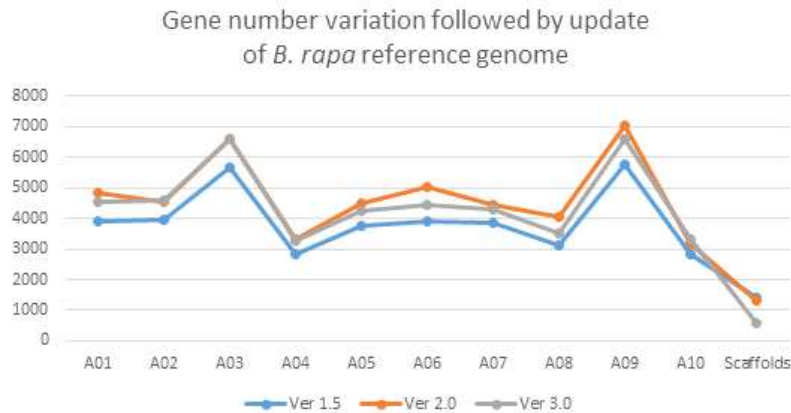


그림 2. 배추 reference genome 버전에 따른 크로모솜 상의 예측된 유전자 수의 변화

본 과제에서는 배추의 reference genome 2.0 ver의 공개 이후 이를 기반으로 한 SNP calling 및 유전분석을 수행하였다. 그러나 2018년 배추 reference genome 3.0 버전의 공개 이후 지금까지 공개된 reference genome이 가지는 특성을 비교한 결과(표 x, 그림 x), reference genome 2.0 ver이 다른 두 버전에 비해 다소 이질적인 양상을 보였다. 실제 배추 reference 버전 2.0은 버전 1.5와 3.0과 다른 배추 계통에서 얻은 sequencing 및 연관지도를 활용한 것으로 인해 위와 같은 결과를 나타내는 것으로 사료된다. 또한 추가적으로 배추 reference 버전 간 시퀀스의 일치도를 mummer software를 통해 확인할 결과, reference genome 1.5와 3.0은 높은 수준의 서열 유사도를 보이지만 reference genome 2.0을 다른 버전과 비교하였을 때 크로모솜 별로 일치도가 상당히 낮게 나타는 것을 확인할 수 있었다. (그림 9, 10)

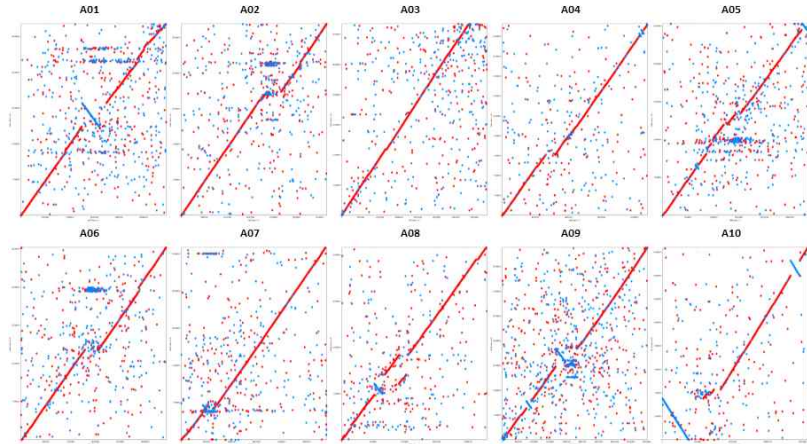


그림 9. 배추 reference genome 1.5와 3.0의 크로모좀별 일치도 비교

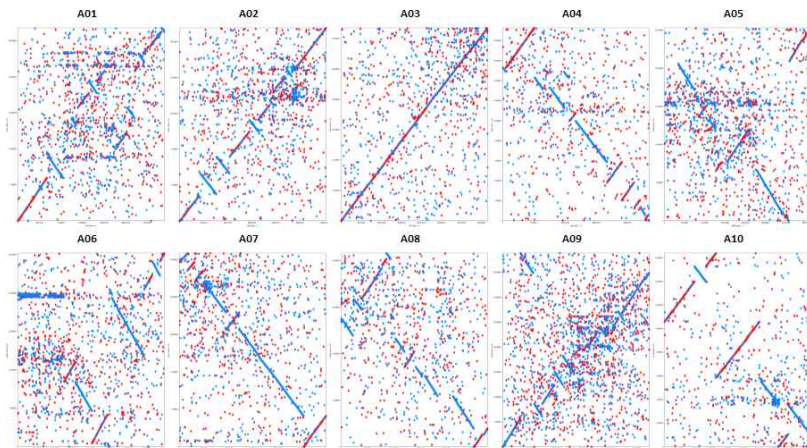


그림 10. 배추 reference genome 1.5와 2.0의 크로모좀별 일치도 비교

위와 같은 배추의 reference genome 특성에 대한 전반적인 특성을 고려하였을 때 배추의 유전연구에 널리 사용되어 왔던 reference genome 1.5와 높은 선형성을 가지며 chromosome length 및 scaffold ordering 문제를 상당 부분 해소한 reference genome 3.0으로 유전분석을 수행하는 것이 기존의 것보다 심도있고 정확한 결과를 얻을 수 있을 것이라 사료된다.

나. 배추 수집단의 Whole genome sequence 데이터의 reference genome 3.0 버전에 대한 alignment 수행 결과의 평가

배추 수집단 201개 계통 중 Pekinensis에 해당하는 175개 계통을 대상으로 하여 구축한 파이프라인을 통해 배추 reference 버전 3.0에 alignment 및 bam file 편집을 수행하였다. 결과 bam file은 Qualimap software를 통해 sequence filtering 및 bam file 편집과정에서 유지된 read sequence가 reference genome 3.0 버전을 depth 수준별로 어느 정도의 breadth를 보이는 지를 측정하였다(그림 11).

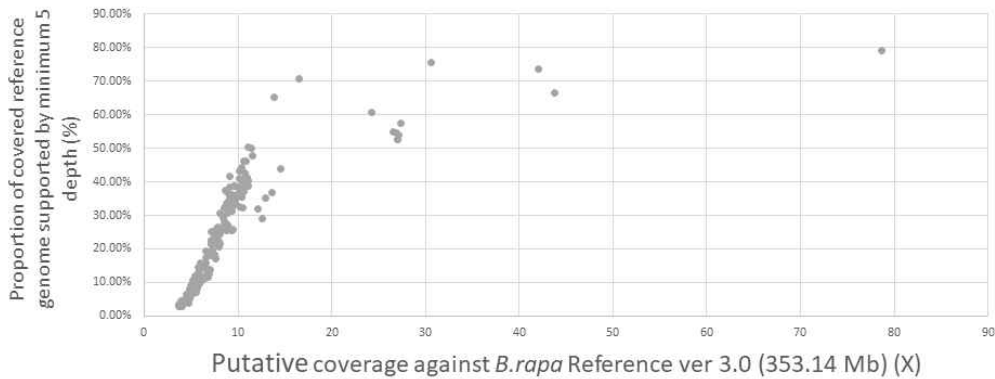


그림 11. 배추 175개 계통의 초기 sequencing coverage 대비 alignment 이후 reference genome을 5 depth 이상 커버하는 sequence 비율의 산포도.

분석에 사용된 175개 계통은 11개의 엘리트 라인을 제외하면 reference genome에 대해 평균 3 ~ 5X 수준의 coverage를 보인다. 이러한 계통들은 구축한 파이프라인을 거치면 주어진 reference를 5 depth 이상으로 커버하는 mapped region이 전체 지놈의 50% 이하인 것으로 나타났다. 또한 초기 sequencing coverage 수준이 10X 이상인 경우 계통의 sequencing quality에 따라 5X 이상 mapped region의 reference genome 상의 비율은 최대 70%를 나타냈으나 매우 큰 분산을 나타냈다. 오직 초기 sequencing coverage가 20X를 넘는 경우만이 안정적으로 50% 이상의 reference 상의 영역을 5 depth 이상으로 커버하는 것으로 나타났다. 뿐만 아니라 계통별 read sequence가 구축한 파이프라인을 거치면서 어느 정도 수준으로 유실이 되는 가를 확인한 결과 초기 sequence coverage 기준으로 60% 내외($61.65 \pm 4.46 \%$)의 read가 variant calling 과정 중에서 필터링되어 오직 40% 내외의 read 데이터로 variant calling이 수행된 것으로 나타났다.

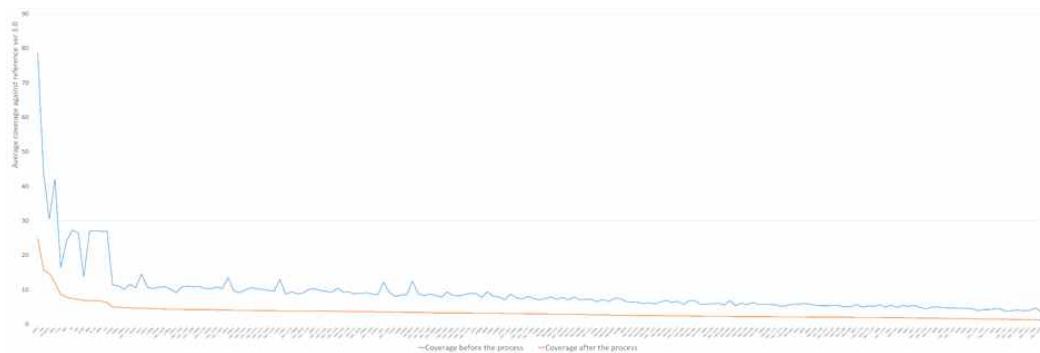


그림 12. 계통별 초기 sequencing coverage 대비 최종 mapped sequence coverage의 비교

채소작물의 variant calling을 위해 구축한 파이프라인의 경우 adapter sequence 및 low-phred score sequence의 제거 뿐만이 아니라 1차 필터링 과정에서 paired end의 양 sequence가 전부 유지된 pair만이 alignment에 반영되었고 alignment 이후에도 두 paired end sequence가 1000bp 이내의 insert size를 보여야만 variant calling에 활용되었기 때문에 상당히 엄격한 수준의 filtering condition을 보인다고 할 수 있다. 그러나 low-sequence depth 위주의 sequence 데이터에서 false positive variant의 발생을 최소화하기 위한 조치였기 때문에 이에 대한 조건을 완화시키기는 어렵다. 따라서 기존의 결과보다 더욱 신뢰성이 있는 variant calling을 위해서는 기존에 배추수집단에 대하여 조사한 표현형을 기준으로 유용 표현형이 이루는 분포의 양 극단에 속 하며 특히 낮은 sequencing depth를 보이는 계통위주로 추가적인 sequencing 작업이 필요할 것으로 사료된다.

다. 파이프라인에 의해 구성되는 SNP matrix 내의 marker loci 개수에 영향을 미치는 요인의 확인

NGS 기반 플랫폼의 Sequencing을 통해 얻은 계통의 Whole genome sequence로부터 reference genome에 대해 variant calling을 수행하는 전 과정은 리눅스 서버 환경에서 command line을 통해 구성되며 각 command는 특정 소프트웨어를 활용하고 각 소프트웨어는 여러 가지 패러미터를 지정 받아 작업을 수행한다. 결국 각 단계의 패러미터의 조합이 누적되어 최종적인 SNP matrix 내에서의 marker loci의 총 개수에 영향을 미치게 되는 것을 상정할 수 있다. 본 작업에서는 조사 계통의 수, reference 버전, 편집이 완료된 bam file을 input으로 Variant calling을 실질적으로 수행하는 소프트웨어, minimum depth, missing rate, minimum allele frequency를 구축한 파이프라인에서 SNP matrix의 marker loci 개수에 영향을 미치는 요인으로 선정하고 각 요인의 조합별로 SNP matrix를 생산하여 어떤 요인이 가장 크게 SNP matrix 구성 결과에 영향을 미치는 지를 확인하고자 하였다 (표 6).

표 6. SNP matrix의 구성에 영향을 미치는 요인과 변수의 일람

| Variables | Case1 | Case2 | Case3 |
|--------------------|----------|-----------------|-------|
| Accession number | 145 | 199 | - |
| Reference | V1.5 | V3.0 | - |
| SNP calling method | Samtools | HaplotypeCaller | - |
| Minimum Depth | 3 | 5 | 10 |
| Missing rate | 10% | 20% | 30% |
| MAF | 0.01 | 0.05 | - |

현재 HaplotypeCaller를 활용한 variant calling은 아직 이루어지지 않아 총 72개 경우(2 X 2 X 1 X 3 X 3 X 2) 를 상정하여 variant calling을 독립적으로 진행하였으며 하나의 reference genome과 선택된 계통수 이하의 경우에 대한 조건은 표 x와 같다. 선택 계통의 수와 reference genome의 조합이 총 4개이므로 표 3에서 나타나는 조합의 수를 곱하면 상정한 72개의 경우가 나타나게 된다.

표 7. minimum depth, missing rate, 그리고 MAF를 통해 조합 가능한 경우에 group을 부여한 결과

| Depth supporting | specific SNP position | Missing rate in one SNP position | Minimum allele frequency | Group |
|------------------|-----------------------|----------------------------------|--------------------------|-------------|
| MinDepth 3 | | | 0.1 | 0.01 Case1 |
| | | | 0.05 Case2 | |
| | | | 0.2 | 0.01 Case3 |
| | | | 0.05 Case4 | |
| | | | 0.3 | 0.01 Case5 |
| | | | 0.05 Case6 | |
| MinDepth 5 | | | 0.1 | 0.01 Case7 |
| | | | 0.05 Case8 | |
| | | | 0.2 | 0.01 Case9 |
| | | | 0.05 Case10 | |
| | | | 0.3 | 0.01 Case11 |
| | | | 0.05 Case12 | |
| MinDepth 10 | | | 0.1 | 0.01 Case13 |
| | | | 0.05 Case14 | |
| | | | 0.2 | 0.01 Case15 |
| | | | 0.05 Case16 | |
| | | | 0.3 | 0.01 Case17 |
| | | | 0.05 Case18 | |

이와 같이 총 72개의 경우를 상징한 후 독립적으로 SNP matrix를 생성하고 각 SNP matrix에서 나타난 SNP marker loci의 개수의 분포는 그림 13과 같다.

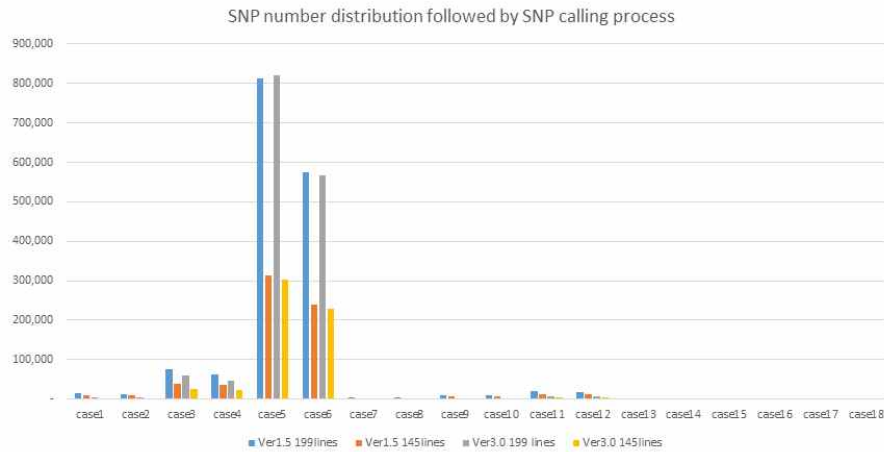


그림 13. 상징한 72개 경우에 따른 SNP matrix의 marker loci 개수의 분포

reference genome의 버전에 관계없이 199개 계통을 대상으로 minimum depth를 3으로 허용하고 missing rate 30%이하, MAF 0.01 이상일 때 최대의 SNP marker loci가 나타나는 것으로 확인되었다. 145개 계통을 사용한 모든 경우는 Pekinensis에 속하는 계통들만 반영되어 chinensis와 같은 다른 아종이 포함되는 경우(199 계통 이용)보다 유전적인 다양성이 떨어져 상대적으로 낮은 개수의 SNP marker loci가 나타나는 것으로 설명할 수 있다. Depth 5와 10 기준의 조건인 경우 SNP matrix 내의 marker loci가 10,000개 이하로 나타나 GWAS 분석과 같은 전장 유전체 상의 변이를 대상으로 하는 유전분석에는 부적합한 것으로 사료된다. 반면 case 5와 case 6와 같은 minimum depth 3 기반 조건의 SNP matrix의 경우, 배추 유전체 전반에 분포하는 SNP marker loci를 보이지만 낮은 depth로 인한 신뢰성이 낮은 SNP의 비중이 큰 것이 추후 분석의 신뢰성에 우려를 미치는 점이라 할 수 있다.

라. 최종 산출되는 SNP 개수에 영향을 미치는 조합에 따른 SNP matrix의 신뢰성 평가

SNP matrix의 구성이후 산출된 matrix 상의 SNP loci는 reference genome 상의 물리적인 위치에 대응된다. 따라서 각 loci가 reference genome상에 예측된 유전자 coding 지역 상의 존재 유무를 SNP annotation을 통해 검증이 가능하다. 뿐만 아니라 작물 지놈의 중복 서열이 갖는 특성 때문에 sequencing 과정에서 fastq 파일을 생산할 때 이러한 영역에서 유래된 read sequence가 파일내 에서 큰 비중을 차지할 가능성이 있다. 이러한 read sequence가 read alignment 과정에서 원래 유래한 영역에 mapping 되지 않으면 mapping 된 위치에서 heterozygous variant로 나타나는 경우를 상정할 수 있다. 따라서 본 작업에서는 72개 조건에 따라 구성된 SNP matrix를 SNP marker loci의 genic/intergenic 유래 여부, heterozygous rate 두 측면에서 관찰하여 유전분석에 활용할 SNP matrix의 신뢰성 수준을 평가하고자 한다(그림 8). SNP의 genic/intergenic 영역 유래의 판별은 계통별 변이 정보를 통합한 VCF 파일을 input으로 하여 snpEFF 소프트웨어를 활용하여 수행하였다. snpEFF 소프트웨어는 배추의 reference genome에 포함된 정보인 크로모솜 별 염기서열과 그 위에 예측된 유전자의 위치를 기준으로 SNP matrix 상의 각 position에 대한 annotation을 수행한다. 또한 SNP matrix 상의 각 marker loci의 genomic position에 대해 각 계통별로 그 위치에 해당하는 genotype을 reference와 동일, alternative allele, missing allele로 구분하고 이에 대한 통계치를 계산하는 perl script를 자체적으로 작성하여 SNP matrix 내에서의 모든 SNP marker loci에 대한 heterozygous rate를 측정하였다.

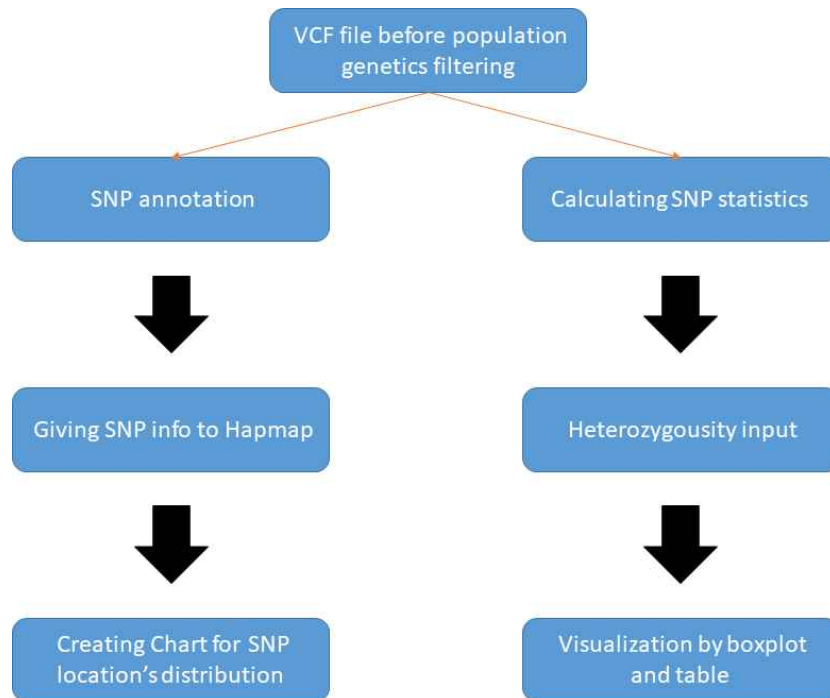


그림 14. SNP matrix의 특성 평가(SNP annotation, heterozygosity calculation)의 도식화

모든 조건에 대하여 구성된 SNP matrix 내의 SNP에 부여된 정보가 이루는 genic/intergenic 유래에 대한 정보는 matrix 내의 SNP를 비율로 표현하여 bar plot으로 나타내었으며 SNP들이 각 loci에서 계통별 genotype으로부터 계산된 heterozygosity의 통계치는 box plot의 형태로 표현하였다(그림 15-18). 이와 같은 데이터 시각화는 R의 ggplot2 패키지를 활용하여 수행되었다.

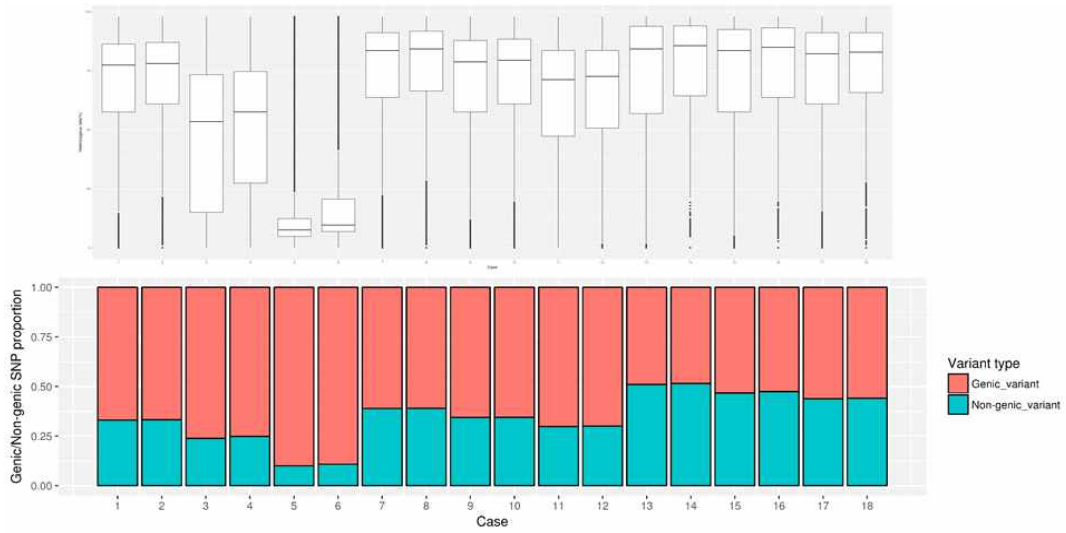


그림 15. Reference 버전 1.5 기반 145개 계통의 SNP matrix의 heterozygosity, genic 유래 SNP 비율의 평가

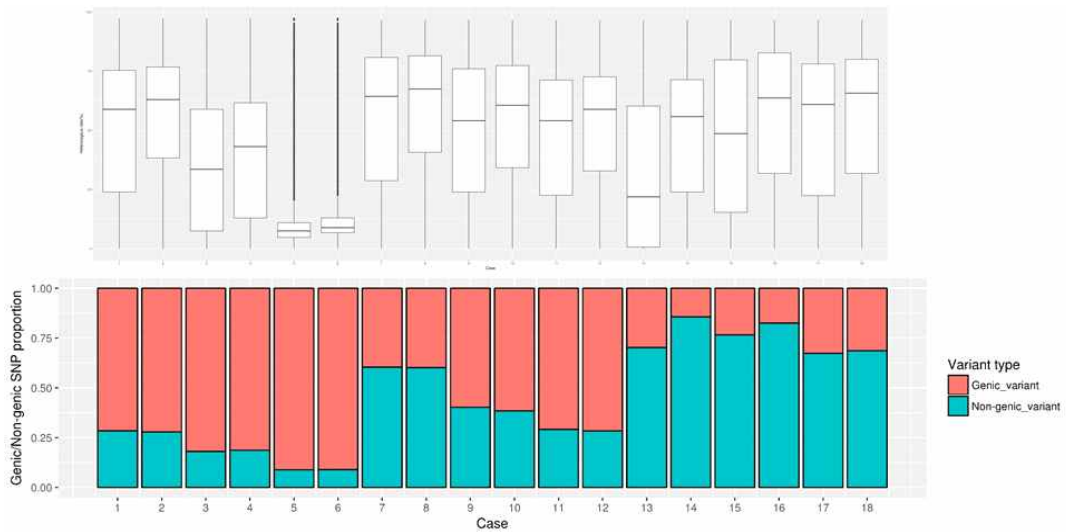


그림 16. Reference 버전 3.0 기반 145개 계통의 SNP matrix의 heterozygosity, genic 유래 SNP 비율의 평가

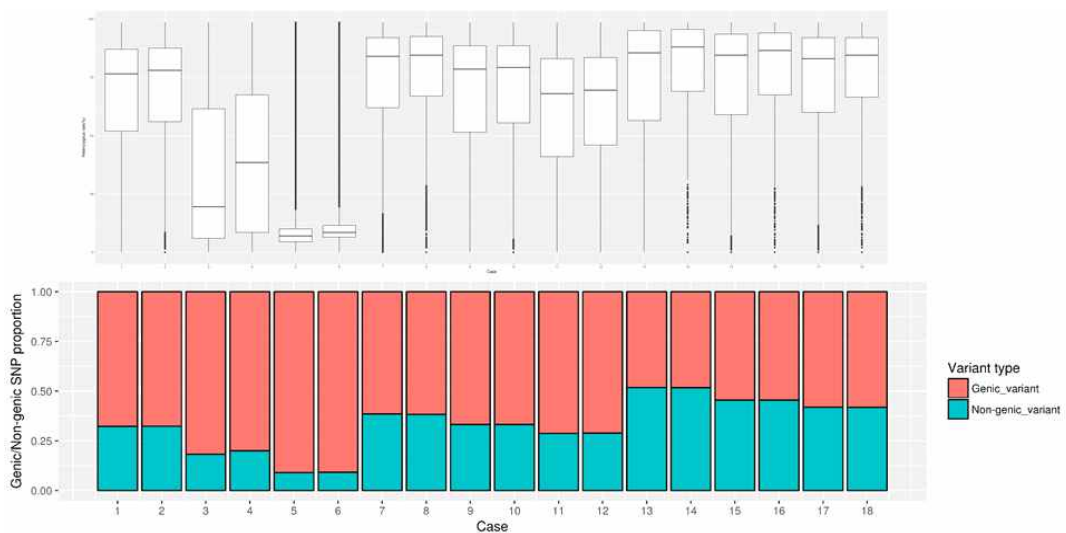


그림 17. Reference 버전 1.5 기반 199개 계통의 SNP matrix의 heterozygosity, genic 유래 SNP 비율의 평가

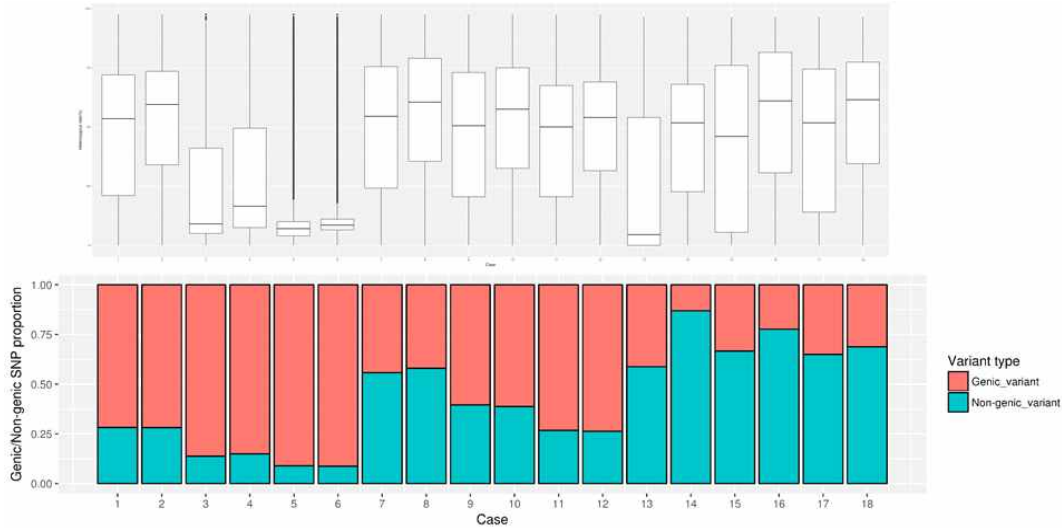


그림 18. Reference 버전 3.0 기반 199개 계통의 SNP matrix의 heterozygosity, genic 유래 SNP 비율의 평가

reference와 조사한 배추 계통의 수를(2 X 2) 반영한 4가지 경우를 바탕으로 각 경우에 해당하는 case들에 대해 주어진 2개의 특성을 조사하였을 때 reference가 1.5에서 3.0으로 바뀌었을 때 현저한 특성치의 차이가 나타났다. 특히 minimum depth를 10으로 통제된 경우에서 reference genome의 intergenic 영역에서 유래한 SNP의 비중이 크게 증가하였는데 이는 실제 intergenic 영역에서 유래한 sequence read들이 1.5 버전에서는 genic 영역에 mapping이 되었다가 reference genome 3.0 버전에서 해당 sequence read에 해당하는 영역이 반영됨으로서 이전 버전에서 잘못 mapping된 read들이 원래의 위치로 mapping 된 것으로 해석할 수 있다. heterozygosity의 경우 버전이 높은 경우에서 boxplot이 보이는 분산이 넓어진 것을 확인할 수 있었다. 이 또한 업데이트된 reference genome이 기존에 커버하지 못한 유전적 영역을 새로 반영하면서 sequence alignment 과정에서 기존에 잘못 mapping되어 heterozygous allele SNP로 보이게 한 read sequence들이 실제 유래한 유전적 영역에 mapping이 되어 heterozygosity가 variant calling 사용한 reference 버전이 높아짐에 따라 줄어든 것으로 해석이 가능하다.

5. 신규 표준 유전체 기반으로 한 배추 수집단 154계통의 잎털 형질의 유전적 구조 분석

배추는 한 작물에 대하여 표준 유전체의 개선이 지속적으로 이루어진 대표적인 예이다. 따라서 작물이 가지는 다양한 형질에 대한 유전분석은 분석방식의 변화뿐 만이 아니라 그 바탕이 되는 표준 유전체의 변화 또한 지속적으로 반영하고 새로 수행이 되어야 할 것이다. 이하의 연구는 배추의 잎털(Trichome)에 대하여 GWAS(Genome-Wide Association Study)분석을 수행한 것으로 이를 위한 다년간의 표현형 데이터(Phenotype)의 작성과 신규 표준 유전체 서열을 반영한 NGS 기반의 유전자형(Genotype)의 데이터 생산이 수반되었다.

가. 배추 수집단 154계통에 대한 잎털 형질의 조사

배추 잎털 형질의 표현형 조사를 위하여 배추 수집단에 포함된 154계통을 충남대학교 시험 포장에서 2016년 2017년에 걸쳐 재배하였다. 선택한 배추 154개 계통은 국내 수집 및 육성 계통뿐만 아니라 중국, 일본, 영국, 그리고 네덜란드에서 도입된 계통을 포함한다. 해당 집단을 이루는 대다수의 계통은 ssp *pekinensis*에 해당하지만 outgroup으로서 비결구성 배추에 해당하는 ssp. *chinensis*, ssp

parachinensis, *ssp perviridis*, *ssp trilocularis*, *ssp niposinica*, 그리고 *ssp. rapa* 또한 함께 포장에서의 재배 및 표현형 평가를 수행하였다.

(1) 2016년 표현형 검정

배추 154개 품종을 가을 작기에 실험 포장에서 재배하여 잎의 앞뒤 표면 및 엽맥과 엽연에 잎털이 존재하는 지를 확인하였다. 또한 3반복의 잎을 샘플링 하고 5cm²의 면적내에 잎털의 발생 빈도에 인덱싱을 부여하여 표현형 데이터를 구성하였다(0= No trichome, 1= 1~5 trichome in the 5cm², 2 = over 20 trichomes in the 5cm², 3 = over 40 trichomes in the 5cm², 4= over 70 trichomes in the 5cm²).

(2) 2017년 표현형 검정

전년도와 동일한 154개 품종을 가을 작기에 실험 포장에서 재배하였으며 3개의 반복을 배추의 결구 형성 시기 이전에 성숙한 외엽을 대상으로 샘플링하였다. 잎털의 발생유무는 잎의 앞뒷면, 중륜, 그리고 엽연에서 확인하였으며 표현형을 촬영한 사진은 포토샵 CS6를 통해 편집하였으며 편집한 사진을 기반으로 주어진 반복 상의 트리콤의 발생 개수를 counting 하였다. 또한 stereomicroscope를 활용하여 잎털의 물리적 구조를 관찰하였다.

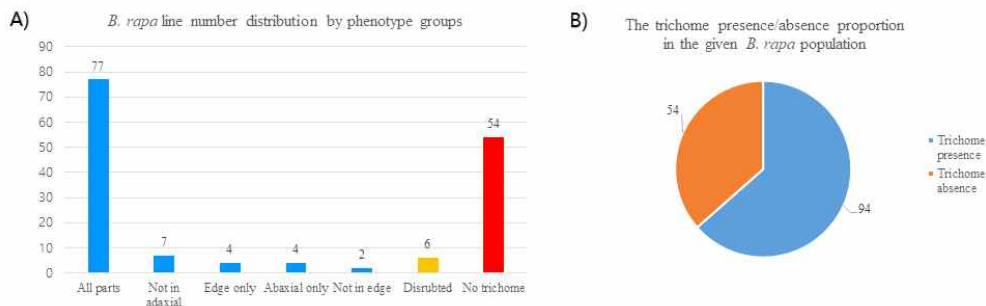


그림 19. 배추 수집단 154 계통의 잎털 표현형 조사 결과. A) 배추 잎털의 생성 유형에 따른 분류에 대한 계통의 분포 B) 잎털 유무에 따른 배추 수집단 계통의 분류 (반복내에서 상이한 표현형을 보이는 6개 계통 제외)

나. NGS 데이터를 활용한 배추 154계통의 genotyping

배추 154개 계통의 sequencing 데이터는 기존에 구성한 배추 수집단의 Illumina paired end sequencing 데이터에서 연구에 필요한 계통에 대응되는 ID를 기준으로 선별하였다. 선별한 sequence 파일은 과거 연차의 작업을 통해 adapter 서열의 제거 및 low-quality read의 제거가 이미 이루어진 상태이다. 각 계통의 sequencing 파일은 새로 공개된 배추의 reference genome 3.0 버전에 대하여 새로 alignment가 이루어 졌으며 기존의 구축한 방식과 마찬가지로 MarkDuplicates, IndelRealignment, AddOrReplaceReadGroup이 수행되었으나 추가적으로 reference genome에 alignment된 paired end read의 insert size가 1000bp 이내로 mapping된 read 만을 Vcftool을 이용한 variant calling에 반영하였다. 선정된 모든 계통에 대하여 개별적인 SNP calling 이후 depth 5 이상 homozygous alternative allele만을 filtering 하고 이 SNP position의 위치를 취합함으로써 reference genome 상에 나타날 수 있는 모든 가능한 변이 위치에 대한 리스트를 작성하였다. 이 unique SNP position에 대한 리스트와 계통별로 편집이 완료된 bam file을 input으로 하여 samtool과 bcftool을 파이프로 연결한 명령어를 통해 joint variant calling을 수행하였다. 이와 같은 방식으로 얻은 배추 154개 계통이 이루는 집단에

대한 1차 VCF file은 minimum depth를 3으로 허용하고 multi-allelic SNP position 제거 및 variant quality ≥ 30 조건으로 필터링한 후 PLINK software를 통해 SNP position의 missing rate 30% 이하, minimum allele frequency 0.05, heterozygosity 20% 이하의 조건으로 최종 선별과정을 거쳐 274,686개의 SNP position을 얻을 수 있었다 (표 8).

표 8. 배추 수집단의 SNP matrix 구성 과정에서의 필터링 조건에 따른 SNP marker 개수의 변화

| | Properties | Statistics |
|--|---|------------|
| | Initial putative SNPs on the primary matrix | 5,208,748 |
| | Bi-allelic only variants | 5,052,372 |
| SNP number fluctuation through filtering process | Filtering process by missing rate over 30% for all SNP positions | 525,294 |
| | Filtering process by MAF under 0.05 for all SNP positions | 322150 |
| | Filtering process by heterozygosity under 20% for all SNP positions | 274,686 |

다. 배추 수집단 잎털 형질에 대한 유전분석

전술한 조건으로부터 산출한 SNP matrix를 활용하여 SNPhylo를 통한 phylogeny 분석 및 조사된 표현형과 결합하여 GAPIT 기반의 GWAS 분석을 수행하였다. 또한 전술한 조건에서 minimum depth를 5로 상향 조정하여 979개의 SNP position을 갖는 matrix를 구성하여 Strauto software를 통해 STRUCTURE 분석을 수행하였다.

(1) 배추 수집단 154계통의 SNP 데이터 기반의 계통도 작성

SNPhylo software는 입력된 모든 SNP position을 대상으로 계통간 서열 비교를 통해 계통도를 작성하는 것이 아니라 주어진 SNP matrix 상에서 LD(Linkage disequilibrium)를 식별하여 해당 영역의 변이를 대상으로 계통간 유전적 거리를 측정하여 Phylogenetic tree를 작성한다. 작업의 종료 이후 자체적으로 작성한 계통도의 png 파일 이외에 newick 포맷으로 구성된 분석결과를 생성한다. 해당 결과를 FigTree (ver1.4.3)에 입력하여 root format으로 설정 후 계통도를 시각적으로 수정하였다(그림 20).

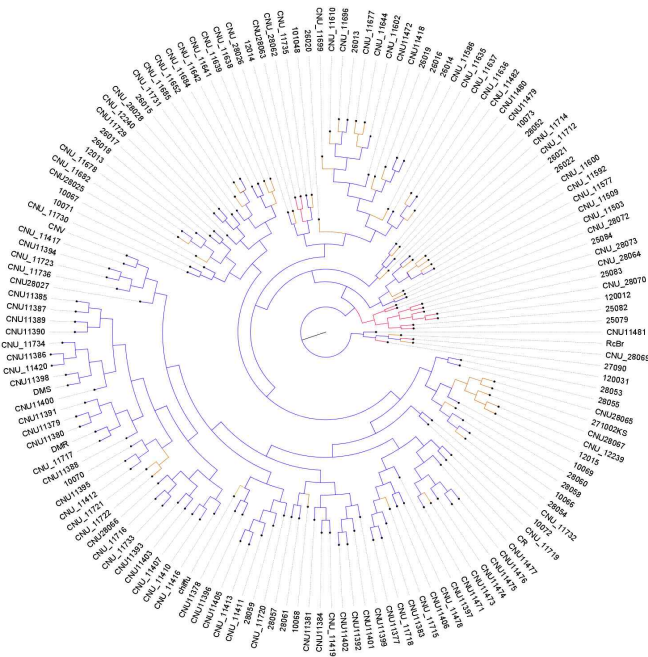


그림 20. 배추의 잎털 형질의 유전분석을 위해 선정된 배추 154 계통의 계통도(Phylogenetic tree) (Red node: Non-pekinensis, Yellow node: Non-trichome pekinensis, Blue node: Normal pekinensis)

계통도의 root에서부터 제일 초기에 분지된 subgroup은 pekinensis 계통과 non-pekinensis 계통이 혼재되어 있지만 전반적으로 잎털이 없는 형질을 보이며 그 다음 분지에서 하나의 subgroup에 대다수의 non-pekinensis 계통들이 clustering 되었다. 그 외에 두 개의 계통이 (CNU_28062, CNU28063) 이 pekinensis의 subgroup에 속해있는 것을 확인할 수 있었으나 이는 해당 계통의 높은 missing rate가 기인하여 나타난 결과라 가정할 수 있다. 계통도에서 전반적으로 보이는 추세를 살펴보면 non-pekinensis group이 계통도 내에 outgroup의 형태로 나타나고 이후 pekinensis group에 속하는 계통들이 출현하였으나 그 대다수가 잎털이 없는 계통이 우세하게 나타났다. 이후 분화가 진행되면서 잎털을 가진 pekinensis의 계통이 점차 증가하게 되고 과반을 차지하는 형세를 보였다. 이를 통해 배추가 처음에는 non-pekinensis 계통에서 분화하여 육종 및 선발과정에서 잎털의 생성 능력을 획득한 것이라는 가설을 세울 수 있다.

(2) STRUCTURE 분석을 통한 배추 수집단 154계통의 유전적 구조의 식별

Joint variant calling을 통해 구성한 154계통의 통합 VCF file 내의 Minimum Depth가 5 이하인 SNP는 missing으로 처리하고 PLINK software를 통해 missing rate 30%, MAF 0.05, 그리고 heterozygosity 20%로 데이터를 통제하여 979개의 SNP position을 얻었다. STRUCTURE 분석은 strauto software를 통해 MCMC = 10,000, burnin=10000 조건으로 subpopulation의 수는 1부터 10까지 가정하였으며 각 subpopulation의 연산은 단계별로 20회 반복하였다. STRUCTURE 연산 결과는 structureHarvester.py를 통해 Delta K를 산출하고 Structure 2.3.4 software를 통해 Q matrix 별 sorting 하여 barplot의 형태로 시각화 하였다. 집단은 크게 두 개의 subpopulation으로 나눌 수 있지만 subpopulation의 수를 가정하는 peak가 K=4와 K=8에서 각각 minor하게 나타났다(그림 21).

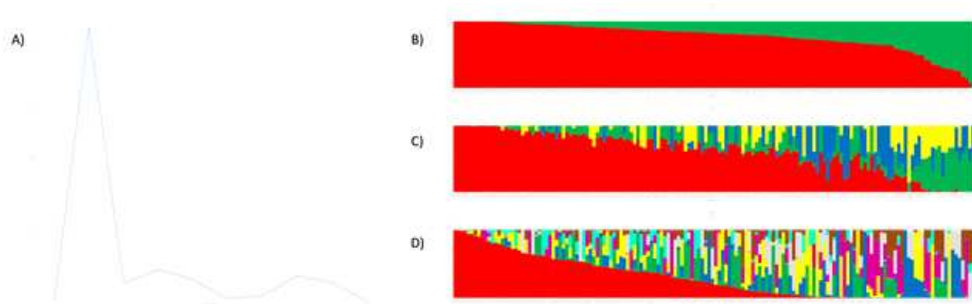


그림 21. 154계통의 939 SNP position으로부터 산출된 STRUCTURE 결과의 시각화. A) 1부터 10까지 가정된 집단내의 subpopulation에 대한 Delta K, B) K=2로 가정되었을 경우의 집단의 Structure, C) K=4로 가정되었을 경우의 집단의 Structure, D) K=8로 가정되었을 경우의 집단의 Structure

(3) 배추 잎의 모용 형질에 대한 GWAS 분석

배추 154 계통의 genotyping 작업으로 산출한 274,686개의 SNP marker position을 갖는 SNP matrix에 잎털의 형질 조사 결과를 GAPIT 기반의 GWAS 분석을 통해 산출된 SNP와 조사 표현형 간의 association을 검정하였다.

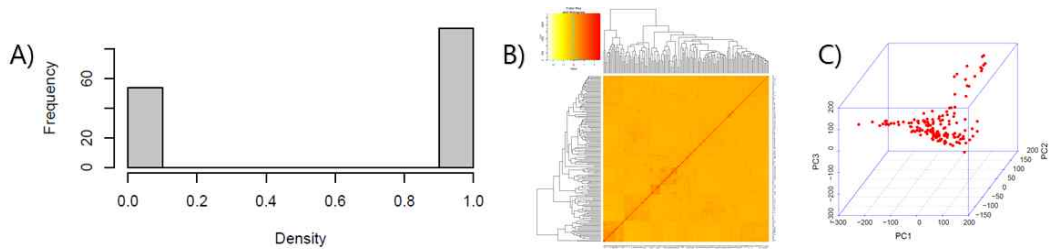


그림 22. A) 배추 154개 계통에서 나타난 잎털 표현형의 질적 분포. B) 배추 154개 계통에 대한 친족 플롯(Kinship matrix), C) 배추 154개 계통별 유전자형에 대한 PCA plot

GAPIT은 GWAS 분석에 앞서 입력된 유전자형을 수치화 한 이후 표현형과 함께 일차적으로 입력된 정보를 우선적으로 분석하여 그림 x의 예시와 같이 조사 집단의 표현형적/유전형적 특성을 요약하는 시각적 자료를 구성한다. 집단 내 계통의 일부가 하나의 계통에서 소포자 배양을 통해 생산된 계통들로 이루어지고 대다수가 pekinensis로 분류되는 계통인 관계로 친족 플롯에서 명확히 구분되는 복수의 블록이 나타나지는 않았으나 친족 플롯의 좌측 하단의 블록 및 3D PCA의 결과를 확인하였을 때 non-pekinensis에 해당하는 계통들이 명확하게 분류되는 것을 확인되었다(그림 22).

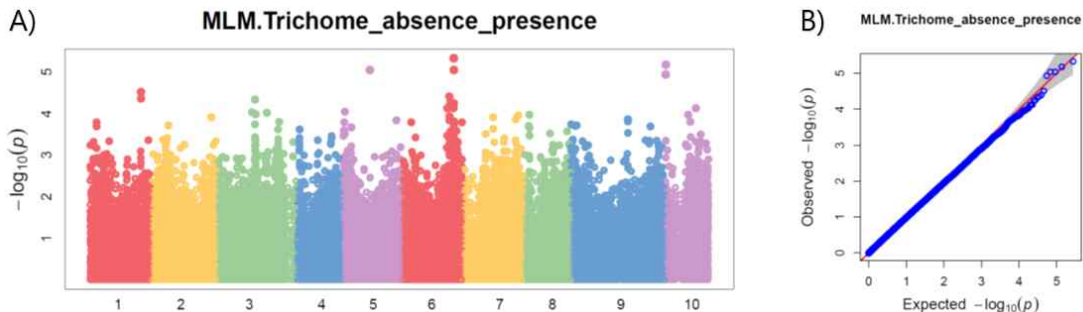


그림 23. 배추 154개 계통의 잎털에 대한 genome-wide association study의 시각적 결과. A) Manhattan plot, B) QQ plot

GAPIT은 compressed Mixed Linear Model(cMLM)을 통계모델로 선정하여, 입력된 계통별 표현형과 SNP로 대변되는 유전자형의 상관관계를 찾고 이에 대한 유의성을 p-value로 산출한다. 이러한 P-value는 로그를 취한 값에 -1을 곱함으로써 양의 실수를 갖게 되고 이는 각 SNP의 포지션 별로 전환된 P-value가 플롯 상에 크로모솜 별로 구분되어 나타나게 된다. SNP matrix내에 마커로 활용이 가능한 SNP position의 개수를 고려하였을 때 맨하탄 플롯의 유의할 만한 SNP position의 기준으로 SNP position 수의 역수에 로그를 취한 값에 -1을 곱한 값을 사용 가능하다. 분석에 사용한 SNP matrix의 경우, 5.43을 그 기준치로 취할 수 있으나 manhattan plot에서 SNP position이 나타내는 변환값의 최대치는 5.33으로 기준치를 넘지 않았다. 또한 전체 SNP가 나타내는 p-value의 분포가 확률적으로 기대되는 그 분포와 크게 다르지 않아 QQ plot에서 점이 붉은 기준선 상에 위치하는 것을 확인할 수 있는데 이 역시 유의성 있게 나타나는 SNP의 수가 분포의 정규성을 무너뜨릴 수준으로 높게 나타나지 않은 것을 의미한다. 단, 크로모솜 6번에서 기준치에 근접하고 다른 크로모솜의 경우에 비해 안정적으로 나타나는 peak가 관찰되었다(그림 24). 따라서 이 peak내에 예측된 유전자들의 리스트를 작성 및 그 기능에 대한 정보를 수집을 수행하여 peak내에 해당하는 유전적 영역의 특성이 배추의 잎털 생성에 대해 연관성이 있는지를 파악할 필요가 있다. 따라서 manhattan plot에서 나타난 상위 100개의 SNP position 중에서 6번째 크로모솜에서 유래한 SNP만을 우선 조사 대상으로 한정하여 SNP position 정보를 기준으로 SNP annotation 결과를 대조하였다(표 9).

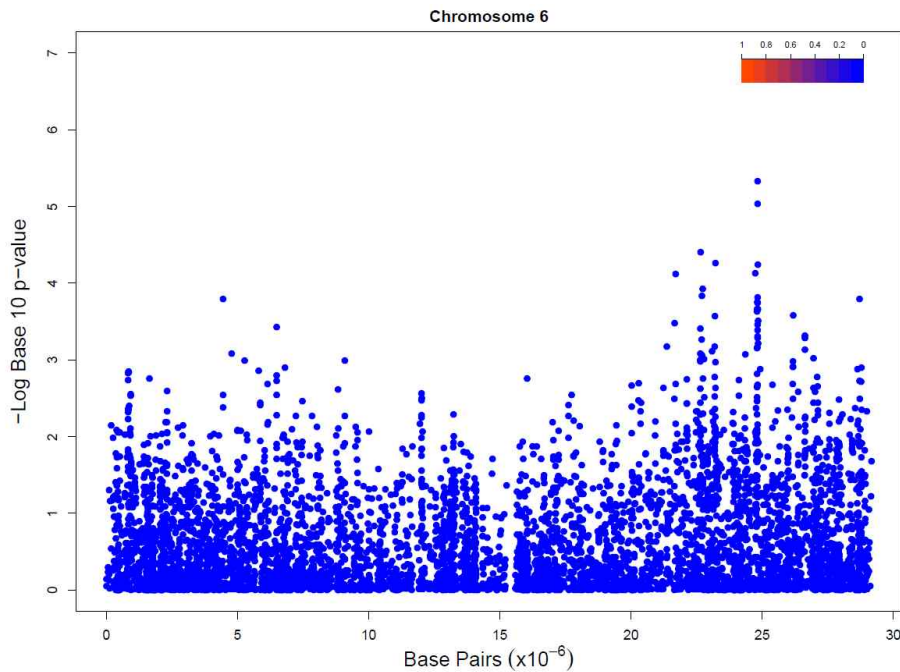


그림 24. 배추 154 계통의 잎털 발현 형질에 대한 배추 표준 유전체의 6번 크로모솜의 맨하탄 플롯

표 9. 배추 표준 유전체 6번 크로모솨 상의 잎털 형질 발현에 대한 잠정적인 유전적 영역 상의 SNP

| Rank | Chr | Position | REF | ALT | GeneID in <i>B. rapa</i> reference ver 3.0 | SNP Annotation | feature type | P value | Dot value | MAF |
|------|-----|----------|-----|-----|--|--------------------------------------|-------------------|-------------|-----------|------|
| 1 | A06 | 24825546 | T | G | BraA06g036880.3C | upstream_gene_variant | transcript | 4.66E-06 | 5.33 | 0.47 |
| 2 | A06 | 24826774 | C | T | BraA06g036880.3C | upstream_gene_variant | transcript | 9.16E-06 | 5.04 | 0.47 |
| 3 | A06 | 22655242 | G | T | BraA06g032800.3C | upstream_gene_variant | transcript | 3.94E-05 | 4.4 | 0.44 |
| 4 | A06 | 23225605 | A | G | BraA06g033810.3C | downstream_gene_variant | transcript | 5.50E-05 | 4.26 | 0.42 |
| 5 | A06 | 24839868 | G | C | BraA06g036890.3C | upstream_gene_variant | transcript | 5.73E-05 | 4.24 | 0.49 |
| 6 | A06 | 24735263 | T | C | BraA06g036640.3C | missense_variant | transcript | 7.46E-05 | 4.13 | 0.24 |
| 7 | A06 | 21704440 | T | A | BraA06g031210.3C | downstream_gene_variant | transcript | 7.54E-05 | 4.12 | 0.48 |
| 8 | A06 | 22734843 | T | G | BraA06g032940.3C | upstream_gene_variant | transcript | 0.000117834 | 3.93 | 0.31 |
| 9 | A06 | 22703525 | T | A | BraA06g032880.3C | upstream_gene_variant | transcript | 0.00014714 | 3.83 | 0.48 |
| 10 | A06 | 24826392 | G | T | BraA06g036880.3C | upstream_gene_variant | transcript | 0.000152366 | 3.82 | 0.42 |
| 11 | A06 | 28716804 | G | A | BraA06g043760.3C | upstream_gene_variant | transcript | 0.000161146 | 3.79 | 0.34 |
| 12 | A06 | 4450692 | C | A | BraA06g007970.3C | splice_region_variant&intron_variant | transcript | 0.000161577 | 3.79 | 0.34 |
| 13 | A06 | 24823493 | T | G | BraA06g036870.3C | synonymous_variant | transcript | 0.000178815 | 3.75 | 0.48 |
| 14 | A06 | 24826505 | C | G | BraA06g036880.3C | upstream_gene_variant | transcript | 0.000182539 | 3.74 | 0.48 |
| 15 | A06 | 24808155 | C | T | BraA06g036850.3C | upstream_gene_variant | transcript | 0.000216105 | 3.67 | 0.34 |
| 16 | A06 | 24839924 | C | T | BraA06g036910.3C | upstream_gene_variant | transcript | 0.000217769 | 3.66 | 0.5 |
| 17 | A06 | 24806178 | T | C | BraA06g036840.3C | downstream_gene_variant | transcript | 0.000231923 | 3.63 | 0.31 |
| 18 | A06 | 26182689 | G | T | BraA06g038950.3C-BraA06g038960.3C | intergenic_region | intergenic_region | 0.000264078 | 3.58 | 0.16 |
| 19 | A06 | 23206589 | C | A | BraA06g033800.3C | downstream_gene_variant | transcript | 0.000267202 | 3.57 | 0.38 |
| 20 | A06 | 24850600 | A | G | BraA06g036920.3C | missense_variant | transcript | 0.000309598 | 3.51 | 0.38 |
| 21 | A06 | 21665152 | T | C | BraA06g031160.3C | downstream_gene_variant | transcript | 0.000335323 | 3.47 | 0.21 |
| 22 | A06 | 24826184 | A | G | BraA06g036880.3C | upstream_gene_variant | transcript | 0.000345626 | 3.46 | 0.47 |
| 23 | A06 | 6493707 | A | T | BraA06g011870.3C | upstream_gene_variant | transcript | 0.000375764 | 3.43 | 0.34 |
| 24 | A06 | 22646160 | A | G | BraA06g032790.3C | synonymous_variant | transcript | 0.000391754 | 3.41 | 0.35 |
| 25 | A06 | 24827071 | C | G | BraA06g036880.3C | upstream_gene_variant | transcript | 0.00040801 | 3.39 | 0.4 |
| 26 | A06 | 26635777 | A | G | BraA06g039750.3C | upstream_gene_variant | transcript | 0.000479873 | 3.32 | 0.36 |

(4) GWAS 분석에서 나타난 잎털 형질 관련 유전적 영역상의 candidate gene의 발굴

그림 18에서 확인된 배추 잎털 형질에 연관된 SNP peak의 범위를 확인한 결과, 표 10의 SNP 그룹은 배추 reference genome ver 3.0 기준으로 6번 크로모솨의 24,735,263-24,850,600 bp 범위에 위치하는 것을 확인하였다. 그리고 해당 범위에서 예측된 모든 유전자들의 단백질 서열을 확보하고 이를 애기장대 표준 유전자의 단백질 데이터베이스에 local blastp를 이용하여 peak내의 배추 유전자들에 대한 애기장대에서의 orthologue 유전자를 찾았다. 이를 통해 해당 peak 내의 범위에서 총 29개의 유전자가 candidate gene으로 식별이 되었다. 이들 중 BraA06g036920.3C는 애기장대의 잎털 형성에 관여하는 것으로 알려진 GL1의 orthologue gene인 것으로 나타났다(표 10).

표 10. 배추 잎털 연관 유전적 영역상의 candidate gene의 정보

| No. | Chr* | Start | End | Dir* | Gene ID | TAIR orthologue | Other name | Gene description |
|-----|------|----------|----------|------|------------------|-----------------|---|---|
| 1 | A06 | 24733561 | 24736121 | - | BraA06g036640.3C | AT3G28460.1 | NA | methyltransferase;(source:Araport11) |
| 2 | A06 | 24736566 | 24738128 | + | BraA06g036650.3C | AT3G28455.1 | CLAVATA3/ESR-RELATED 25, CLE25 | Member of a large family of putative ligands homologous to the Clavata3 gene. Consists of a single exon. Can not replace CLV3 function in vivo.CLE25 participates in long distance signaling in response to dehydration. It produces a graft transmissible signal from root to shoot that induces ABA synthesis and results in stomatal closure. The BAM1 and BAM3 receptor-kinases are likely receptors for CLE25 as they are required for this signaling. |
| 3 | A06 | 24739465 | 24740967 | - | BraA06g036660.3C | AT3G28450.1 | BAK1-INTERACTING RECEPTOR-LIKE KINASE 2, BIR2 | Leucine-rich repeat protein kinase family protein;(source:Araport11) |
| 4 | A06 | 24742450 | 24743871 | - | BraA06g036670.3C | AT1G19010.1 | NA | hypothetical protein;(source:Araport11) |
| 5 | A06 | 24752571 | 24754043 | - | BraA06g036680.3C | AT3G28420.1 | NA | Putative membrane lipoprotein;(source:Araport11) |
| 6 | A06 | 24757010 | 24758694 | + | BraA06g036690.3C | AT3G28390.1 | ABCB18, ATP-BINDING CASSETTE B18, P-GLYCOPROTEIN 18, PGP18 | P-glycoprotein 18;(source:Araport11) |
| 7 | A06 | 24759982 | 24761559 | - | BraA06g036700.3C | AT3G28370.1 | NA | spindle assembly checkpoint component;(source:Araport11) |
| 8 | A06 | 24764528 | 24764968 | - | BraA06g036710.3C | AT3G28345.1 | ABCB15, ATP-BINDING CASSETTE B15, MDR13, MULTI-DRUG RESISTANCE 13 | Encodes an ATP-binding cassette (ABC) transporter. Expressed in the vascular tissue of primary stem. |
| 9 | A06 | 24766261 | 24768204 | + | BraA06g036720.3C | AT3G28360.1 | ABCB16, ATP-BINDING CASSETTE B16, P-GLYCOPROTEIN 16, PGP16 | P-glycoprotein 16;(source:Araport11) |
| 10 | A06 | 24768674 | 24769812 | + | BraA06g036730.3C | AT3G28415.1 | ABCB22, ATP-BINDING CASSETTE B22 | ABC transporter family protein;(source:Araport11) |
| 11 | A06 | 24770114 | 24771833 | + | BraA06g036740.3C | N/A | N/A | N/A |
| 12 | A06 | 24773512 | 24773754 | + | BraA06g036750.3C | AT3G28415.1 | ABCB22, ATP-BINDING CASSETTE B22 | ABC transporter family protein;(source:Araport11) |
| 13 | A06 | 24775162 | 24776946 | - | BraA06g036760.3C | AT3G28370.1 | NA | spindle assembly checkpoint component;(source:Araport11) |
| 14 | A06 | 24778419 | 24780247 | + | BraA06g036770.3C | N/A | N/A | N/A |
| 15 | A06 | 24782136 | 24782729 | + | BraA06g036780.3C | AT3G28345.1 | ABCB15, ATP-BINDING CASSETTE B15, MDR13, MULTI-DRUG RESISTANCE 13 | Encodes an ATP-binding cassette (ABC) transporter. Expressed in the vascular tissue of primary stem. |
| 16 | A06 | 24784677 | 24789008 | + | BraA06g036790.3C | AT3G28340.1 | GALACTINOL SYNTHASE 8, GALACTURONOSYLTRANSFERASE-LIKE 10, GATL10, GOLS8 | Encodes a protein with putative galacturonosyltransferase activity. |
| 17 | A06 | 24789376 | 24791934 | - | BraA06g036800.3C | AT3G28210.1 | PMZ, SAP12, STRESS-ASSOCIATED PROTEIN 12 | Encodes a putative zinc finger protein (PMZ). |
| 18 | A06 | 24793020 | 24793800 | + | BraA06g036810.3C | N/A | N/A | N/A |

| | | | | | | | | |
|----|-----|----------|----------|---|------------------|-------------|--|--|
| 19 | A06 | 24794927 | 24799794 | + | BraA06g036820.3C | AT3G28180.1 | ATCSLC04, ATCSLC4, CELLULOSE-SYNTHASE LIKE C4, CELLULOSE-SYNTHASE-LIKE C4, CSLC04, CSLC4 | encodes a gene similar to cellulose synthase The mRNA is cell-to-cell mobile. |
| 20 | A06 | 24800166 | 24800567 | + | BraA06g036830.3C | AT3G28150.1 | ALTERED XYLOGLUCAN 4-LIKE, AXYL4L, TBL22, TRICHOME BIREFRINGENCE-LIKE 22 | Encodes a member of the TBL (TRICHOME BIREFRINGENCE-LIKE) gene family containing a plant-specific DUF231 (domain of unknown function) domain. A putative xyloglucan O-acetyltransferase. TBL gene family has 46 members, two of which (TBR/AT5G06700 and TBL3/AT5G01360) have been shown to be involved in the synthesis and deposition of secondary wall cellulose, presumably by influencing the esterification state of pectic polymers. A nomenclature for this gene family has been proposed (Volker Bischoff & Wolf Scheible, 2010, personal communication). |
| 21 | A06 | 24801965 | 24804865 | + | BraA06g036840.3C | AT3G28040.1 | NA | Leucine-rich receptor-like protein kinase family protein;(source:Araport11) |
| 22 | A06 | 24805670 | 24807701 | - | BraA06g036850.3C | AT3G28030.1 | ULTRAVIOLET HYPERSENSITIVE 3, UV REPAIR DEFECTIVE 1, UVH3, UVR1 | Required for repair of pyrimidine-pyrimidinone (6-4) dimers. The mRNA is cell-to-cell mobile. |
| 23 | A06 | 24815868 | 24816878 | + | BraA06g036860.3C | AT3G28007.1 | ATSWEET4, SWEET4 | Nodulin MtN3 family protein;(source:Araport11) |
| 24 | A06 | 24820058 | 24825374 | + | BraA06g036870.3C | AT4G32050.1 | NA | neurochondrin family protein;(source:Araport11) |
| 25 | A06 | 24828083 | 24829174 | + | BraA06g036880.3C | AT3G27970.1 | NA | Exonuclease family protein;(source:Araport11) |
| 26 | A06 | 24830171 | 24830824 | - | BraA06g036890.3C | AT3G27960.1 | KINESIN LIGHT CHAIN-RELATED 2, KLRC2 | Tetratricopeptide repeat (TPR)-like superfamily protein;(source:Araport11) |
| 27 | A06 | 24831207 | 24832271 | + | BraA06g036900.3C | AT5G61760.1 | ARABIDOPSIS THALIANA INOSITOL POLYPHOSPHATE KINASE 2 BETA, ATIPK2BETA, INOSITOL POLYPHOSPHATE KINASE 2 BETA, IPK2B, IPK2BETA | Encodes an inositol polyphosphate 3-/6-/5-kinase that is localized to the nucleus. Able to complement a mutation in a yeast transcriptional regulator gene (ARG82/IPK2). Acts redundantly with ATIPK2alpha during pollen development, pollen tube guidance and embryogenesis. |
| 28 | A06 | 24832805 | 24835904 | - | BraA06g036910.3C | AT3G27925.1 | DEG1, DEGP PROTEASE 1, DEGPI, DEGRADATION OF PERIPLASMIC PROTEINS 1 | Encodes a DegP protease; nuclear gene encoding chloroplast-targeted protease that can degrade two luminal proteins, plastocyanin and OE33, suggesting a role as a general-purpose protease in the thylakoid lumen. Involved in the degradation of D1 protein of PS II, hence participating in the repair of PS II damages caused by photoinhibition. The mRNA is cell-to-cell mobile. |
| 29 | A06 | 24849933 | 24851553 | + | BraA06g036920.3C | AT3G27920.1 | ATGL1, ATMYB0, GL1, GLABRA 1, MYB DOMAIN PROTEIN 0, MYB0 | Encodes GL1, a Myb-like protein that is required for induction of trichome development. Interacts with JAZ and DELLA proteins to regulate trichome initiation. |

Chr*: Abbreviation of the Chromosome, Dir*: Abbreviation of the gene coding direction on the chromosome

6. 배추 NGS 데이터를 활용한 집단 SNP의 산출

가. 배추의 High-quality 변이 데이터의 구성을 위한 NovaSeq sequencing 수행

○ Whole genome resequencing의 경우, 공개된 genome assembly상의 변이를 탐색할 수 있는 장점이 있으나 sequencing depth가 낮은 경우 false positive인 변이 정보를 얻을 수 있다는 단점이 있다. 한쌍의 부모본에서 유래한 bi-parental mapping 집단의 경우, 부모본의 sequencing 데이터를 높여 후대 집단의 낮은 sequencing depth를 imputation 방식을 통해 보완하는 방법이 있다. 그러나 계통간의 유연관계가 불분명한 수집된 집단의 경우 이와 같은 방식을 적용할 수 없다. 따라서 배추 육종 집단의 신뢰성 높은 변이의 발굴을 위하여 한국 육종 집단을 중심으로 Illumina NovaSeq platform을 활용한 high-coverage resequencing project를 수행하였다(표 11). 신규 resequencing project를 통해 총 114개 계통에 대하여 3.42 Tbp에 달하는 sequencing data가 fastq 포맷으로 구성이 되었다.

표 11. 배추 육종 소재 114개 계통에 대한 NovaSeq sequencing 수행 결과

| # | Accession ID | Scientific name | Total base pair (bp) | Sequencing depth (X) |
|----|--------------|--|----------------------|----------------------|
| 1 | 25083 | <i>Brassica rapa</i> ssp. <i>chinensis</i> | 32,874,109,736 | 65.75 |
| 2 | 120030 | <i>Brassica rapa</i> ssp. <i>nipposinica</i> | 35,910,782,552 | 71.82 |
| 3 | 120031 | <i>Brassica rapa</i> ssp. <i>nipposinica</i> | 31,955,709,012 | 63.91 |
| 4 | CNU_28072 | <i>Brassica rapa</i> ssp. <i>parachinensis</i> | 34,036,140,806 | 68.07 |
| 5 | 10068 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 24,483,720,444 | 48.97 |
| 6 | 10069 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 26,744,081,120 | 53.49 |
| 7 | 12013 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 29,844,773,406 | 59.69 |
| 8 | 12015 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 30,259,317,632 | 60.52 |
| 9 | 26013 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 31,498,614,194 | 63 |
| 10 | 26014 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 33,413,794,004 | 66.83 |
| 11 | 26015 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 28,198,572,614 | 56.4 |
| 12 | 26016 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 27,361,514,684 | 54.72 |
| 13 | 26017 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 28,846,755,516 | 57.69 |
| 14 | 26018 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 32,900,407,594 | 65.8 |
| 15 | 26019 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 27,287,251,676 | 54.57 |
| 16 | 26020 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 30,350,312,044 | 60.7 |
| 17 | 26021 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 28,210,283,570 | 56.42 |
| 18 | 26022 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 31,602,093,890 | 63.2 |
| 19 | 27142 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 35,170,833,326 | 70.34 |
| 20 | 28052 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 21,458,666,474 | 42.92 |
| 21 | 28053 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 30,320,800,604 | 60.64 |
| 22 | 28058 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 22,183,748,542 | 44.37 |
| 23 | 28059 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 25,699,250,814 | 51.4 |
| 24 | 28060 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 25,543,973,890 | 51.09 |
| 25 | 28061 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 33,565,886,942 | 67.13 |
| 26 | 101048 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 21,728,527,332 | 43.46 |
| 27 | 279064 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 33,908,085,256 | 67.82 |
| 28 | CNU_11377 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 27,976,146,896 | 55.95 |
| 29 | CNU_11378 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 25,575,623,792 | 51.15 |
| 30 | CNU_11379 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 28,448,382,182 | 56.9 |
| 31 | CNU_11380 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 28,844,036,610 | 57.69 |
| 32 | CNU_11381 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 33,342,286,142 | 66.68 |
| 33 | CNU_11383 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 27,296,787,628 | 54.59 |
| 34 | CNU_11384 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 30,672,172,772 | 61.34 |
| 35 | CNU_11385 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 30,003,064,592 | 60.01 |

| | | | | |
|----|---------------|---|----------------|-------|
| 36 | CNU_11386 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 28,383,813,072 | 56.77 |
| 37 | CNU_11387 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 27,615,329,376 | 55.23 |
| 38 | CNU_11388 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 32,410,371,522 | 64.82 |
| 39 | CNU_11389 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 33,136,966,912 | 66.27 |
| 40 | CNU_11390 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 30,107,966,104 | 60.22 |
| 41 | CNU_11392 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 35,447,689,712 | 70.9 |
| 42 | CNU_11393 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 29,695,542,522 | 59.39 |
| 43 | CNU_11394 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 32,822,456,864 | 65.64 |
| 44 | CNU_11395 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 47,309,637,860 | 94.62 |
| 45 | CNU_11396 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 37,497,411,730 | 74.99 |
| 46 | CNU_11397 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 33,482,716,746 | 66.97 |
| 47 | CNU_11398 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 22,792,719,462 | 45.59 |
| 48 | CNU_11399 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 33,752,603,274 | 67.51 |
| 49 | CNU_11400 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 37,154,920,476 | 74.31 |
| 50 | CNU_11401 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 43,281,099,196 | 86.56 |
| 51 | CNU_11402 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 30,266,238,868 | 60.53 |
| 52 | CNU_11403 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 37,904,642,724 | 75.81 |
| 53 | CNU_11405 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 31,157,563,782 | 62.32 |
| 54 | CNU_11406 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 28,313,700,752 | 56.63 |
| 55 | CNU_11407 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 28,075,222,526 | 56.15 |
| 56 | CNU_11410 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 27,131,979,886 | 54.26 |
| 57 | CNU_11411 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 33,685,680,678 | 67.37 |
| 58 | CNU_11412 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 24,694,480,808 | 49.39 |
| 59 | CNU_11416 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 26,569,417,004 | 53.14 |
| 60 | CNU_11419 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 18,984,553,822 | 37.97 |
| 61 | CNU_11420 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 25,134,811,304 | 50.27 |
| 62 | CNU_11471 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 30,319,023,032 | 60.64 |
| 63 | CNU_11472 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 19,843,480,478 | 39.69 |
| 64 | CNU_11473 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 24,273,118,328 | 48.55 |
| 65 | CNU_11474 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 21,977,016,858 | 43.95 |
| 66 | CNU_11475 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 29,447,681,458 | 58.9 |
| 67 | CNU_11476 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 25,617,478,274 | 51.23 |
| 68 | CNU_11477 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 27,613,027,532 | 55.23 |
| 69 | CNU_11479 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 30,760,044,202 | 61.52 |
| 70 | CNU_11480 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 26,823,018,484 | 53.65 |
| 71 | CNU_11540 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 38,923,500,426 | 77.85 |
| 72 | CNU_11600 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 22,912,078,318 | 45.82 |
| 73 | CNU_11635 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 28,992,547,224 | 57.99 |
| 74 | CNU_11636 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 33,572,920,522 | 67.15 |
| 75 | CNU_11638 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 32,286,932,344 | 64.57 |
| 76 | CNU_11639 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 25,430,712,414 | 50.86 |
| 77 | CNU_11652 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 39,640,574,964 | 79.28 |
| 78 | CNU_11672 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 29,329,801,496 | 58.66 |
| 79 | CNU_11682 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 26,209,057,618 | 52.42 |
| 80 | CNU_11684 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 28,878,796,508 | 57.76 |
| 81 | CNU_11685 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 22,868,896,244 | 45.74 |
| 82 | CNU_11688 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 32,006,991,632 | 64.01 |
| 83 | CNU_11696 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 31,594,333,396 | 63.19 |
| 84 | CNU_11698 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 28,699,824,966 | 57.4 |
| 85 | CNU_11699_1-1 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 20,867,549,190 | 41.74 |
| 86 | CNU_11699_1-2 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 29,063,207,070 | 58.13 |
| 87 | CNU_11714 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 34,522,484,928 | 69.04 |
| 88 | CNU_11716 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 21,209,360,340 | 42.42 |
| 89 | CNU_11717 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 36,661,828,466 | 73.32 |

| | | | | |
|-----|------------|---|----------------|-------|
| 90 | CNU_11719 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 32,273,330,868 | 64.55 |
| 91 | CNU_11721 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 28,029,823,168 | 56.06 |
| 92 | CNU_11722 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 20,801,539,540 | 41.6 |
| 93 | CNU_11723 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 28,399,681,964 | 56.8 |
| 94 | CNU_11731 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 31,822,471,746 | 63.64 |
| 95 | CNU_11732 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 27,258,770,056 | 54.52 |
| 96 | CNU_11735 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 38,037,572,554 | 76.08 |
| 97 | CNU_12239 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 29,789,834,472 | 59.58 |
| 98 | CNU_130153 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 36,604,910,526 | 73.21 |
| 99 | CNU_130156 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 26,345,156,032 | 52.69 |
| 100 | CNU_28025 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 25,209,423,122 | 50.42 |
| 101 | CNU_28026 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 34,448,583,716 | 68.9 |
| 102 | CNU_28027 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 22,546,684,894 | 45.09 |
| 103 | CNU_28066 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 18,303,914,678 | 36.61 |
| 104 | CNU_28067 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 25,051,881,500 | 50.1 |
| 105 | GG-1 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 37,403,673,648 | 74.81 |
| 106 | L210-3 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 39,564,031,856 | 79.13 |
| 107 | L226-1 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 40,347,699,810 | 80.7 |
| 108 | L272-1 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 35,284,745,008 | 70.57 |
| 109 | rr-1 | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 30,661,367,514 | 61.32 |
| 110 | Rr-1_a | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 38,782,317,238 | 77.56 |
| 111 | RR-1_b | <i>Brassica rapa</i> ssp. <i>pekinensis</i> | 30,570,513,230 | 61.14 |
| 112 | CNU_11481 | <i>Brassica rapa</i> ssp. <i>rapa</i> | 27,018,540,834 | 54.04 |
| 113 | CNU_28062 | <i>Brassica rapa</i> ssp. <i>rapa</i> | 37,391,337,250 | 74.78 |
| 114 | 27090 | <i>Brassica rapa</i> ssp. <i>trilocularis</i> | 31,736,996,384 | 63.47 |

나. 배추의 genome assembly ver 3.0에 대한 변이 데이터의 생산

○ 생산된 114개 계통 중 표현형 조사 결과를 갖는 96개 계통 및 30X 이상의 sequencing depth를 갖는 지부와 핵배추를 대상 (그림 25)으로 SNP calling을 수행하였다. 배추의 표준 유전체는 *Brassica rapa* genome assembly ver 3.0을 활용하였으며 계통 별 SNP calling을 수행하였다. 그러나 배추의 SNP를 개체 수준이 아닌 집단을 대상으로 구성하는 과정에서 추가적으로 heterozygosity와 linkage disequilibrium을 고려한 필터링 과정을 추가하였다. 배추의 유전체 상에서 나타나는 반복서열로 인해 실제 서로 다른 영역에서 유래한 read가 genome assembly 상의 동일한 위치에 mapping되어 나타나는 높은 heterozygosity는 Tassel 5.0 software로 제거하였다. 또한 집단을 이루는 계통의 수에 따른 linkage disequilibrium으로 인해 나타나는 동일한 minimum allele frequency를 보이는 인접한 SNP의 제거를 위해 Plink를 활용한 LD pruning을 적용하였다. 집단 SNP의 필터링 이후 후속 유전 분석을 수행하는 프로그램이 요구하는 format에 따라 결과 파일의 형식을 변환하였다 (그림 26).

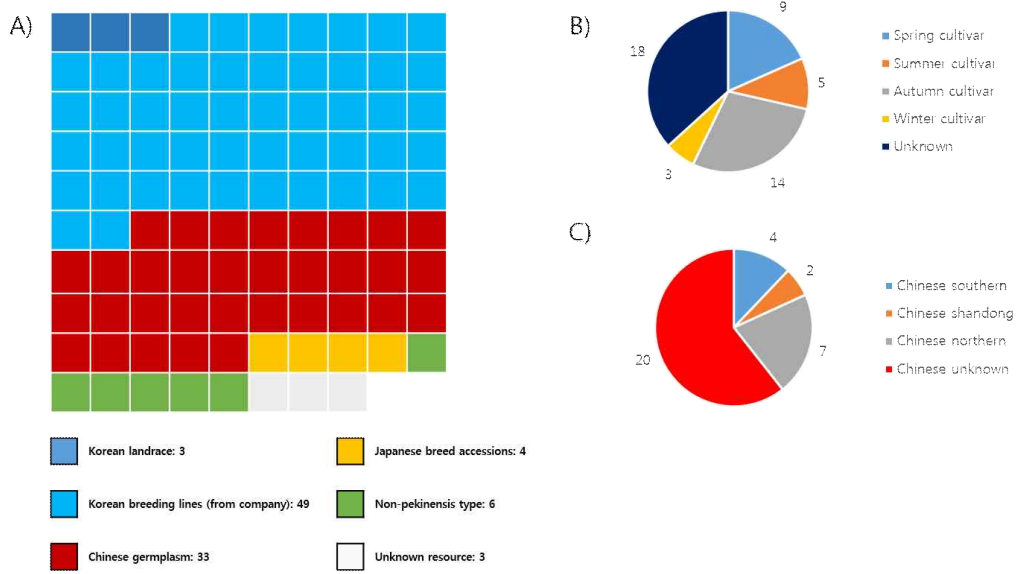


그림 25. 배추 high-sequencing depth(30X 이상) 98개 계통의 유형. A) 배추 98개 계통의 수집 지역의 출처, B) 국내 육종 집단의 작형 분류, C) 중국계 배추의 수집 지역 분류

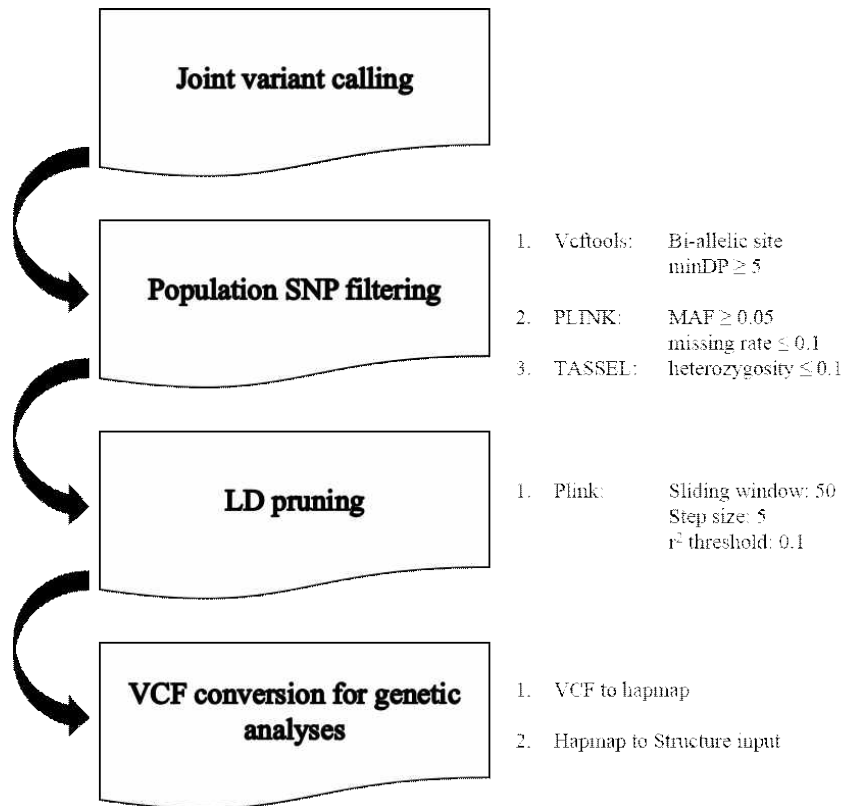


그림 26. 집단 SNP의 filtering 과정의 개요도

7. 배추 육종사 연구에 기반한 배추의 농업적 형질 연관 변이의 탐색

가. 배추의 육종사 확인에 따른 국내 육성 배추의 형성

○ 배추(*Brassica rapa*)의 기원지는 지중해 연안이며 원시적인 형태가 실크로드를 통해 중국으로 전래된 것으로 알려져있다. 현재 우리가 소비하고 있는 형태의 배추(*Brassica rapa* spp. *pekinensis*)는 양쯔강 지역에서 순무와 청경채에 해당하는 두 아종이 교잡되어 형성된 것으로 추정된다. 한국의 경우는 고려 시대의 문헌인 향약구급방에서 송이라는 이름으로 배추의 기록이 최초로 나타난다. 이후 조선시대에는 배추의 생산은 중국으로부터 수입된 종자에 의존하였다. 20세기 초에 들어서야 중국으로부터 도입한 배추의 상업적인 육종 및 유통이 시작된다.

○ 일본의 경우 중국의 산동계 배추가 보이는 결구 형질에 주목하고 이를 도입하여 19세기 후반부터 결구 배추의 육종이 시작된다. 이후 배추의 결구형질의 고정 및 작형의 다변화(조생, 중생, 만생)의 달성에 이른다. 이와 같이 중국으로부터 도입된 계통 및 일본을 거쳐 육성된 계통들이 1950년 이후 배추 육종 프로그램에서 서로 교잡되어 현생 한국 배추 육성종의 근간을 이루게 된다(그림 27).

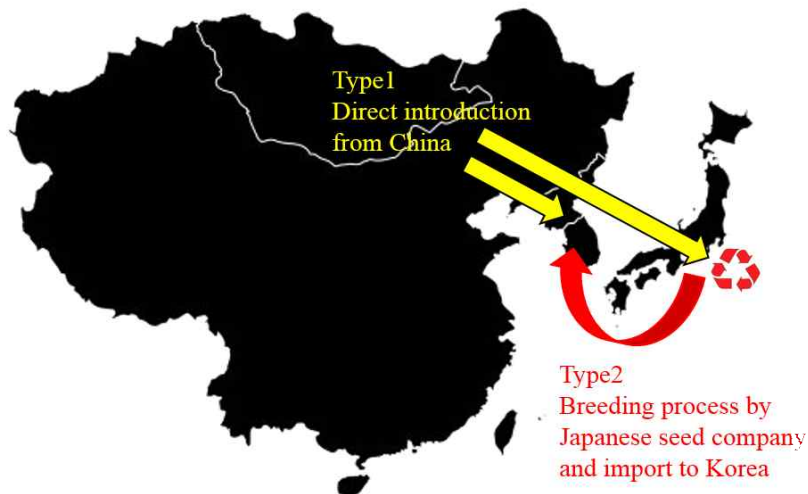


그림 27. 20세기 이후 한국의 배추(*Brassica rapa* spp. *pekinensis*) 전래 경로

○ 한국 전쟁이후 배추의 육종방향은 수량 증대에 초점이 맞추어지게 되어 F₁ hybrid 중심의 배추 육종이 수행된다. 이후 배추의 주년 생산의 달성을 위해 봄철의 저온에 의한 추대방지를 목표로 추대 저항성 계통이 육성되어 봄 및 여름철 고랭지 재배를 위한 계통들이 보급된다. 또한 한 여름 평지재배를 위해 내서성 계통 및 남부 지역 월동 재배를 위해 내한성 계통들이 각각 육종에 반영되어 1980년대에 한국의 배추 생산은 주년화를 달성하게 된다(그림 28). 이후 국내의 주요 배추 생산지에서 수확량을 크게 저하시키는 요인인 뿌리혹병에 대한 대응으로 일본에서 저항성 계통이 도입되어 활용되었으며 현재에는 기능성 2차 대사 물질에 초점이 맞추어진 배추 육종이 진행되고 있다.

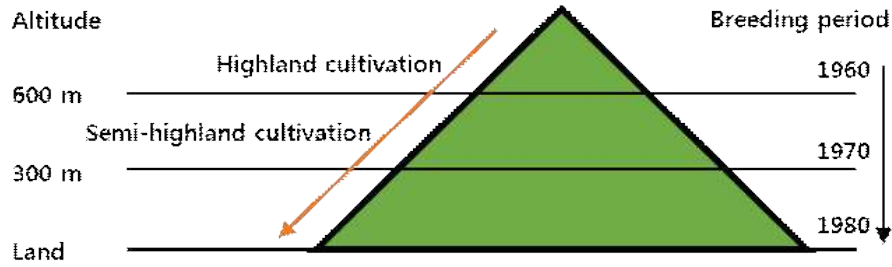
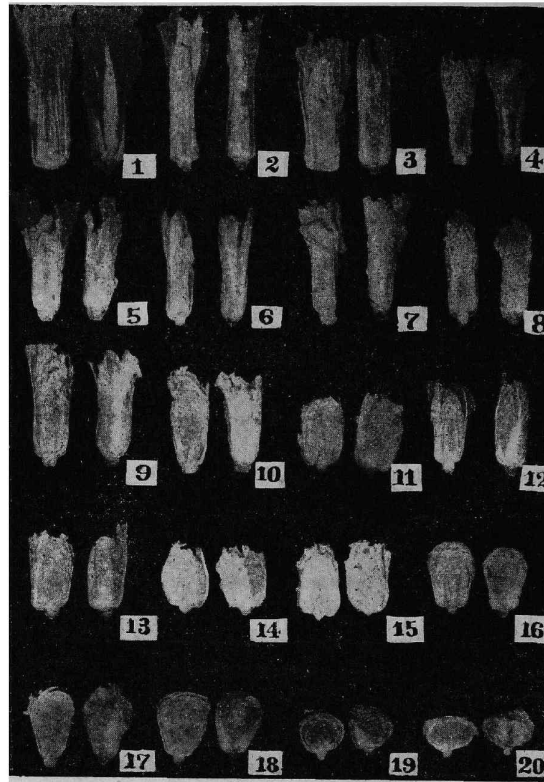


그림 28. 1950년대 이후 육종 시기에 따른 여름철의 배추 재식 가능 지역의 확대

나. 배추의 기원지에 따른 표현형 변이의 확인

○ 국내로 유입된 배추는 중국으로부터 직접 도입된 품종 및 일본에서 고정된 품종으로 양분되어 그림 29와 같이 크게 결구 형태에서 큰 차이를 나타낸다(그림 29).



第 2 圖 草型別に見た華北白菜の代表的葉球

- | | | |
|-----------------|------------------|-------------------|
| 北方群 | 8. 花心類 青芽小白菜(濟寧) | 15. 膠縣類 膠州白帮(青島) |
| 1. 大葉型 大葉青(昌黎) | 9. 河頭型 河頭(北平) | 16. 菊花心類 菊花心(張店) |
| 2. " 八葉青(唐山) | 10. " 大白口(南苑) | 南方群 |
| 核桃紋類 | 山東群 | 17. 二莊型 窩心(石家莊) |
| 3. 核桃紋類 核桃紋(唐山) | 山東型 | 18. " 二莊青帮(正定) |
| 4. " 核桃紋(德縣) | 11. 芝罘類 牛心(北平) | 19. 三莊型 小窩心早生(保定) |
| 5. 麻葉類 麻葉青(辛集) | 12. " 芝罘白菜(日本) | 20. " 三莊(保定) |
| 6. " 青麻葉(軍糧城) | 13. " 京都三號(日本) | |
| 7. 花心類 黃芽白菜(泰安) | 14. 膠縣類 浮山包頭(青島) | |

그림 29. 중국 화북 지역 재래종 배추 및 일본 고정종의 결구 형질 조사

○ 이와 같은 사실 및 30X 이상으로 sequencing된 배추 유전자원 98점의 수집 출처가 크게 일본계 배추의 결구 형질의 영향을 강하게 받은 국내 육종 집단과 중국계 배추인 것에 착안하여 수집된 계통 중 pekinensis type에 해당하는 82개 계통의 표현형 형질을 국내 및 중국 그룹으로 나누어 그 분포를 대조하였다(그림 30).

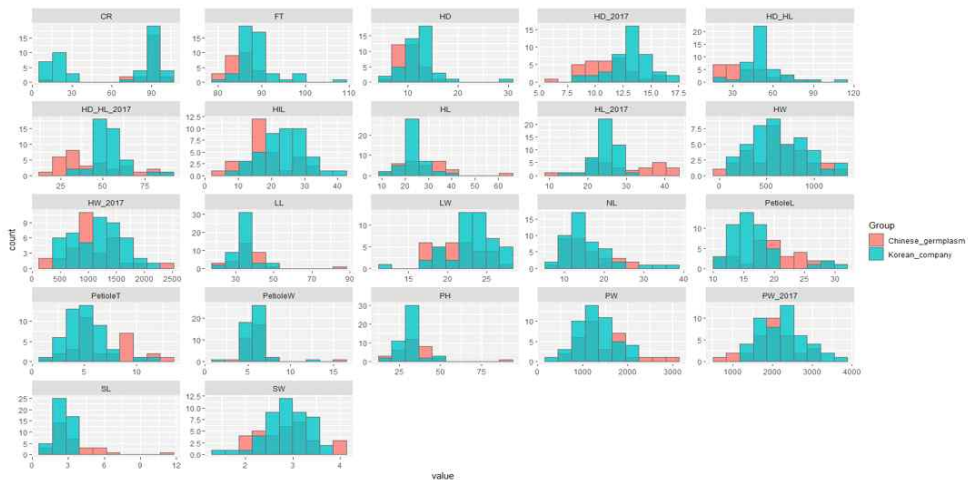


그림 30. 배추 수집단의 수집 출처에 따른 표현형 분포. CR 뿌리혹병 저항성, FT 개화기, HD 구폭(2013년 조사), HD_2017 구폭(2017년 조사), HD_HL 결구형태(2013년 조사), HD_HL_2017 결구형태(2017년 조사), HIL 내엽의 수(2013년 조사), HL 구고(2013년 조사), HL_2017 구고(2017년 조사), HW 구중(2013년 조사), HW_2017 구중(2017년 조사), LL 엽장(2013년 조사), LW 엽폭(2013년 조사), NL 외엽의 수(2013년 조사), PetioleL 중륵 길이(2013년 조사), PetioleT 중륵 두께(2013년 조사), PetioleW 중륵 폭(2013년 조사), PH 초장(2013년 조사), PW 생체중(2013년 조사), PW_2017 생체중(2017년 조사), SL 줄기 길이(2013년 조사), SW 줄기 두께(2013년 조사)

○ 표현형 분포의 대조 결과 주로 구의 관련 형질에서 두 집단 간의 차이가 나타났다. 특히 내엽의 수(HIL)는 한국 육종 집단이 중국 집단보다 높게 나타났으며 구의 직경 및 높이를 의미하는 구폭(HD)과 구고(HL)의 경우 각각 중국과 한국의 배추가 높게 나타났다. 이는 중국계 배추에서 특이적으로 나타나는 실린더형과 한국계 배추에서 일반적으로 발견되는 계란형 결구 형질을 반영한다. 또한 한국 육종 집단에 적극적으로 도입된 뿌리혹병 저항성과 만추대성으로 인해 한국 육종 집단은 중국 집단보다 뿌리혹병에 대해 높은 저항성 및 낮은 개화기 양상을 나타낸다.

○ 조사된 표현형을 대상으로 shapiro-wilk test를 수행하였을 때 총 8개의 표현형에 대해서 표현형 분포의 정규성이 나타났다(표 13). (HD_2017, HD_HL_2017, HIL, HW, HW_2017, LW, PW_2017, SW)

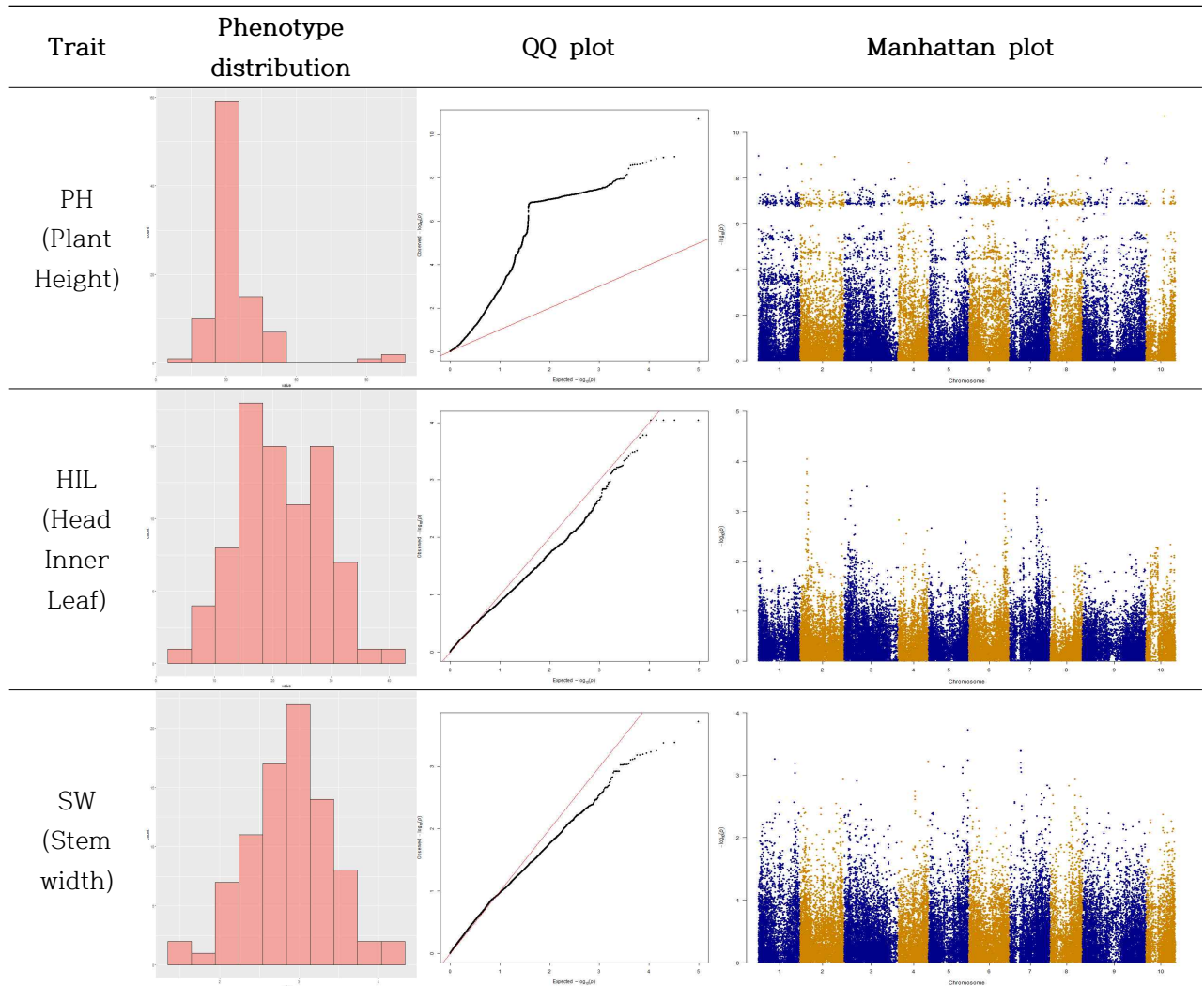
표 12. 조사된 양적 형질 분포의 정규성 검정 결과

| # | Trait | Description | Number of Surveyed accessions | P-value | W |
|----|------------|--|-------------------------------|-----------|---------|
| 1 | CR | Clubroot disease resistance | 71 | 4.37E-10 | 0.73134 |
| 2 | FT | Flowering time | 74 | 2.64E-06 | 0.8748 |
| 3 | HD | Head diameter | 75 | 1.91E-07 | 0.84279 |
| 4 | HD_2017 | Head diameter (surveyed in 2017) | 75 | 0.4832 | 0.98434 |
| 5 | HD_HL | Proportion between head diameter and length | 75 | 0.01342 | 0.95766 |
| 6 | HD_HL_2017 | Proportion between head diameter and length (surveyed in 2017) | 75 | 0.08151 | 0.9709 |
| 7 | HIL | Number of inner leaf in the head | 75 | 0.7093 | 0.98807 |
| 8 | HL | Head length | 76 | 6.44E-05 | 0.91235 |
| 9 | HL_2017 | Head length (surveyed in 2017) | 75 | 0.0008368 | 0.9352 |
| 10 | HW | Head weight | 74 | 0.5243 | 0.9849 |
| 11 | HW_2017 | Head weight (surveyed in 2017) | 77 | 0.4283 | 0.98367 |
| 12 | LL | Leaf length | 80 | 1.25E-09 | 0.77869 |
| 13 | LW | Leaf weight | 80 | 0.07177 | 0.9716 |
| 14 | NL | Number of outer leaf | 80 | 2.53E-05 | 0.90735 |
| 15 | PetioleL | Petiole length | 80 | 0.0003469 | 0.93149 |
| 16 | PetioleT | Petiole thickness | 80 | 0.0006741 | 0.93708 |
| 17 | PetioleW | Petiole width | 80 | 2.23E-12 | 0.65691 |
| 18 | PH | Plant height | 80 | 1.06E-08 | 0.8121 |
| 19 | PW | Plant weight | 80 | 0.006743 | 0.95501 |
| 20 | PW_2017 | Plant weight (surveyed in 2017) | 80 | 0.9397 | 0.99283 |
| 21 | SL | Stem length | 80 | 2.50E-10 | 0.74795 |
| 22 | SW | Stem width | 79 | 0.9177 | 0.9922 |

다. High-coverage SNP 기반의 배추 조사 표현형의 genome-wide association study

○ 국내로 유입된 배추는 중국으로부터 직접 도입된 품종 및 일본에서 고정된 품종으로 양분되어 결구 형태에서 큰 차이를 나타낸다.

표 12. 표현형 분포에 따른 MLM GWAS 분석 결과의 예시



○ 조사된 표현형은 크게 정규분포를 따른 경우와 그렇지 않은 경우로 구분이 가능하였다. 정규분포를 따르지 않는 표현형은 QQ Plot을 확인하였을 때 모든 마커의 유의성이 예측된 유의수준보다 과대평가되는 경향을 나타내었다. 이는 SNP의 유의성을 측정하는 MLM이 표현형의 분포를 정규분포로 가정하기 때문에 나타나는 현상이라 추정할 수 있다. 그 외의 정규분포를 따르는 표현형의 경우 크로모솜 상에 peak를 일부 나타내었으나 유의성의 기준으로 삼은 4.69를 넘지 못하는 양상을 보였다. 이는 현재 98개로 한정된 집단의 대다수가 국내의 육종회사에서 수집된 자원으로 국내 시장 환경에서 선호되는 표현형의 강한 선발 및 유전적 다양성의 감소를 유발하여 genomewide association study에서 유의한 SNP의 식별을 수행하지 못한 것으로 판단된다. 추후 현재 집단에서 나타나지 않는 표현형을 나타내지 않는 계통의 추가 및 SNP 필터링 조건의 완화, 별도의 통계 모델의 적용으로 표현형에 연관된 SNP의 발굴을 위한 파이프라인을 개선할 예정이다(표 12).

8. 분자마커 개발을 위한 유전연구

가. DH 집단을 이용한 뿌리혹병 저항성에 관한 QTL mapping

○ 뿌리혹병은 환물기생균인 *Plasmodiophora brassicae*에 의해 야기되는 토양 감염병이다. 이는 재배과정에서 감염된 배추의 뿌리에 *P. brassicae* 증식하면서 배추의 생육을 저해시키고 상품성에 심각한 피해를 주는 병이다. 뿌리혹병을 방제하기위한 여러 가지 방법이 있으나, 현재 마커를 이용한 저항성 품종개발이 중요하게 대두되고 있다. 저항성 연관 마커를 개발하기 위하여 QTL분석을 하였으며, 분석을 기반으로 high-throughput가 가능한 SNP/Indel 마커를 개발 하였다.

나. 방립 저항성 집단을 이용한 표현형 검정

○ 방립 균주에 대응하는 배추의 저항성 연구를 위하여, 배추(*brassica rapa*)의 DH mapping population (n=81) 과 저항성 라인 (09CR500)과 감수성 라인 (09CR501)을 사용하였다. 뿌리혹병 저항성 테스트를 위하여 안성과 대전에서 표현형 검정을 진행하였다. 파종 후 2주된 배추 묘에 1×10^7 농도의 방립 균주를 접종 한 후 8주 후에 뿌리를 세척하여, 표준화된 프로토콜에 따라 disease index를 평가 하였다 (그림 31).

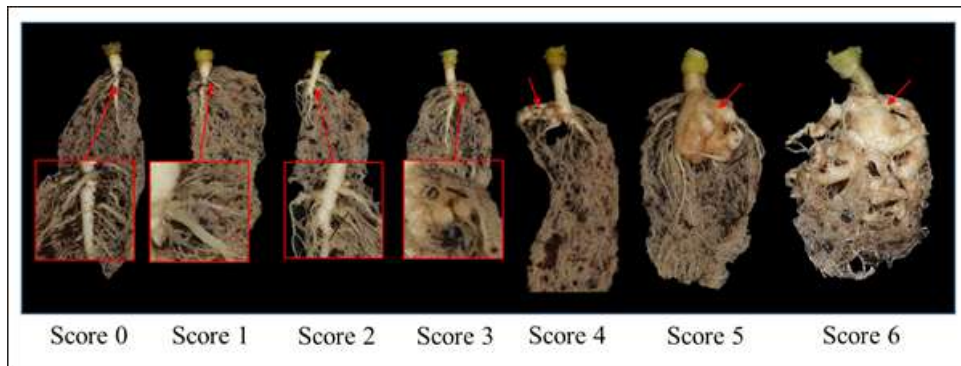


그림 31. *Plasmodiophora brassicae*에 감염된 뿌리의 증상 및 뿌리 형태 변형의 정도에 따른 평가 기준.

○ 대전에서의 표현형 검정 결과 09CR500은 전부 저항성을 보였으며, 09CR501은 감수성을 보였다. 그리고 DH 집단에서는 저항성과 감수성 식물이 각각 40개와 1개의 missing data로 저항성과 감수성의 비율이 1:1을 나타냈다. 안성에서의 표현형 검정 결과 대전과 동일하게 09CR500에서 전부 저항성이며, 09CR501에서는 전부 감수성이었다. DH집단에서는 39개의 저항성과 40개의 감수성 그리고 2개의 missing 결과를 보였다. 대전과 동일하게 안성에서도 저항성과 감수성의 비율이 1:1을 나타내었다(그림 32, 표13).

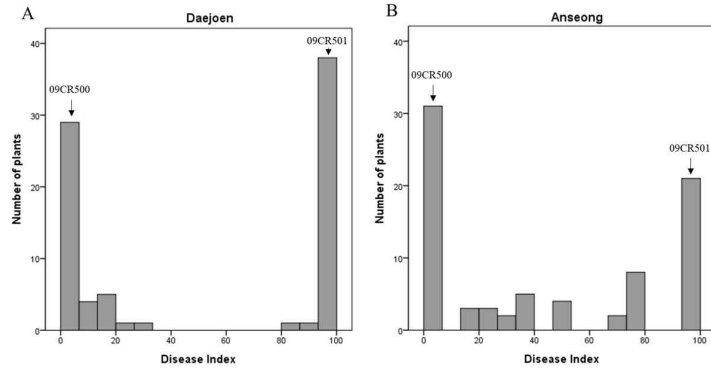


그림 32. '09CR-DH' 집단의 (A)대전과 (B)안성에서의 2년간의 Disease index 값의 분포

표 13. 방림 균주에서 반응하는 DH 집단의 뿌리혹병 저항성관련 유전분석

| Location | Plant materials | Total No. | Phenotype of | | | Expected ratio (R:S) | χ^2 |
|----------|-----------------|-----------|----------------|-----------------|-------------|----------------------|----------|
| | | | Resistance (R) | Susceptible (S) | Missing (-) | | |
| Daejeon | 09CR500 | 10 | 10 | 0 | 0 | - | - |
| | 09CR501 | 10 | 0 | 10 | 0 | - | - |
| | DH population | 81 | 40 | 40 | 1 | 1:1 | 0.01 |
| Anseong | 09CR500 | 10 | 10 | 0 | 0 | - | - |
| | 09CR501 | 10 | 0 | 10 | 0 | - | - |
| | DH population | 81 | 39 | 40 | 2 | 1:1 | 0.06 |

다. 배추의 뿌리혹병 저항성에 관한 QTL 분석

○ 유전자 지도 작성을 위하여 총 808개의 primer를 이용하여 부모계통간의 변이 Screening을 진행하였다. 총 222개의 마커를 이용하여 Joinmap version 4.0를 이용하여 연관지도를 작성하였다. 총 731.7cM 길이의 10개의 연관지도를 만들었다. QTL 분석을 위하여, WinQTL cartographer 2.5와 IciMapping 4.1의 프로그램을 사용하였다.

○ WinQTL cartographer 2.5의 분석결과 대전과 안성에서 모두 8번 염색체에서 QTLs를 발견하였다. 각각의 QTL을 $PbBrA08^{Banglim-1}$ 과 $PbBrA08^{Banglim-1}$ 로 표시하였다. $PbBrA08^{Banglim-1}$ 은 18.86에서 38.86 cM 사이에 위치하며, 29.12cM에 flanking marker인 "09CR.11390652"가 위치한다. $PbBrA08^{Banglim-2}$ 의 경우 $PbBrA08^{Banglim-1}$ 의 안쪽에 위치한다. phenotypic variation 값은 26.0%이며, LOD 값은 7.9이다. IciMapping 4.1 프로그램을 통한 분석결과 WinQTL cartographer의 QTL 범위 안의 ~1Mb 가량의 QTL이 동일하게 분석되었다(그림 33, 표 14).

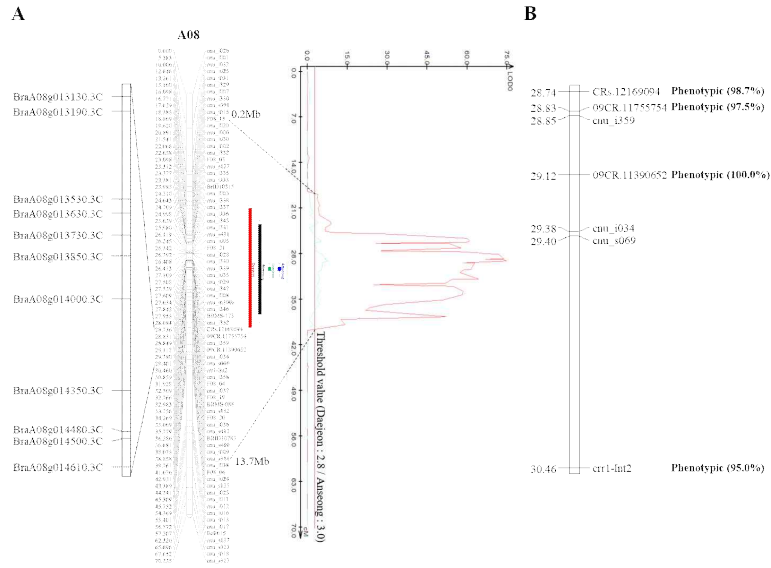


그림 33. 배추의 8번염색체의 연관지도와 방립균주의 저항성 형질과 연관되는 *PbBrA08^{Banglim}* QTL mapping 결과.(A) 연관지도에서의 *PbBrA08^{Banglim}*의 위치와 후보유전자. 연관지도의 왼쪽은 유적거리를 centi-morgans (cM)으로 표시하였으며, 오른쪽은 마커를 표시하였다. 오른쪽에 QTL 결과를 표시하였다. (B)는 저항성 loci에 근접한 연관마커를 표시하였다.

표 14. 방립균주에 대항하는 QTL 결과 요약

| Loci name | Location | Chr No. | Closest Marker | LOD | R^2 (%) | Additive |
|-------------------------------------|----------|---------|----------------|------|-----------|----------|
| <i>qPbBrA08^{Banglim}-1</i> | Daejeon | 8 | 09CR.11390652 | 74.8 | 97.1 | -47.0 |
| <i>qPbBrA08^{Banglim}-2</i> | Anseong | 8 | 09CR.11390652 | 7.9 | 26.0 | -26.1 |

라. 배추 방립 균주에 대한 뿌리혹병 마커 개발

○ 두 개의 마커(09CR.11390652, 09CR.11755754)는 *PbBrA08^{Banglim}*와 밀접하게 연관되었다. 마커 09CR.11390652는 High-resolution melting (HRM) 분석용 SNP 마커로 개발되었으며, 09CR.11755754는 indel마커로 개발되었다(그림 35). 위 두 개의 마커는 DH 집단에서 저항성과 감수성을 구별할 수 있으며, 리를 통해 배추의 뿌리혹병 저항성 품종 개발을 위한 marker-assisted selection에 도움을 줄 것이다.

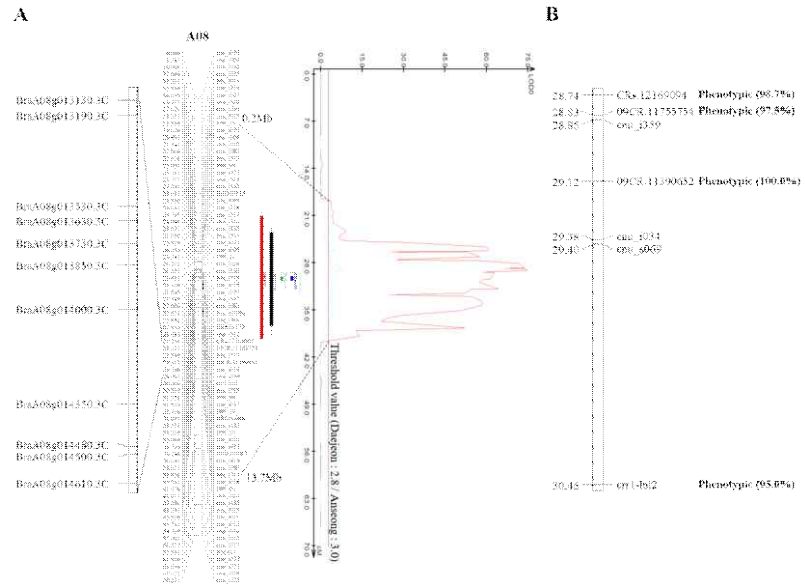


그림 34. *PbBrA08^{Banglim}* 유전자좌와 연관된 SNP, InDel 마커.(A) LightScanner System를 이용하여 SNP 마커 '09CR.11390629'을 분석한 것과 표현형 값 (B) LabChip® GX TouchTM을 사용하여 indel 마커인 '09CR.11755630' 전기영동 사진과 표현형 값

제6절 분자마커의 high-throughput selection을 위한 마커 제공

1. MABC를 위한 배추의 Fluidigm genotyping assays용 SNP 마커 세트 개발

가. 실험 재료 및 육종 소재 확보

○ 배추의 background selection marker 및 F₁ 순도 검정 SNP marker 선발을 위해 Hankookseed Co., Ltd.와 Koregon Co., Ltd.에서 분양 받은 고정종 45계통, 그리고 이들을 양친으로 사용하여 교배로 만든 29 교배 조합의 F₁ 종자를 사용하였다. 또한 개발한 marker 로 F₁ 순도 검정의 유용성을 검정하기 위해 코레곤 종묘회사에서 채종을 의뢰한 3곳(채종포 A, 채종포 B, 채종포 C)의 농가로부터 생산된 4개의 F₁ hybrid 종자와 그의 양친을 확보하여 순도검정을 수행하였다(표 1).

표 1. 32개의 배추 F₁ 품종과 그의 교배 조합

| Hybrid ID | Cross combinations (P ₁ XP ₂) ^z | Company |
|-------------|---|-----------------------|
| HNC-1 | C-284 X C-287 | Hankookseed Co., Ltd. |
| HNC-2 | 2b X 2a | |
| HNC-3 | C-307 X C-188 | |
| HNC-4 | C-3 X 2b | |
| HNC-5 | C-316 X C-317 | |
| HNC-6 | C-3 X C-4 | |
| HNC-7 | C-284 X 7b | |
| HNC-8 | C-172 X 8b | |
| HNC-9 | 9a X 9b | |
| HNC-10 | C-241 X C-64 | |
| HNC-11 | C-90 X C-238 | |
| HNC-12 | C-287 X C-190 | |
| HNC-13 | 13a X C-289 | |
| HNC-14 | C-284 X C-289 | |
| HNC-15 | 15a X C-289 | |
| HNC-16 | C-290 X C-289 | |
| HNC-17 | C-292 X C-289 | |
| HNC-18 | C-186 X C-287 | |
| HNC-19 | C-287 X 19b | |
| HNC-20 | 20a X C-329 | |
| HNC-21 | C-252 X C-3 | |
| HNC-22 | C-237 X C-64 | |
| HNC-23 | C-265 X C-64 | |
| HNC-24 | C-259 X C-64 | |
| 5725 | 5725_2 X 5725_1 | Koregon Co., Ltd |
| 5028 | 5028_2 X 5028_1 | |
| BN226 | BN226_2 X BN226_1 | |
| BN227 | BN227_2 X BN227_1 | |
| ND | ND_2 X ND_1 | |
| Seed farm A | | Koregon Co., Ltd |
| Seed farm B | SF-1 X SF-2 | |
| Seed farm C | | |

나. SNP 콜링 및 SNP 변이 분석

○ 국내외에서 수집한 배추 192 계통을 기반으로 3X에서 생산된 리드를 가지고 표준유전체 서열에 맵핑 및 SNP 콜링(calling)을 하기 위하여 아래와 같은 작업을 하였다. 트리밍된 서열을 BWA(Burrows-Wheeler Alignment tool)을 이용하여, BRAD v1.2(<http://brassicadb.org/brad/index.php>)에 있는 표준유전체 서열에 맵핑하였다. 위 생성물을 SAM(Sequence Alignment Map) 파일로

생성하였다. 생성된 SAM 파일을 Samtools(<http://samtools.sourceforge.net/>)을 이용하여 BAM(Binary Alignment Map) 파일로 변환하였다. 생성된 SAM을 이용하여 맵핑된 리드들의 SNP 콜링 정확도를 높이기 위하여 게놈 분석 툴인 Picard(<https://broadinstitute.github.io/picard/>), GATK(Genome Analysis Tool Kit), Samtools를 이용하여, 분류(sorted), 듀플리케이트의 제거(removed duplicates), 그룹 고정(fixed group), 및 SNP 재정렬(realigned)을 수행하였다. 위 과정의 생산물을 Bcftools, vcftools를 이용해 SNP을 예측하였으며, SNPs는 In-home Perl script를 이용하여 stringent condition과 grouped를 토대로 선별하였다. 5X 사이의 NGS 리드를 생산하였다. 생산된 리드는 퀄리티가 낮은 서열을 트리밍(trimming)하기 위해 FastX 소프트웨어를 이용하여 낮은 복잡도 및 퀄리티(low complexity and quality)를 트리밍하였다.

○ 192계통에서 만들어진 SNP을 염색체의 좌(loci)를 기반으로 병합한 후 SNP 매트릭스(matrix)를 생성하였다. SNP 매트릭스를 생성하기 위한 파라미터로 SNP 콜링 값은 MAF(Minor Allele Frequency) 값이 5% 이상인 값을 선별하였고, 그 결과 1,678,054개의 SNPs를 선별하였다. 선별된 SNPs를 PIC(Polymorphism Information Content) 값이 0.3 이상인 것으로 추가 선별하였다. 그 결과, 총 573,747개의 SNPs를 선별되었다. 선별된 573,747개의 SNPs가 분석 실수 또는 서열 오류(sequence error)로 인한 SNP가 아닌지 확인하기 위하여 HRM(High Resolution Melt) 방식으로 PCR(Polymerase Chain Reaction)하기 위한 SNP 후보 선정 후, 각 염색체에 고루 분포하게 하기 위해 2Mb 간격으로 프라이머를 제작하고, 프라이머가 단일 위치좌만 증폭하는지를 평가한 후 단일 염기좌 맵핑 영역만 이용하여 프라이머를 제작 후 검증하였다. 증폭 및 SNP가 확인된 위치를 표적으로 Fluidigm (<https://www.fluidigm.com/>) 플랫폼에 맞게 디자인을 하였다. 이를 통하여 177개의 Fluidigm용 SNP 마커를 생산하였다.

다. SNP 마커의 효율성 검증

○ 분양 받은 시료를 개발한 177개의 마커를 이용하여 SNP분석을 실시하였다. 그림36에서 각각의 A의 경우 M(모본)에서 XY 타입의 유전형을 발견할 수 있으며 이는 M의 순도가 떨어짐을 의미한다. 반면 B의 경우 M과 F(부본)의 유전형이 모두 XX, YY로 동형접합의 유전형을 가지고 있음을 알 수 있고, 그에 따른 F₁ 또한 XY 타입으로 순도가 높음을 알 수 있다(그림 1).

○ 177개의 마커를 이용하여 두 종자회사의 내혼계 45점을 분석한 결과, 종자회사로부터 육종가들이 평가할 때 계통 고정이 되었다고 판단된 시료임에도 불구하고 동형접합의(homozygous) SNP는 82.5%에서 98.9%사이의 값을 나타내었다(표 2). 이를 통하여 본 발명의 177개의 마커는 배추의 순도 검증에 사용될 수 있음을 확인하였다.

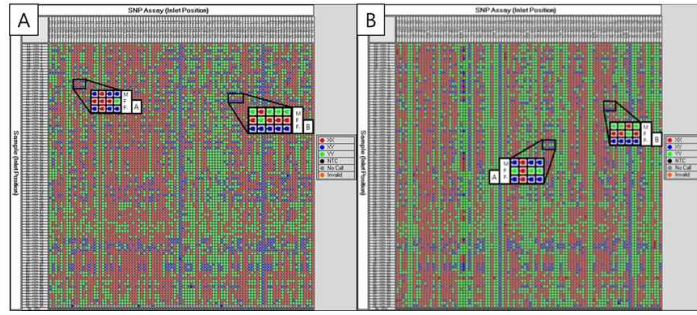


그림 1. Fluidigm SNP assays 분석 결과

표 2. Fluidigm SNP genotyping 분석을 이용한 45개의 내혼계 계통의 순도 분석

| Inbred lines | SNP type | | | Purity (%) |
|--------------|----------------------|-----------------|------------------|------------|
| | XX & YY ^z | XY ^y | N/A ^x | |
| 5725_1 | 173 | 4 | 0 | 97.7 |
| 5725_2 | 165 | 11 | 1 | 93.2 |
| 5028_1 | 172 | 5 | 0 | 97.2 |
| 5028_2 | 172 | 4 | 1 | 97.2 |
| BN226_1 | 173 | 3 | 1 | 97.7 |
| BN226_2 | 175 | 2 | 0 | 98.9 |
| BN227_1 | 174 | 3 | 0 | 98.3 |
| BN227_2 | 175 | 2 | 0 | 98.9 |
| ND_1 | 174 | 3 | 0 | 98.3 |
| ND_2 | 173 | 2 | 2 | 97.7 |
| 2a | 174 | 3 | 0 | 98.3 |
| 2b | 173 | 4 | 0 | 97.7 |
| 7b | 174 | 3 | 0 | 98.3 |
| 8b | 169 | 7 | 1 | 95.5 |
| 9a | 153 | 24 | 0 | 86.4 |
| 9b | 174 | 3 | 0 | 98.3 |
| 13a | 172 | 5 | 0 | 97.2 |
| 15a | 174 | 3 | 0 | 98.3 |
| 19b | 168 | 8 | 1 | 94.9 |
| 20a | 146 | 31 | 0 | 82.5 |
| C-3 | 173 | 2 | 2 | 97.7 |
| C-4 | 174 | 3 | 0 | 98.3 |
| C-64 | 173 | 4 | 0 | 97.7 |
| C-90 | 174 | 3 | 0 | 98.3 |
| C-172 | 174 | 2 | 1 | 98.3 |
| C-186 | 152 | 23 | 2 | 85.9 |
| C-190 | 173 | 4 | 0 | 97.7 |
| C-237 | 155 | 21 | 1 | 87.6 |
| C-238 | 173 | 2 | 2 | 97.7 |
| C-241 | 168 | 8 | 1 | 94.9 |
| C-252 | 149 | 27 | 1 | 84.2 |
| C-259 | 152 | 24 | 1 | 85.9 |
| C-265 | 173 | 3 | 1 | 97.7 |
| C-284 | 173 | 4 | 0 | 97.7 |
| C-287 | 172 | 4 | 1 | 97.2 |
| C-289 | 169 | 7 | 1 | 95.5 |
| C-290 | 173 | 3 | 1 | 97.7 |
| C-292 | 171 | 4 | 2 | 96.6 |
| C-329 | 173 | 4 | 0 | 97.7 |
| C-307 | 174 | 3 | 0 | 98.3 |
| C-188 | 157 | 18 | 2 | 88.7 |
| C-316 | 174 | 3 | 0 | 98.3 |
| C-317 | 172 | 5 | 0 | 97.2 |
| SF-1 | 153 | 24 | 0 | 86.4 |
| SF-2 | 172 | 4 | 1 | 97.2 |

^z XX & YY is refer to a specific homozygous SNP

^y XY is refer to a specific heterozygous SNP

^x NA means that it was not amplified.

라. SNP 마커의 효율성 검증

○ 177개의 Fluidigm SNP type assays 이용하여 45종류의 inbred lines의 순도를 분석하였다(그림 2) 유전형이 XX와 YY인 것을 homozygous type으로 이 빈도가 높을수록 상대적으로 순도가 높은 inbred line이며, XY인 것을 heterozygous type으로 분석하여, 이 빈도가 높을수록 상대적으로 순도가 낮은 inbred line으로 판단하였다. Inbred lines의 평균 순도는 95.4%로 나왔다. BN226_2와 BN227_2의 시료의 경우, 175개의 marker에서 homozygous type이며, 2개의 marker에서 heterozygous type으로 나타내었고 98.9%로 순도가 가장 높았다. 20a 시료는 159개의 marker에서 homozygous type이며, 31개의 marker에서 heterozygous type으로 나타나 82.5%로 순도가 가장 낮았다(표 3). 이는 전통적 육종에서는 계통을 고정하기 위해서는 육종가의 경험과 표현형에 대한 평가로 이루어졌으나, marker를 이용한 유전형 분석으로 정량적인 순도 분석이 가능함을 나타낸다.

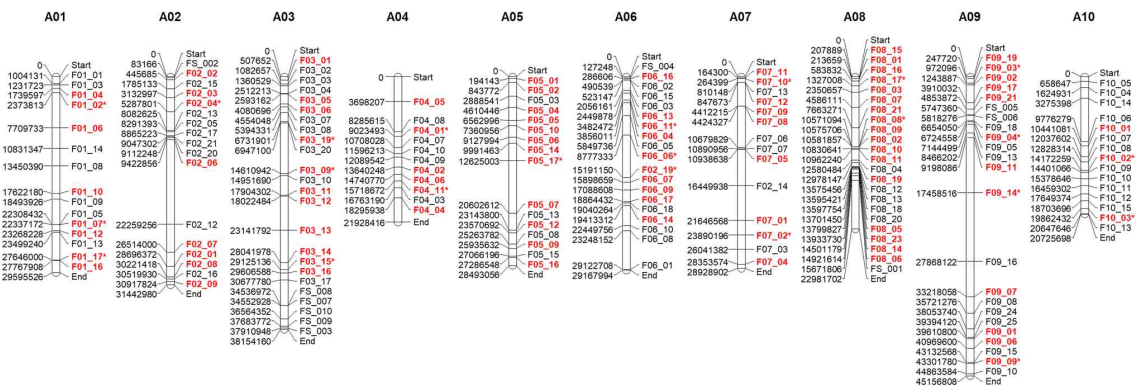


그림 2. 배추에서 177 개의 SNP (Single nucleotide polymorphism) 마커의 위치 (막대 왼쪽 : 물리적 위치, 오른쪽 : 연마커 이름, 빨간색 글자 : background 선택 마커, *는 F₁ 순도 테스트 마커)

마. MABC와 F₁ 순도 검정을 위한 마커 세트 선발

○ 177개의 marker 중 background selection marker로 유용한 marker 96개 set를 선발하고자 하였다. 선발 조건은 다음과 같다. 1)부분과 모본의 유전형이 homozygous type이다. 2)F₁의 유전형이 양친의 유전형을 따른다. 3)marker는 교배 조합 내에서 다형성을 많이 나타내는 marker이다. 4)10개의 염색체에 고루 분포하도록 하여야 한다(그림 3).

○ 96개의 MABC 마커 중 F₁ 순도 검정용 마커 24개를 선발하였다(그림 3). 마커는 96개의 마커내에서 교배조합간의 다형성을 많이 나타내며 선발 효율이 높은 마커 순으로 선발하였다. 이를 통하여 다량의 marker를 사용하여 비용 증가와 시간 소모를 줄이고, 특정 유전자만 보는 것이 아닌 유전체단위의 검정을 통하여 높은 선발 효과를 기대 할 수 있다.

바. F₁ 종자 순도 검정용을 위한 마커 세트 적용

○ 선발된 24개의 marker를 이용하여 교배 조합과 채종포 종자를 대상으로 marker의 유용성 검증 시험을 하였다. 선발된 marker를 이용하여 29개의 교배 조합의 F₁ 순도 검정하였다(그림 3,4).



그림 4. 96.96 IFC (Integrated Fluidic Circuit)을 이용한 24개 마커의 genotype 산포도

○ 채종포 종자의 순도검정은 코레콘 종묘에서 각기 다른 3곳의 위탁 농가에서 수확한 종자를 대상으로 검정하였다. 각 농가별로 채종한 종자 100립씩을 사용하여, 3개의 SSR marker인 ACMP00133, ACMP00153, ACMP00586를 이용하여 순도 검정하였다. 채종포 A, B, C의 각각 92.0%, 96.0%, 97.3% 순도를 보였다. 채종포 A, B, C에서 각각 1개의 순도가 불량하다고 판단되는 시료를 대상으로 24개 SNP marker의 유용성 검정에 사용하였다(표 3).

표 3. 24개 마커를 사용한 3개의 채종포 농장의 순도 분석 결과

| Chromosome | Marker | P ₁ | P ₂ | Seed farm A | Seed farm B | Seed farm C |
|------------|--------|----------------|----------------|-------------|-------------|-------------|
| A01 | F01_02 | AA | GG | GG | GG | AG |
| | F01_07 | GG | TT | TT | GT | GT |
| | F01_17 | CC | TT | TT | CT | CT |
| A02 | F02_04 | GG | TT | TT | TT | GT |
| | F02_19 | GG | TT | TT | TG | TG |
| A03 | F03_19 | GG | GG | GG | GG | GG |
| | F03_09 | GG | TT | TT | TT | TG |
| | F03_15 | GG | AA | AA | GA | GA |
| A04 | F04_01 | TT | CC | CC | CT | CT |
| | F04_11 | TT | TT | TT | TT | TT |
| A05 | F05_10 | GA | AA | AA | AA | AA |
| | F05_17 | CC | CC | CC | CC | CC |
| A06 | F06_11 | AA | GG | GG | GA | GA |
| | F06_06 | AG | GG | GG | AG | GG |
| A07 | F07_10 | AA | AA | AA | AA | AG |
| | F07_02 | AA | GG | GG | AG | AG |
| A08 | F08_17 | AA | CC | CC | CA | CA |
| | F08_08 | GG | TT | TT | TG | TG |
| A09 | F09_03 | AA | AA | AA | AA | AA |
| | F09_04 | AG | GG | AG | AG | AG |
| | F09_14 | TT | CC | CC | CT | CT |
| | F09_09 | GG | TT | TT | TT | GT |
| A10 | F10_02 | GG | GG | GG | GG | GG |
| | F10_03 | GG | GG | GG | GG | GG |

○ 채종포 A의 경우 24개의 마커중에서 14개의 마커가 이형접합이 아닌 부모의 유전형을 따라가는 것을 볼 수 있었다. 채종포 B의 경우 24개의 마커 중 4개의 마커에서 부모와 같은 유전형을 나타냈다. 채종포 C의 경우 F07_10 마커에서 양측 부모의 유전형이 AA인데 반해 F₁에서는 AG를 보였다. 이는 부모에 없는 유전형을 나타내는 결과이다.

○ 이를 통하여 본 24개의 marker는 염색체별 marker를 세트화 하여 전체 유전체 수준으로 유전형을 분석 할 수 있어 한 유전자위에 편향된 순도 검정 결과를 얻는 문제점을 보완 할 수 있으며, 다형성이 높은 marker로 이루어졌기 때문에 경제적, 시간적 효율을 확보할 수 있다. 또한 F₁순도 검정뿐만 아니라 모본의 자가합성이나 원종의 오염을 선별해 낼 수 있는 장점이 있다.

○ NGS를 통해 대량의 SNP를 생산하여 이를 육종에 적용하는 시스템이 발달하고 있다. 하지만 배추의 육종에서는 SNP를 이용한 시스템은 미비한 상태로 남아 있다. 이를 극복하기 위해 SNP를 이용한 대량 분석 시스템 플랫폼 중 Fluidigm SNP Type Assay 선택하여 배추 육종에 적합한 시스템을 만들고자 하였다. Fluidigm SNP Type Assay를 이용한 SNP genotyping 방식은 기존의 SSR 방식의 gel-based 분석법에 비해 자동화가 가능하고 분석이 용이하며 시간을 절감할 수 있어, high-throughput에 적합하다. 또한 SSR marker의 경우 육종회사에서 품종 육성에 사용되는 계통간의 유전적 거리가 가까운 경우 다형성을 나타내는 marker를 대량으로 선별하기 어렵다는 문제점이 있다. 이에 비해 SNP는 유전체 전역에 가장 많이 분포하는 다형성 지표이다. 기존의 marker로는 선별하기 힘든 background selection 및 F₁순도 검정에 적합하다. 본 연구에서는 NGS를 통하여 생산한 SNP marker의 정확성을 검증하고, 선별한 SNP를 실제 육종에 적용 할 수 있는지 판단하기 위해 육종회사 소재를 통하여 Fluidigm SNP Type Assay system를 구축하였다.

○ 총 177개의 SNP marker를 선별하였으며 그 중 MABC를 위한 96개의 marker set과 F₁순도 검정을 위한 24개의 marker set를 선별하였다. 이는 배추의 여교배집단에서의 조기 고정 선별 계통을 효율적으로 선별 할 수 있을 뿐만 아니라 F₁ 품종의 순도 검정에 매우 유용하게 활용될 수 있는 기술로 판단된다.

사. 개발된 마커를 이용한 육종 소재에 적용

○ MABC 및 순도 검정으로 개발된 마커(192개)를 이용하여 한국종묘의 여교배 집단 BC₃S₁의 Background selection과 F₁순도 검정을 하였다. 이 집단은 뿌리혹병 저항성 육성 소재를 양성하기 위한 것으로 사용된 식물은 저항성이 있는 모계 5개와 저항성이 없는 모계 5개 그리고 부계 5개를 사용하였으며, 이로부터 파생된 BC₃S₁ 60개를 사용하였다. F₁순도 검정용으로 190개의 시료를 사용하였다. 총 54,720 rxn을 하였다(표 4).

표 4. MABC를 위한 Background selection과 F₁ 순도 검정 결과

| Sample ID | Sample type | Purity score | | |
|-----------|--------------------------------|--------------|-------|--------|
| | | Low | High | Mean |
| A17 | P ₂ | 98.95 | 100 | 99.6 |
| mab20-01 | P ₁ | 100 | 100 | 100 |
| mab20-02 | BC ₃ S ₁ | 98.43 | 99.47 | 98.95 |
| mab20-03 | BC ₃ S ₁ | 97.91 | 98.95 | 98.54 |
| mab20-04 | BC ₃ S ₁ | 96.87 | 97.91 | 97.08 |
| mab20-05 | BC ₃ S ₁ | 96.87 | 97.91 | 97.6 |
| mab20-06 | BC ₃ S ₁ | 97.91 | 100 | 99.16 |
| mab20-07 | BC ₃ S ₁ | 98.43 | 100 | 99.37 |
| mab20-08 | P ₁ | 94.27 | 95.31 | 94.53 |
| mab20-09 | BC ₃ S ₁ | 93.22 | 94.27 | 93.7 |
| mab20-10 | BC ₃ S ₁ | 92.7 | 93.75 | 93.125 |
| mab20-11 | BC ₃ S ₁ | 91.14 | 92.18 | 91.35 |
| mab20-12 | BC ₃ S ₁ | 92.7 | 93.22 | 93.12 |
| mab20-13 | BC ₃ S ₁ | 93.75 | 94.79 | 94.16 |
| mab20-14 | BC ₃ S ₁ | 92.7 | 94.27 | 93.33 |
| mab20-15 | F ₁ | 69.27 | 100 | 97.94 |
| mab20-17 | F ₁ | 68.2 | 98.4 | 94.6 |
| mab20-16 | F ₁ | 93.75 | 94.27 | 94.18 |
| mab20-18 | F ₁ | 92.19 | 93.23 | 92.71 |

2. 배추 NGS 데이터를 활용한 특정 패스웨이 유전자 상의 변이 대량 발굴

가. 애기장대에서의 카로티노이드 패스웨이 유전자 식별

○ 카로티노이드는 식물에서 광합성을 돕고 자외선의 유해 작용을 막는 식물 색소이며, 베타카로틴과 같은 일부 카로티노이드는 동물에서는 비타민 A의 모체로서 도움을 주고, 시력 과도 관계한다. 이와 같이 항산화성 물질 및 비타민 전구체로서의 특성을 갖는 카로티노이드는 애기장대(*Arabidopsis thaliana*)에서 패스웨이 기반 연구가 활발하게 이루어졌다. KEGG 데이터베이스에서의 검색(그림 5) 및 관련된 문헌 수집을 통해 총 54개의 카로티노이드 합성에 관계된 애기장대 유전자를 확인할 수 있었다(표 5).

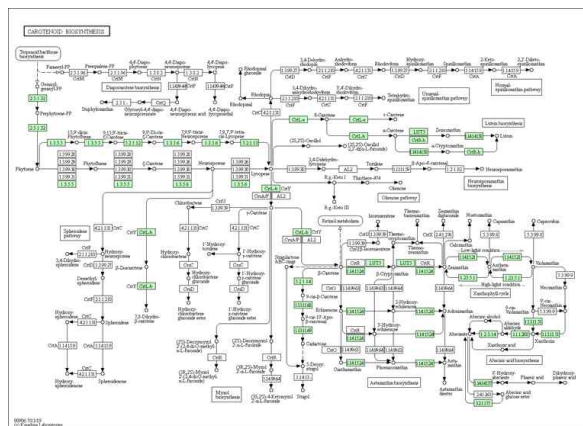


그림 5. KEGG에서의 애기장대 카로티노이드 패스웨이 검색 결과

표 5. 문헌조사를 통해 식별한 카로티노이드 패스웨이 유전자

| # | TAIR ID | Related pathway | Symbol | Protein length (aa) |
|----|-----------|-------------------------------|--|---------------------|
| 1 | AT1G49530 | MEP pathway to GGPP genes | GERANYLGERANYL PYROPHOSPHATE SYNTHASE 6 (GGPS6) | 336 |
| 2 | AT1G63970 | MEP pathway to GGPP genes | ISOPRENOID F (ISPF) | 231 |
| 3 | AT2G02500 | MEP pathway to GGPP genes | ISOPRENOID D (ISPD) | 302 |
| 4 | AT2G18620 | MEP pathway to GGPP genes | GERANYLGERANYL PYROPHOSPHATE SYNTHASE 2 (GGPS2) | 347 |
| 5 | AT2G18640 | MEP pathway to GGPP genes | GERANYLGERANYL PYROPHOSPHATE SYNTHASE 4 (GGPS4) | 372 |
| 6 | AT2G23800 | MEP pathway to GGPP genes | GERANYLGERANYL PYROPHOSPHATE SYNTHASE 2 (GGPS2) | 376 |
| 7 | AT2G26930 | MEP pathway to GGPP genes | 4-(CYTIDINE 5'-PHOSPHO)-2-C-METHYL-D-ERITHRITOL KINASE (CDPMEK) | 383 |
| 8 | AT3G02780 | MEP pathway to GGPP genes | ISOPENTENYL PYROPHOSPHATE:DIMETHYLALLYL PYROPHOSPHATE ISOMERASE 2 (IPP2) | 284 |
| 9 | AT3G14510 | MEP pathway to GGPP genes | - | 284 |
| 10 | AT3G14530 | MEP pathway to GGPP genes | GERANYL-FARNESYL PYROPHOSPHATE SYNTHASE 1 (GFPPS1) | 360 |
| 11 | AT3G14550 | MEP pathway to GGPP genes | GERANYLGERANYL PYROPHOSPHATE SYNTHASE 3 (GGPS3) | 360 |
| 12 | AT3G20160 | MEP pathway to GGPP genes | POLYPRENYL PYROPHOSPHATE SYNTHASE 2 (PPPS2) | 344 |
| 13 | AT3G29430 | MEP pathway to GGPP genes | GERANYL(GERANYL)DIPHOSPHATE SYNTHASE 9 (ATGGPS9) | 357 |
| 14 | AT3G32040 | MEP pathway to GGPP genes | GERANYL(GERANYL)DIPHOSPHATE SYNTHASE 1 (ATGGPS10) | 360 |
| 15 | AT4G15560 | MEP pathway to GGPP genes | CLOROPLASTOS ALTERADOS 1 (CLA1) | 717 |
| 16 | AT4G34350 | MEP pathway to GGPP genes | 4-HYDROXY-3-METHYLBUT-2-ENYL DIPHOSPHATE REDUCTASE (HDR) | 466 |
| 17 | AT4G36810 | MEP pathway to GGPP genes | GERANYLGERANYL PYROPHOSPHATE SYNTHASE 1 (GGPS1) | 371 |
| 18 | AT4G38460 | MEP pathway to GGPP genes | GERANYLGERANYL REDUCTASE (GGR) | 326 |
| 19 | AT5G16440 | MEP pathway to GGPP genes | ISOPENTENYL DIPHOSPHATE ISOMERASE 1 (IPP1) | 291 |
| 20 | AT5G60600 | MEP pathway to GGPP genes | 4-HYDROXY-3-METHYLBUT-2-ENYL DIPHOSPHATE SYNTHASE (HDS) | 741 |
| 21 | AT5G62790 | MEP pathway to GGPP genes | 1-DEOXY-D-XYLULOSE 5-PHOSPHATE REDUCTOISOMERASE (DXR) | 497 |
| 22 | AT1G06820 | Carotenoid biosynthetic genes | CAROTENOID ISOMERASE (CRTISO) | 595 |
| 23 | AT1G08550 | Carotenoid biosynthetic genes | NON-PHOTOCHEMICAL QUENCHING 1 (NPQ1) | 462 |
| 24 | AT1G10830 | Carotenoid biosynthetic genes | 15-CIS-ZETA-CAROTENE ISOMERASE (Z-ISO) | 367 |
| 25 | AT1G30100 | Carotenoid biosynthetic genes | NINE-CIS-EPOXYCAROTENOID DIOXYGENASE 5 (NCED5) | 589 |
| 26 | AT1G31800 | Carotenoid biosynthetic genes | CYTOCHROME P450, FAMILY 97, SUBFAMILY A, POLYPEPTIDE 3 (CYP97A3) | 595 |
| 27 | AT1G52340 | Carotenoid biosynthetic genes | ABA DEFICIENT 2 (ABA2) | 285 |
| 28 | AT1G57770 | Carotenoid biosynthetic genes | - | 574 |
| 29 | AT1G67080 | Carotenoid biosynthetic genes | ABSCISIC ACID (ABA)-DEFICIENT 4 (ABA4) | 220 |
| 30 | AT1G78390 | Carotenoid biosynthetic genes | NINE-CIS-EPOXYCAROTENOID DIOXYGENASE 9 (NCED9) | 657 |
| 31 | AT2G27150 | Carotenoid biosynthetic genes | ABSCISIC ALDEHYDE OXIDASE 3 (AAO3) | 1332 |
| 32 | AT2G44990 | Carotenoid biosynthetic genes | CAROTENOID CLEAVAGE DIOXYGENASE 7 (CCD7) | 618 |
| 33 | AT3G04870 | Carotenoid biosynthetic genes | ZETA-CAROTENE DESATURASE (ZDS) | 558 |
| 34 | AT3G10230 | Carotenoid biosynthetic genes | LYCOPENE CYCLASE (LYC) | 501 |
| 35 | AT3G14440 | Carotenoid biosynthetic genes | NINE-CIS-EPOXYCAROTENOID DIOXYGENASE 3 (NCED3) | 599 |
| 36 | AT3G24220 | Carotenoid biosynthetic genes | NINE-CIS-EPOXYCAROTENOID DIOXYGENASE 6 (NCED6) | 577 |
| 37 | AT3G53130 | Carotenoid biosynthetic genes | LUTEIN DEFICIENT 1 (LUT1) | 539 |
| 38 | AT4G14210 | Carotenoid biosynthetic genes | PHYTOENE DESATURASE 3 (PDS3) | 566 |
| 39 | AT4G15110 | Carotenoid biosynthetic genes | CYTOCHROME P450, FAMILY 97, SUBFAMILY B, POLYPEPTIDE 3 (CYP97B3) | 580 |
| 40 | AT4G18350 | Carotenoid biosynthetic genes | NINE-CIS-EPOXYCAROTENOID DIOXYGENASE 2 (NCED2) | 583 |
| 41 | AT4G19170 | Carotenoid biosynthetic genes | NINE-CIS-EPOXYCAROTENOID DIOXYGENASE 4 (NCED4) | 595 |
| 42 | AT4G25700 | Carotenoid biosynthetic genes | BETA-HYDROXYLASE 1 (BETA-OHASE 1) | 310 |
| 43 | AT4G32810 | Carotenoid biosynthetic genes | CAROTENOID CLEAVAGE DIOXYGENASE 8 (CCD8) | 570 |
| 44 | AT5G17230 | Carotenoid biosynthetic genes | PHYTOENE SYNTHASE (PSY) | 437 |
| 45 | AT5G52570 | Carotenoid biosynthetic genes | BETA-CAROTENE HYDROXYLASE 2 (BETA-OHASE 2) | 303 |
| 46 | AT5G57030 | Carotenoid biosynthetic genes | LUTEIN DEFICIENT 2 (LUT2) | 524 |
| 47 | AT5G67030 | Carotenoid biosynthetic genes | ABA DEFICIENT 1 (ABA1) | 667 |
| 48 | AT1G03055 | - | DWARF27 (D27) | 264 |
| 49 | AT1G52400 | - | BETA GLUCOSIDASE 18 (BGLU18) | 528 |
| 50 | AT2G29090 | - | CYTOCHROME P450, FAMILY 707, SUBFAMILY A, POLYPEPTIDE 2 (CYP707A2) | 482 |
| 51 | AT3G19270 | - | CYTOCHROME P450, FAMILY 707, SUBFAMILY A, POLYPEPTIDE 4 (CYP707A4) | 443 |
| 52 | AT3G63520 | - | - | 538 |
| 53 | AT4G19230 | - | CYTOCHROME P450, FAMILY 707, SUBFAMILY A, POLYPEPTIDE 1 (CYP707A1) | 484 |
| 54 | AT5G45340 | - | CYTOCHROME P450, FAMILY 707, SUBFAMILY A, POLYPEPTIDE 3 (CYP707A3) | 463 |

나. 배추에서의 카로티노이드 패스웨이 유전자 식별

○ 애기장대에서 확인한 카로티노이드 패스웨이 유전자와의 단백질 서열 상동성에 근거하여 배추의 표준유전체에서 카로티노이드 유전자를 식별하는 작업을 수행하였다. 애기장대 카로티노이드 단백질 서열을 subject로 배추에서 예측된 45,169개 단백질 서열은 query로 하여, BlastP 분석을 수행하여 배추의 orthologous 유전자를 탐색하였다. BLASTP 작업이 완

료된 이후 pident < 60, e-value < 1.0 x 10⁻¹⁰, query coverage > 60, subject coverage > 60 기준으로 필터링 및 정렬을 수행한 결과, 배추에서 총 79개의 카로티노이드 유전자를 탐색하였다(표 5).

표 5. 배추(*Brassica rapa* spp. *pekinensis*)에서 식별된 카로티노이드 패스웨이 유전자

| # | Gene ID | Gene description | Homologue gene IDs in <i>Brassica rapa</i> 3.0 annotation | Number of homologues in <i>Brassica rapa</i> |
|----|-----------|--|--|--|
| 1 | AT5G45340 | CYTOCHROME P450, FAMILY 707, SUBFAMILY A, POLYPEPTIDE 3 (CYP707A3) | BraA01g010660.3C, BraA02g031780.3C, BraA03g038810.3C, BraA04g020810.3C, BraA05g028020.3C, BraA06g042600.3C, BraA09g021180.3C | 7 |
| 2 | AT2G27150 | ABSCISIC ALDEHYDE OXIDASE 3 (AAO3) | BraA04g019480.3C, BraA06g022060.3C, BraA06g023170.3C, BraA10g003130.3C, BraA10g019670.3C | 5 |
| 3 | AT3G02780 | ISOPENTENYL PYROPHOSPHATE-DIMETHYLALLYL PYROPHOSPHATE ISOMERASE 2 (IPP2) | BraA01g044250.3C, BraA03g007480.3C, BraA05g041420.3C | 3 |
| 4 | AT3G14440 | NINE-CIS-EPOXYCAROTENOID DIOXYGENASE 3 (NCED3) | BraA01g037270.3C, BraA03g036470.3C, BraA05g032610.3C | 3 |
| 5 | AT3G63520 | - | BraA04g000040.3C, BraA09g053400.3C, BraA09g053420.3C | 3 |
| 6 | AT4G15560 | CLOROPLASTOS ALTERADOS 1 (CLA1) | BraA01g021140.3C, BraA03g039860.3C, BraA03g046310.3C | 3 |
| 7 | AT5G17230 | PHYTOENE SYNTHASE (PSY) | BraA02g006890.3C, BraA03g007870.3C, BraA10g022730.3C | 3 |
| 8 | AT5G52570 | BETA-CAROTENE HYDROXYLASE 2 (BETA-OHASE 2) | BraA01g016430.3C, BraA03g052720.3C, BraA10g010930.3C | 3 |
| 9 | AT1G52400 | BETA GLUCOSIDASE 18 (BGLU18) | BraA02g023150.3C, BraA05g017770.3C | 2 |
| 10 | AT1G63970 | ISOPRENOID F (ISPF) | BraA09g013520.3C, BraA09g014810.3C | 2 |
| 11 | AT1G78390 | NINE-CIS-EPOXYCAROTENOID DIOXYGENASE 9 (NCED9) | BraA02g024670.3C, BraA07g041520.3C | 2 |
| 12 | AT2G18620 | GERANYLGERANYL PYROPHOSPHATE SYNTHASE 2 (GGPPS2) | BraA03g036510.3C, BraA07g002120.3C | 2 |
| 13 | AT2G23800 | GERANYLGERANYL PYROPHOSPHATE SYNTHASE 2 (GGPS2) | BraA04g017500.3C, BraA09g054090.3C | 2 |
| 14 | AT3G20160 | POLYPRENYL PYROPHOSPHATE SYNTHASE 2 (PPPS2) | BraA03g039240.3C, BraA05g027110.3C | 2 |
| 15 | AT4G18350 | NINE-CIS-EPOXYCAROTENOID DIOXYGENASE 2 (NCED2) | BraA01g009810.3C, BraA03g048210.3C | 2 |
| 16 | AT4G19170 | NINE-CIS-EPOXYCAROTENOID DIOXYGENASE 4 (NCED4) | BraA01g010560.3C, BraA08g013370.3C | 2 |
| 17 | AT4G34350 | 4-HYDROXY-3-METHYLBUT-2-ENYL DIPHOSPHATE REDUCTASE (HDR) | BraA01g003910.3C, BraA08g016030.3C | 2 |
| 18 | AT5G57030 | LUTEIN DEFICIENT 2 (LUT2) | BraA03g012640.3C, BraA10g014920.3C | 2 |
| 19 | AT5G62790 | 1-DEOXY-D-XYLULOSE 5-PHOSPHATE REDUCTOISOMERASE (DXR) | BraA06g026430.3C, BraA09g007410.3C | 2 |
| 20 | AT5G67030 | ABA DEFICIENT 1 (ABA1) | BraA07g016890.3C, BraA09g009220.3C | 2 |
| 21 | AT1G03055 | DWARF27 (D27) | BraA09g065400.3C | 1 |
| 22 | AT1G06820 | CAROTENOID ISOMERASE (CRTISO) | BraA09g063710.3C | 1 |
| 23 | AT1G08550 | NON-PHOTOCHEMICAL QUENCHING 1 (NPQ1) | BraA06g005700.3C | 1 |
| 24 | AT1G10830 | 15-CIS-ZETA-CAROTENE ISOMERASE (Z-ISO) | BraA06g007670.3C | 1 |
| 25 | AT1G30100 | NINE-CIS-EPOXYCAROTENOID DIOXYGENASE 5 (NCED5) | BraA09g034930.3C | 1 |
| 26 | AT1G31800 | CYTOCHROME P450, FAMILY 97, SUBFAMILY A, POLYPEPTIDE 3 (CYP97A3) | BraA08g009810.3C | 1 |
| 27 | AT1G57770 | - | BraA09g018330.3C | 1 |

| | | | | |
|-------|-----------|--|------------------|----|
| 28 | AT1G67080 | ABSCISIC ACID (ABA)-DEFICIENT 4 (ABA4) | BraA02g017410.3C | 1 |
| 29 | AT2G02500 | ISOPRENOID D (ISPD) | BraA02g035640.3C | 1 |
| 30 | AT2G18640 | GERANYLGERANYL PYROPHOSPHATE SYNTHASE 4 (GGPS4) | BraA09g011510.3C | 1 |
| 31 | AT2G26930 | 4-(CYTIDINE 5'-PHOSPHO)-2-C-METHYL-D-ERITHRITOL KINASE (CDPMEK) | BraA07g017900.3C | 1 |
| 32 | AT2G44990 | CAROTENOID CLEAVAGE DIOXYGENASE 7 (CCD7) | BraA04g030840.3C | 1 |
| 33 | AT3G04870 | ZETA-CAROTENE DESATURASE (ZDS) | BraA05g040310.3C | 1 |
| 34 | AT3G10230 | LYCOPENE CYCLASE (LYC) | BraA05g036490.3C | 1 |
| 35 | AT3G14550 | GERANYLGERANYL PYROPHOSPHATE SYNTHASE 3 (GGPS3) | BraA01g002000.3C | 1 |
| 36 | AT3G24220 | NINE-CIS-EPOXYCAROTENOID DIOXYGENASE 6 (NCED6) | BraA07g009070.3C | 1 |
| 37 | AT3G29430 | GERANYL(GERANYL)DIPHOSPHATE SYNTHASE 9 (ATGGPS9) | BraA05g032530.3C | 1 |
| 38 | AT4G14210 | PHYTOENE DESATURASE 3 (PDS3) | BraA04g007910.3C | 1 |
| 39 | AT4G15110 | CYTOCHROME P450, FAMILY 97, SUBFAMILY B, POLYPEPTIDE 3 (CYP97B3) | BraA08g011140.3C | 1 |
| 40 | AT4G19230 | CYTOCHROME P450, FAMILY 707, SUBFAMILY A, POLYPEPTIDE 1 (CYP707A1) | BraA03g048740.3C | 1 |
| 41 | AT4G32810 | CAROTENOID CLEAVAGE DIOXYGENASE 8 (CCD8) | BraA01g005230.3C | 1 |
| 42 | AT4G36810 | GERANYLGERANYL PYROPHOSPHATE SYNTHASE 1 (GGPS1) | BraA08g021280.3C | 1 |
| 43 | AT4G38460 | GERANYLGERANYL REDUCTASE (GGR) | BraA01g000080.3C | 1 |
| 44 | AT5G16440 | ISOPENTENYL DIPHOSPHATE ISOMERASE 1 (IPP1) | BraA03g031130.3C | 1 |
| 45 | AT5G60600 | 4-HYDROXY-3-METHYLBUT-2-ENYL DIPHOSPHATE SYNTHASE (HDS) | BraA10g018340.3C | 1 |
| 46 | AT1G49530 | GERANYLGERANYL PYROPHOSPHATE SYNTHASE 6 (GGPS6) | - | 0 |
| 47 | AT1G52340 | ABA DEFICIENT 2 (ABA2) | - | 0 |
| 48 | AT2G29090 | CYTOCHROME P450, FAMILY 707, SUBFAMILY A, POLYPEPTIDE 2 (CYP707A2) | - | 0 |
| 49 | AT3G14510 | - | - | 0 |
| 50 | AT3G14530 | GERANYL-FARNESYL PYROPHOSPHATE SYNTHASE 1 (GFPPS1) | - | 0 |
| 51 | AT3G19270 | CYTOCHROME P450, FAMILY 707, SUBFAMILY A, POLYPEPTIDE 4 (CYP707A4) | - | 0 |
| 52 | AT3G32040 | GERANYL(GERANYL)DIPHOSPHATE SYNTHASE 1 (ATGGPS10) | - | 0 |
| 53 | AT3G53130 | LUTEIN DEFICIENT 1 (LUT1) | - | 0 |
| 54 | AT4G25700 | BETA-HYDROXYLASE 1 (BETA-OHASE 1) | - | 0 |
| Total | | | | 79 |

다. 카로티노이드 함량에 대한 엘리트 라인에서의 변이 데이터 추출

○ 카로티노이드 함량에 대하여 엘리트 라인으로 선발된 세 개의 배추 계통(L210-3, L226-1, L272-1)에 대하여 카로티노이드 패스웨이 상의 변이를 산출하기 위하여 Illumina NovaSeq 6000 sequencing platform을 통해 생산된 sequencing 데이터를 활용하였다(표 23). 3개 계통에 대하여 총 115.2 Gbp의 sequence 데이터가 생산이 되었으며 배추 유전체에 대하여 계통별로 평균 76.8X의 sequencing depth를 나타내었다. fastqc 분석을 통한 raw read의 quality에 대한 평가를 수행하였다(그림 5).

표 6. 배추 카로티노이드 엘리트 라인 3개 계통의 Sequencing 결과

| # | Accession ID | Carotenoid contents | Sequencing platform | Sequencing size (bp) | Total number of sequence | Sequencing depth(X) |
|---|--------------|---------------------|---------------------|------------------------------------|--------------------------|---------------------|
| 1 | L210-3 | High | NovaSeq 6000 | 39,564,031,856 | 262,013,456 | 79.13 |
| 2 | L226-1 | High | | 40,347,699,810 | 267,203,310 | 80.7 |
| 3 | L272-1 | Low | | 35,284,745,008 | 233,673,808 | 70.57 |
| | | | | Total: 115,196,476,674 (115.2 Gbp) | Total: 762,890,574 | Average: 76.8 |

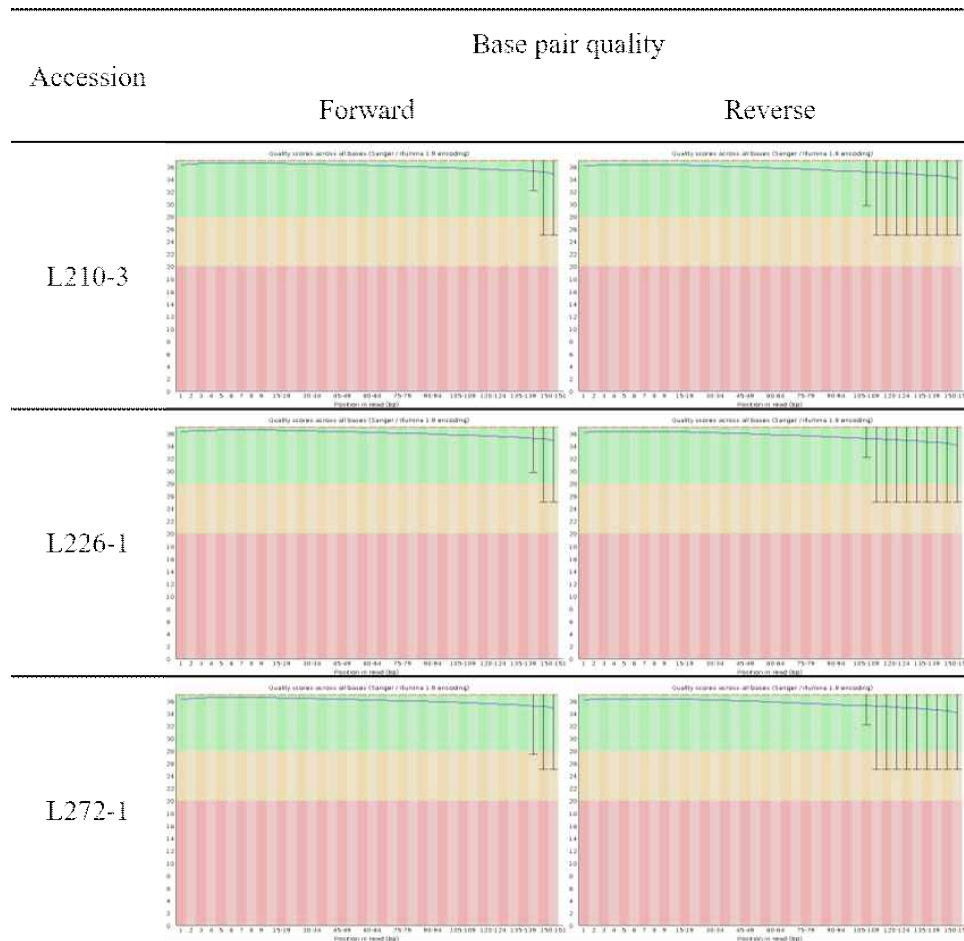


그림 5. 배추 카로티노이드 엘리트 라인 3개 계통의 raw sequence read의 quality

○ 1차 QC 이후 Trimmomatic을 통해 sequence의 필터링 및 adaptor 서열을 제거하였으며 3개 계통 각각에서 얻은 filtered read sequence 데이터를 배추 표준유전체(ver 3.0)에 alignment를 수행하여 sam 파일을 구성하였다. 이후, GATK 파이프라인을 활용하여 sam 파일의 bam 파일로 전환 및 편집을 수행하였으며 bcftools를 활용하여 계통별로 변이 정보를 생산한 이후 filtering 및 joint variant calling을 수행하여 3개 계통에 대해서 1,913,911개의 변이 위치가 나타나는 것을 확인하였다. 해당 변이 정보를 filtering하고 배추에서 식별한 카로티노이드 유전자가 위치한 영역에서의 변이만을 최종적으로 식별하였다(표 7).

표 7. 배추 카로티노이드 엘리트 라인 3개 계통의 SNP calling 결과

| # | Filtering condition | Number of variants | Proportion(%) |
|---|-------------------------------|--------------------|---------------|
| 1 | Raw variant | 1,913,911 | 100 |
| 2 | Bi-allelic site | 1,908,105 | 99.7 |
| 3 | Homozygous site only | 1,723,865 | 90.07 |
| 4 | LD pruning | 31,576 | 1.65 |
| 5 | Carotenoid gene coding region | 126 | 0.01 |

라. 마커 이전을 위한 마커의 육종 소재 적용

○ 수집 및 생산된 유용형질 관련 마커를 이전하기 전에 마커를 육종소재에 적용함으로써 마커의 실용성을 확인하였다. 육종소재는 국내 한국종묘(232개), 대일종묘(22개), 코레곤종묘(34개)의 육종 소재를 사용하였다. 사용된 마커는 뿌리혹병 저항성 선발 마커로 각각의 회사마다 4개의 유전자좌에서 16개의 유전자좌를 검정하였다(그림 6).

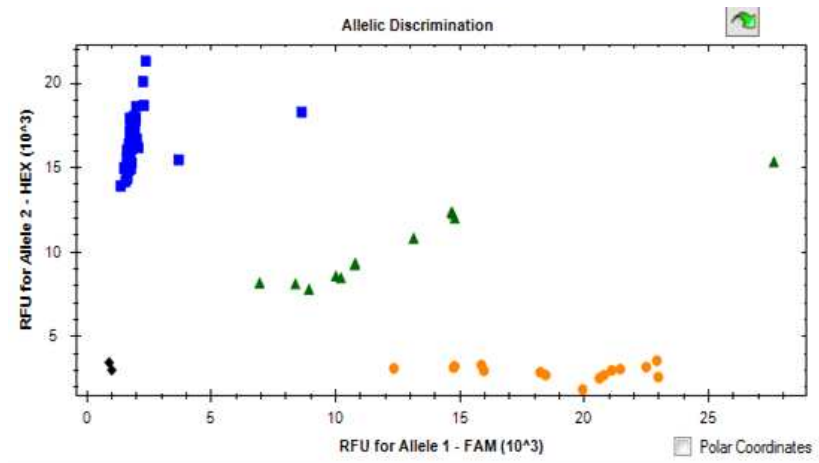


그림 6. 뿌리혹병 저항성 마커 검정

제7절 종자산업 발전 및 마케팅 전략 구축을 위한 기반 연구

1. 종자기업의 품목 전문화를 통한 종자기업 경쟁력 강화 및 종자산업 발전을 위한 종자업체 현황조사

가. 종자업체의 산업 현황

(1) 종자업체 등록 현황

○ 2015년 기준 종자업체 등록현황은 총 1,699개소로서, 이는 2014년보다 약 14% 증가한 수치임. 종자업체 등록업체는 최근 높은 증가추세에 있으며 2006년 대비 약 161%, 5개년 평균 약 16%의 증가율을 보이고 있음. 과수업이 29.1%로 가장 많은 비중을 차지하고 있으며 기타(27.3%), 채소(16.3%), 화훼(13.6%), 버섯(7.7%), 빵(2.1%) 순으로 나타남.

○ 종자업체 등록업체가 매년 늘어나고 있는 반면 채소종자 등록업체의 경우 매년 비중이 줄어들고 있는 상황임. 특히 2015년 종자업체 등록현황 중 채소 종자업체 등록업체의 비중은 약 16.3%로서 2006년 22.1%보다 약 4.8%p 감소함.

표 1. 품목의 연도별 종자업체 등록현황

단위 : 개소

| 연도 | 계 | 식량 | 과수 | 채소 | 화훼 | 버섯 | 빵 | 기타 |
|------|------------------|-------------|---------------|---------------|---------------|---------------|-------------|---------------|
| 2006 | 736 (100.0) | 18 (2.4) | 235 (31.9) | 163 (22.1) | 110 (14.9) | 115 (15.6) | 26 (3.5) | 69 (9.4) |
| 2007 | 813 (100.0) | 25 (3.1) | 249 (30.6) | 182 (22.4) | 126 (15.5) | 123 (15.1) | 26 (3.2) | 82 (10.1) |
| 2008 | 871 (100.0) | 33 (3.8) | 257 (29.5) | 189 (21.7) | 139 (16.0) | 126 (14.5) | 28 (3.2) | 99 (11.4) |
| 2009 | 819 (100.0) | 37 (4.5) | 250 (30.5) | 173 (21.1) | 121 (14.8) | 110 (13.4) | 30 (3.7) | 98 (12.0) |
| 2010 | 924 (100.0) | 32 (3.5) | 283 (30.6) | 181 (19.6) | 149 (16.1) | 121 (13.1) | 29 (3.1) | 129 (14.0) |
| 2011 | 932 (100.0) | 39 (4.2) | 300 (32.2) | 185 (19.8) | 139 (14.9) | 112 (12.0) | 30 (3.2) | 127 (13.6) |
| 2012 | 1,073 (100.0) | 39 (3.6) | 328 (30.6) | 202 (18.8) | 160 (14.9) | 118 (11.0) | 33 (3.1) | 193 (18.0) |
| 2013 | 1,279 (100.0) | 49 (3.8) | 393 (30.7) | 224 (17.5) | 195 (15.2) | 108 (8.4) | 46 (3.6) | 264 (20.6) |
| 2014 | 1,491 (100.0) | 58 (3.9) | 445 (29.8) | 255 (17.1) | 214 (14.4) | 125 (8.4) | 47 (3.2) | 347 (23.3) |
| 2015 | 1,699 (100.0) | 69 (4.1) | 494 (29.1) | 277 (16.3) | 231 (13.6) | 130 (7.7) | 35 (2.1) | 463 (27.3) |

자료 : 국립종자원 홈페이지(http://www.seed.go.kr/distribution/sail/inquiry_condition.jsp) 접속일 : 17.02.27)

주 : ()안은 각 품목이 차지하는 비중임.

○ 시도별 종자업체 등록현황을 살펴보면 경기도가 대부분의 품목에서 가장 높은 비중을 차지하며 전체 18.8%의 높은 비중을 차지하고 있으며, 다음으로 경상북도, 각각 11.6%를 차지하고 있음.

표 2. 품목의 시도별 종자업 등록현황

단위 : 개소

| 구분 | 계 | 식량 | 과수 | 채소 | 화훼 | 버섯 | 빵 | 기타 |
|-------|------------------|---------------|----------------|----------------|----------------|----------------|---------------|----------------|
| 합 계 | 1,699 (100.0) | 69 (100.0) | 494 (100.0) | 277 (100.0) | 231 (100.0) | 130 (100.0) | 35 (100.0) | 463 (100.0) |
| 서울특별시 | 9(0.5) | 1(1.4) | 1(0.2) | 5(1.8) | 0(0.0) | 0(0.0) | 0(0.0) | 2(0.4) |
| 부산광역시 | 21(1.2) | 1(1.4) | 3(0.6) | 10(3.6) | 4(1.7) | 0(0.0) | 1(2.9) | 2(0.4) |
| 대구광역시 | 9(0.5) | 0(0.0) | 2(0.4) | 2(0.7) | 0(0.0) | 3(2.3) | 0(0.0) | 2(0.4) |
| 인천광역시 | 3(0.2) | 0(0.0) | 0(0.0) | 0(0.0) | 2(0.9) | 0(0.0) | 0(0.0) | 1(0.2) |
| 광주광역시 | 13(0.8) | 0(0.0) | 3(0.6) | 4(1.4) | 4(1.7) | 0(0.0) | 0(0.0) | 2(0.4) |
| 대전광역시 | 13(0.8) | 0(0.0) | 4(0.8) | 1(0.4) | 2(0.9) | 0(0.0) | 0(0.0) | 6(1.3) |
| 울산광역시 | 2(0.1) | 0(0.0) | 0(0.0) | 0(0.0) | 0(0.0) | 1(0.8) | 0(0.0) | 1(0.2) |
| 세종시 | 65(3.8) | 0(0.0) | 8(1.6) | 3(1.1) | 2(0.9) | 1(0.8) | 1(2.9) | 50(10.8) |
| 경기도 | 319(18.8) | 20(29.0) | 41(8.3) | 66(23.8) | 85(36.8) | 23(17.7) | 0(0.0) | 84(18.1) |
| 강원도 | 88(5.2) | 7(10.1) | 15(3.0) | 7(2.5) | 7(3.0) | 6(4.6) | 1(2.9) | 45(9.7) |
| 충청북도 | 212(12.5) | 11(15.9) | 95(19.2) | 21(7.6) | 13(5.6) | 13(10.0) | 3(8.6) | 56(12.1) |
| 충청남도 | 148(8.7) | 4(5.8) | 40(8.1) | 17(6.1) | 32(13.9) | 18(13.8) | 0(0.0) | 37(8.0) |
| 전라북도 | 158(9.3) | 5(7.2) | 37(7.5) | 22(7.9) | 18(7.8) | 3(2.3) | 15(42.9) | 58(12.5) |
| 전라남도 | 197(11.6) | 8(11.6) | 71(14.4) | 32(11.6) | 15(6.5) | 19(14.6) | 6(17.1) | 46(9.9) |
| 경상북도 | 249(14.7) | 5(7.2) | 126(25.5) | 32(11.6) | 20(8.7) | 27(20.8) | 3(8.6) | 36(7.8) |
| 경상남도 | 131(7.7) | 3(4.3) | 21(4.3) | 47(17.0) | 18(7.8) | 15(11.5) | 4(11.4) | 23(5.0) |
| 제주도 | 62(3.6) | 4(5.8) | 27(5.5) | 8(2.9) | 9(3.9) | 1(0.8) | 1(2.9) | 12(2.6) |

자료 : 국립종자원 홈페이지(http://www.seed.go.kr/distribution/sail/inquiry_condition.jsp 접속일 : 17.02.27)

주 : ()안은 각 품목이 시도별로 차지하는 비중임.

(2) 품종보호 출원 및 등록 현황

○ 품종보호제도는 1998년 도입 이래 2016년 기준 총 8,867건이 출원되어 6,392건이 등록되었다. 2011년~2015년 누계기준 출원건수는 EU, 미국, 중국, 우크라이나, 일본, 네덜란드에 이어 세계에서 7번째로 많은 것으로 나타남(사단법인 한국 종자협회, 2017).

○ 2016년까지 전체 출원실적을 살펴보면 화훼류가 52.5%로 가장 높은 비중을 차지하고 있으며 채소류 21.7%, 식량작물 13.1%, 과수류 6.2% 등의 순으로 나타났다. 최근 채소류의 출원비중이 매년 늘어나는 추세임.

○ 채소류의 경우 국내의 소비량이 집중되고 있는 고추, 배추, 무, 수박 등의 품종 출원은 50% 이상일 정도로 높은 비중을 차지하고 있다. 특히 이들 품목은 현재 GSP 사업에서 품종 개발을 집중적으로 연구하고 있으며, 2017년 GSP 사업이 2단계로 들어가면서 좀 더 활발하게 신품종 개발이 이루어 질 것으로 기대됨.

표 3. 부류별 품종보호 출원 및 등록 현황 (2016.12.31. 현재)

단위 : 건수

| 작물명 | 출원실적 | | | | | | 등록실적 |
|------|-------|-------|------|------|------|------|-------|
| | 계 | 12까지 | 2013 | 2015 | 2015 | 2016 | |
| 계 | 8,867 | 6,233 | 599 | 621 | 757 | 657 | 6,392 |
| 식량작물 | 1,166 | 893 | 70 | 62 | 80 | 61 | 954 |
| 과수류 | 547 | 376 | 37 | 45 | 38 | 51 | 337 |
| 화훼류 | 4,658 | 3,342 | 282 | 322 | 391 | 311 | 3,493 |
| 특용 | 330 | 244 | 25 | 20 | 25 | 16 | 244 |
| 사료 | 67 | 45 | 3 | 3 | 11 | 5 | 42 |
| 버섯 | 178 | 122 | 15 | 14 | 17 | 10 | 130 |
| 채소류 | 1,921 | 1,211 | 157 | 155 | 195 | 203 | 1,192 |

자료 : 국립종자원 홈페이지(http://www.seed.go.kr/protection/situation/register_01.jsp 접속일 : 17.02.27)

표 4. 채소류의 품종보호 출원 및 등록현황

단위 : 건수

| 작물명 | 출원실적 | | | | | | 등록실적 |
|-----|-------|-------|------|------|------|------|-------|
| | 계 | 12까지 | 2013 | 2014 | 2015 | 2016 | |
| 채소류 | 1,921 | 1,211 | 157 | 155 | 195 | 203 | 1,192 |
| 고추 | 457 | 270 | 31 | 50 | 52 | 54 | 274 |
| 무 | 205 | 146 | 16 | 10 | 18 | 15 | 132 |
| 배추 | 236 | 162 | 15 | 17 | 21 | 21 | 159 |
| 수박 | 160 | 109 | 13 | 7 | 15 | 16 | 84 |
| 기타 | 863 | 524 | 82 | 71 | 89 | 97 | 543 |

자료 : 국립종자원 홈페이지(http://www.seed.go.kr/protection/situation/register_01.jsp 접속일 : 17.02.27)

(3) 매출규모별 종자업체 현황

○ 2016년 종자업체실태조사(2015년 기준)결과 종자업체는 1,207개로 나타났으며 이 중 5억 원 미만 매출규모 업체는 1,061개소로 87.9%의 비중을 차지하고 있음. 이어 5~15억 원 미만 이 7.3%, 15~40억 원이 3.4%, 40억 원 이상 업체가 1.4%로 대부분의 업체가 상대적으로 소 규모 인 것으로 나타남.

표 5. 매출규모별 종자업체 현황(2015년 기준)

단위 : 개소, %

| 구분 | 업체 수 | 비율 |
|---------|-------------|-------|
| 계 | 1,207 | 100.0 |
| 판매액 규모별 | 5억 원 미만 | 87.9 |
| | 5~15억 원 미만 | 7.3 |
| | 15~40억 원 미만 | 3.4 |
| | 40억 원 이상 | 1.4 |

자료 : 국립종자원 보도자료(2017. 4. 12)

주 : 2015년 기준 종자업체 등록업체는 1,699개소이나 폐업 등을 제외하고 1,405업체 대상으로 조사하였으며, 적격률 86.1%로 무응답 3개소를 제외한 1,207개소 업체를 대상으로 한 조사결과임.

(4) 종자업체의 종자 취급 현황

○ 전체 취급종자를 살펴보면 과수가 31.5%로 가장 많은 비중을 차지하고 있는 가운데 산림작물(19.4%), 채소(16.2%) 등의 순으로 나타남.

○ 판매액 규모별로 살펴보면 소규모 업체일수록 과수종자 취급비중이 높은 반면 규모가 커질수록 채소종자의 비중이 높아지는 것을 볼 수 있음.

표 5. 종자업체 규모별 취급종자(2015년 기준)

단위 : 개수, %

| 구분 | 식량작물 | 채소 | 과수 | 화훼 | 버섯 | 산림작물 | 특용/사료/기타 | 계 | |
|------------|-------------|---------------|---------------|---------------|---------------|---------------|---------------|------------------|----------------|
| 전체 | 59 (3.9) | 247 (16.2) | 482 (31.5) | 185 (12.1) | 94 (6.2) | 296 (19.4) | 165 (10.8) | 1,528 (100.0) | |
| 판매액 규모별 | 소규모 | 42 (3.2) | 177 (13.4) | 464 (35.2) | 159 (12.1) | 60 (4.6) | 269 (20.4) | 1,318 (100.0) | |
| | 중소규모 | 8 (6.7) | 34 (28.3) | 10 (8.3) | 16 (13.3) | 25 (20.8) | 18 (15.0) | 9 (7.5) | 120 (100.0) |
| | 중규모 | 5 (8.1) | 23 (37.1) | 6 (9.7) | 4 (6.5) | 9 (14.5) | 8 (12.9) | 7 (11.3) | 62 (100.0) |
| | 대규모 | 4 (14.3) | 13 (46.4) | 2 (7.1) | 6 (21.4) | 0 (0.0) | 1 (3.6) | 2 (7.1) | 28 (100.0) |

자료 : 국립종자원 보도자료(2017. 4. 12)

주 : 1. 2015년 기준 종자업 등록업체는 1,699개소이나 폐업 등을 제외하고 1,405업체 대상으로 조사하였으며, 적격률 86.1% 무응답 3개소를 제외한 1,207개소 업체를 대상으로 한 중복응답 조사결과임.

2. ()안은 비중임.

○ 종자의 부류별 취급하는 품목은 화훼류가 평균 15.6개로 가장 많았고 다음으로 채소 6.6개, 과수 3.2개 순으로 나타남. 채소종자의 경우 판매액 규모가 클수록 취급하는 작목의 수가 높은 것으로 나타남.

표 6. 종자 부류별 평균 취급 작목 수(2015년 기준)

단위 : 개

| 구분 | 식량작물 | 채소 | 과수 | 화훼 | 버섯 | 산림작물 | 특용/사료/기타 | |
|------------|------|-----|------|------|-------|------|----------|------|
| 전체 평균 | 2.2 | 6.6 | 3.2 | 15.6 | 2.3 | 3.8 | 4.3 | |
| 판매액 규모별 | 소규모 | 2.2 | 4.0 | 3.1 | 10.8 | 2.5 | 3.7 | 3.4 |
| | 중소규모 | 2.5 | 11.0 | 7.4 | 34.3 | 1.9 | 4.5 | 18.8 |
| | 중규모 | 2.0 | 11.0 | 3.7 | 120.3 | 1.9 | 4.9 | 4.4 |
| | 대규모 | 1.5 | 22.7 | 5.5 | 23.2 | 0.0 | 5.0 | 8.0 |

자료 : 국립종자원 보도자료(2017. 4. 12)

주 : 2015년 기준 종자업 등록업체는 1,699개소이나 폐업 등을 제외하고 1,405업체 대상으로 조사하였으며, 적격률 86.1%로 무응답 3개소를 제외한 1,207개소 업체를 대상으로 한 조사결과임.

표 7. 종자업체의 상시근무자 중 평균 종사자 수 - 직무별, 2015년 기준 단위 : 명

| 구분 | 육종 | 국외영업 | 국내외영업 | 관리 및 기타 | 계 | |
|---------|-----------|------------|-----------|------------|------------|-------------|
| 전체 | 0.7(11.0) | 0.0(0.7) | 1.3(20.5) | 4.3(67.8) | 6.4(100.0) | |
| 판매액 규모별 | 소규모 | 0.4(8.3) | 0.0(0.0) | 0.9(18.8) | 3.6(75.0) | 4.8(100.0) |
| | 중소규모 | 0.9(12.9) | 0.1(1.4) | 1.9(27.1) | 4.2(60.0) | 7.0(100.0) |
| | 중규모 | 2.3(18.3) | 0.2(1.6) | 3.5(27.8) | 6.6(52.4) | 12.6(100.0) |
| | 대규모 | 15.5(18.3) | 2.1(2.5) | 21.3(25.1) | 46.0(54.2) | 84.8(100.0) |

주1. 판매액 규모별로 소규모는 5억 원 미만, 중소규모는 5억 원~15억 원 미만, 중규모는 15억 원~40억 원 미만, 대규모는 40억 원 이상 기준임.

주2. ()안은 비중임.

자료 : 2016 종자업 실태조사(2015년 기준). 농림축산식품부, 2017.

나. 종자 생산 현황

(1) 채소종자 생산 및 매출현황

○ 2016년기준 채소종자 생산량은 약 211톤으로 이는 전년도 보다 약 22.6% 감소한 수치임. 최근 우리나라 채소 생산량과 더불어 채소종자 생산량은 감소추세에 있음. 채소종자 중 가장 많은 비중을 차지하는 품목은 배추로서, 2016년 전체 채소 생산량 중 약 53.9%의 비중을 차지함. 다음으로는 고추가 35.4%의 비중을 차지하고 있음.

표 8. 채소종자 생산현황

단위 : Kg

| 구분 | 2012 | 2013 | 2014 | 2015 | 2016 |
|-----|--------------------|--------------------|--------------------|--------------------|--------------------|
| 채소류 | 336,045 (100.0) | 326,451 (100.0) | 405,513 (100.0) | 272,804 (100.0) | 211,182 (100.0) |
| 고추 | 1,932 (0.6) | 2,118 (0.6) | 556 (0.1) | 675 (0.2) | 687 (0.3) |
| 배추 | 132,435 (39.4) | 141,214 (43.3) | 180,913 (44.6) | 118,387 (43.4) | 74,763 (35.4) |
| 무 | 163,718 (48.7) | 155,302 (47.6) | 192,914 (47.6) | 127,682 (46.8) | 113,890 (53.9) |
| 수박 | 965 (0.3) | 1,532 (0.5) | 1,596 (0.4) | 1,315 (0.5) | 1,230 (0.6) |
| 기타 | 36,995 (11.0) | 26,285 (8.1) | 29,534 (7.3) | 24,745 (9.1) | 20,612 (9.8) |

주1. (사)한국종자협회 회원사 제출 자료를 기초로 함.

2. ()안은 전체 채소류에서 해당품목이 차지하는 비중임.

자료 : (사)한국종자협회 홈페이지(www.kosaseed.or.kr 접속일 : 17.03.03)

○ 2016년도 채소종자 매출액은 약 260,500 백만 원으로서 이는 전년도보다 약 1.6% 감소한 수치임. 채소종자 매출액 중 고추, 배추, 무, 수박이 각각 17.5%, 6.6%, 16.1%, 4.6%를 차지하고 있음.

표 9. 채소종자 매출현황

단위 : 백만 원

| 구분 | 2012 | 2013 | 2014 | 2015 | 2016 |
|-----|--------------------|--------------------|--------------------|--------------------|--------------------|
| 채소류 | 239,451 (100.0) | 261,251 (100.0) | 253,600 (100.0) | 264,760 (100.0) | 260,500 (100.0) |
| 고추 | 41,407 (17.3) | 49,009 (18.8) | 44,282 (17.5) | 50,416 (19.0) | 45,603 (17.5) |
| 배추 | 18,947 (7.9) | 19,519 (7.5) | 18,610 (7.3) | 17,384 (6.6) | 17,261 (6.6) |
| 무 | 38,445 (16.1) | 45,804 (17.5) | 43,247 (17.1) | 45,926 (17.3) | 42,070 (16.1) |
| 수박 | 11,125 (4.6) | 12,317 (4.7) | 12,611 (5.0) | 12,764 (4.8) | 11,992 (4.6) |
| 기타 | 129,527 (54.1) | 134,602 (51.5) | 134,850 (53.2) | 138,270 (52.2) | 143,574 (55.1) |

주 : (사)한국종자협회 회원사 제출 자료를 기초로 함.

자료 : (사)한국종자협회 홈페이지(www.kosaseed.or.kr 접속일 : 17.03.03)

(2) 부류별 종자시장 규모

○ 2015년 기준 우리나라의 종자시장 규모는 약 5,008억 원의 규모이다. 이 중 채소가 3,147억 원 규모로 가장 높은 비중(62.8%)을 차지하고 있으며 과수가 567억 원(11.3%), 화훼 486억 원(9.7%) 등의 순으로 나타남. 가장 많은 비중을 차지하고 있는 채소종자 산업의 경우 대부분 민간업체에 의해 공급이 이루어지는 구조를 취하고 있음(김설중, 2014).

표 10. 종자 부류별 종자시장 규모(2015년 기준)

단위 : 억 원, %

| 구분 | 합 계 | 식량 | 채소 | 과수 | 화훼 | 버섯 | 산림 | 특용/사료 |
|--------|------------------|-------------|-----------------|---------------|--------------|--------------|--------------|--------------|
| 종자 판매액 | 5,008 (100.0) | 52 (1.0) | 3,147 (62.8) | 567 (11.3) | 486 (9.7) | 408 (8.1) | 200 (4.0) | 150 (3.0) |

자료 : 국립종자원 보도자료(2017. 4. 12)

주 : 1. 2015년 기준 종자업 등록업체는 1,699개소이나 폐업 등을 제외하고 1,405업체 대상으로 조사하였으며, 적격률 86.1% 무응답 3개소를 제외한 1,207개소 업체를 대상으로 한 조사결과임.

2. ()안은 해당 부류가 전체 판매액에서 차지하는 비중임.

(3) 나라별 종자시장 규모

○ EU(2013)에 따르면 2012년도 세계 종자시장 규모는 34,967 백만 유로 규모로 세계1위의 종자업체 Monsanto가 있는 미국이 26.7%로 가장 큰 비중을 차지하고 있음. 점유율 기준으로 미국에 이어 중국(22.1%), EU(20.3%)가 그 뒤를 잇고 있음. 미국, 중국, EU가 전세계 종자시장의 약 70%를 점유하고 있어 세계 시장을 주도하고 있음을 알 수 있음.

○ 이들 국가가 종자시장에서 차지하는 비중이 큰 이유는 GMO종자의 확산이 가장 큰 이유인 것으로 알려져 있음. 특히 중국의 경우 최근들어 종자시장의 규모가 급격히 늘어나고 있는데, 이는 중국의 경제성장으로 인한 소득 증가로 고품질 농산물의 수요 증가가 주 원인인 것으로 판단됨(김설중, 2014).

표 10. 나라별 국내 종자시장 규모

단위 : 백만 유로, %

| 구분 | 05 | 06 | 07 | 08 | 09 | 10 | 11 | 12 | 12년도 점유율 |
|-------|--------|--------|--------|--------|--------|--------|--------|--------|----------|
| 미국 | 4,582 | 5,575 | 6,202 | 8,159 | 8,603 | 9,052 | 8,621 | 9,340 | 26.7 |
| 중국 | 2,411 | 3,584 | 2,919 | 4,079 | 4,302 | 7,166 | 6,490 | 7,744 | 22.1 |
| EU | 4,903 | 4,972 | 5,839 | 5,849 | 6,511 | 6,974 | 6,968 | 7,106 | 20.3 |
| 브라질 | 965 | 1,195 | 1,459 | 1,360 | 1,434 | 1,509 | 1,886 | 2,043 | 5.8 |
| 캐나다 | 442 | 438 | 401 | 374 | 394 | 415 | 395 | 1,650 | 4.7 |
| 인도 | 452 | 1,035 | 1,094 | 1,020 | 1,075 | 1,509 | 1,437 | 1,557 | 4.5 |
| 일본 | 2,009 | 1,991 | 1,094 | 816 | 896 | 1,056 | 1,114 | 1,051 | 3.0 |
| 아르헨티나 | 748 | 741 | 693 | 501 | 498 | 453 | 542 | 771 | 2.2 |
| 터키 | 137 | 199 | 255 | 255 | 287 | 302 | 287 | 584 | 1.7 |
| 기타 | 3,148 | 2,870 | 2,752 | 2,599 | 2,709 | 2,619 | 2,888 | 3,121 | 8.9 |
| 계 | 19,827 | 22,600 | 22,711 | 25,012 | 26,710 | 31,054 | 30,626 | 34,967 | 100.0 |

자료 : The EU seed and plant material market in perspective : a focus on companies and market shares. EU, 2013. pp23.

- 주 1. 밭작물, 채소, 화훼종자를 포함한 수치임.
- 2. 씨감자는 제외함.

(4) EU의 종자시장 규모

○ EU 회원국의 종자시장 규모는 매년 증가추세에 있으며 2012년 기준 7,106백만 유로로서 이는 전년도 보다 약 2% 증가한 수치임. 세계적 종자회사인 Limagrain이 있는 프랑스가 약 31%를 차지하고 있고, 그 뒤로 독일(12.8%), 이탈리아(8.4%), 스페인(7.2%), 네덜란드(6.5%) 순으로 나타남.

○ 특히 스페인과 네덜란드의 경우 최근 5년간 평균 점유율이 각각 16.2%, 26.9% 성장하는 등 높은 성장률을 보이고 있음.

표 11. EU 회원국의 국내 종자시장 규모

단위 : 백만 유로, %

| 구분 | 05 | 06 | 07 | 08 | 09 | 10 | 11 | 12 | 12년도 점유율 |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|
| 프랑스 | 1,101 | 1,537 | 1,532 | 2,040 | 2,294 | 2,338 | 2,586 | 2,179 | 30.7 |
| 독일 | 804 | 796 | 673 | 628 | 811 | 951 | 841 | 911 | 12.8 |
| 이탈리아 | 522 | 621 | 730 | 510 | 513 | 588 | 514 | 597 | 8.4 |
| 스페인 | 241 | 239 | 328 | 306 | 323 | 339 | 323 | 514 | 7.2 |
| 네덜란드 | 241 | 166 | 219 | 204 | 384 | 441 | 420 | 459 | 6.5 |
| 영국 | 458 | 205 | 292 | 272 | 287 | 302 | 323 | 350 | 4.9 |
| 체코 | 121 | 159 | 219 | 204 | 215 | 226 | 219 | 237 | 3.3 |
| 헝가리 | 161 | 159 | 219 | 204 | 215 | 226 | 216 | 233 | 3.3 |
| 폴란드 | 322 | 207 | 255 | 238 | 186 | 196 | 187 | 218 | 3.1 |
| 스웨덴 | 161 | 123 | 175 | 163 | 172 | 181 | 180 | 195 | 2.7 |
| 루마니아 | - | - | 161 | 150 | 158 | 166 | 158 | 171 | 2.4 |
| 덴마크 | 161 | 135 | 182 | 136 | 118 | 140 | 165 | 170 | 2.4 |
| 그리스 | 113 | 112 | 175 | 163 | 172 | 181 | 172 | 156 | 2.2 |
| 벨기에 | 104 | 104 | 139 | 129 | 133 | 140 | 133 | 144 | 2.0 |
| 핀란드 | 64 | 82 | 117 | 109 | 115 | 121 | 115 | 125 | 1.8 |
| 오스트리아 | 137 | 135 | 109 | 102 | 108 | 113 | 108 | 113 | 1.6 |
| 불가리아 | - | - | 88 | 82 | 86 | 91 | 86 | 93 | 1.3 |
| 슬로바키아 | 72 | 72 | 80 | 75 | 79 | 83 | 79 | 86 | 1.2 |
| 아일랜드 | 48 | 48 | 58 | 54 | 57 | 60 | 57 | 62 | 0.9 |
| 포르투갈 | 48 | 48 | 58 | 54 | 57 | 60 | 57 | 62 | 0.9 |
| 슬로베니아 | 24 | 24 | 29 | 27 | 29 | 30 | 29 | 31 | 0.4 |
| 계 | 4,903 | 4,972 | 5,839 | 5,849 | 6,511 | 6,974 | 6,968 | 7,106 | 100.0 |

자료 : The EU seed and plant material market in perspective : a focus on companies and market shares. EU, 2013. pp 24.

- 주 1. 밭작물, 채소, 화훼종자를 포함한 수치임.
- 2. 씨감자는 제외함.

(5) 세계 주요 종자기업 매출 규모

○ 2014년 기준 매출액이 가장 높은 종자기업은 미국의 Monsanto사로서 약 8,420백만 유로의 매출액을 기록함. 이는 전 세계 종자시장 매출액의 약 23.7%를 차지하는 비중으로 단일 기업으로서는 매우 높은 수치임을 나타내고 있음. 그 다음으로는 DuPont사로서 약 16.8%의 비중을 차지하고 있음.

○ 세계 종자산업은 상위 5개 기업이 이미 2010년 이미 50% 이상의 점유율을 차지하고 있고, 2014년은 무려 66%에 달할 정도로 특정 기업에 의해 주도되어가고 있고 이들의 시장지배력은 더욱 커지고 있는 상황임. 이런 현상의 주요한 요인으로는 기술력과 자본력이 막강한 일부 기업에 의해 무차별적인 인수합병이 주요 요인으로 거론되고 있음.

표 12. 세계 10대 종자회사 매출액

단위 : 백만 유로, %

| 구분 | 2010 | 2011 | 2012 | 2013 | 2014 |
|--------------------------|--------|--------|--------|--------|--------|
| Monsanto | 5,975 | 6,419 | 7,919 | 8,096 | 8,420 |
| DuPont/Popneer Hi-Breed | 4,204 | 4,543 | 5,914 | 6,435 | 5,989 |
| Syngenta | 2,202 | 2,382 | 2,619 | 2,509 | 2,474 |
| Group Limagrain/Vilmorin | 974 | 1,109 | 1,256 | 1,392 | 1,419 |
| DOW | 442 | 804 | 1,084 | 1,285 | 1,257 |
| KWS | 754 | 855 | 986 | 1,147 | 1,178 |
| Land O'Lakes/Winfield | 848 | 811 | 1,034 | 1,007 | 1,041 |
| Bayer Crop Science | 687 | 820 | 897 | 921 | 978 |
| DLF Trifolium | 255 | 274 | 292 | 330 | 427 |
| Sakata | 291 | 319 | 341 | 296 | 301 |
| Top 10 seed market | 16,632 | 18,338 | 22,343 | 23,419 | 23,483 |
| Total world seed market | 32,342 | 31,790 | 33,014 | 35,235 | 35,570 |
| CR3 | 38 | 42 | 50 | 48 | 47 |
| CR5 | 44 | 48 | 57 | 55 | 54 |
| CR10 | 51 | 58 | 68 | 66 | 66 |

자료 : Policy Department Structural and Cogesion Policies-Overview of the Agricultural Inputs Sector in the EU. 2015. EU. pp 26.

○ 채소종자 판매액 1위 기업은 미국의 Monsanto로서 2015-2016 회계연도 기준 판매액은 약 728백만 유로로 나타남. 이어 프랑스의 Vilmorin% Cie, 스위스의 Syngenta, 독일의 Nunhems, 네덜란드의 Rijk Zwaan 순임.

표 13. 세계 주요 종자회사의 채소종자 판매액

단위 : 백만 유로

| 구분 | 국가 | 판매액 |
|---------------------------------|------|-----|
| Monsanto | 미국 | 728 |
| Vilmorin & Cie(Group Limagrain) | 프랑스 | 689 |
| Syngenta | 스위스 | 560 |
| Nunhems(Bayer Crop Science) | 독일 | 399 |
| Rijk Zwaan | 네덜란드 | 340 |

자료 : Vilmorin & Cie Annual Report 2015-2016. Vilmorin & Cie. 2017. pp 21.

주 : 2015-2016 회계연도 기준임.

2. 종자 수출 증가를 통한 종자 판로확대 등 종자산업 발전을 위한 수출관련 현황조사

가. 종자 유통 및 수출입 현황

(1) 유통현황

○ 1992년 이후 모든 품목의 종자 수입이 자유롭게 되면서 수입 뿐만 아닌 국내에서 육성된 품종의 종자를 해외에서 채종할 수 있는 여건이 마련됨. 전체적으로 봤을 때 해외채종을 경험한 국내 종자업체는 많은 편은 아니지만(73개 업체, 6.0%), 규모가 큰 업체일수록 해외채종의 비중이 높은 것을 알 수 있음. 특히 중규모, 대규모 업체의 경우 해외채종 경험이 각각 39.0%, 52.9%로 규모가 커질수록 해외채종 경험이 많은 것으로 나타남.

○ 해외 채종을 하는 가장 큰 이유로는 채종비용의 절감 측면이 45.4%로 가장 높은 것으로 나타남. 그 밖에 기후조건의 유리함 때문에 해외 채종을 하며, 채종인력의 부족으로 인한 해외채종 등의 이유로 해외채종을 하는 것으로 나타남.

표 14. 종자업체의 해외채종 현황(2015년 기준)

단위 : 개소, %

| 구분 | 해외채종 경험 | | | 해외채종 사유 | | | | | | |
|------------|-------------|-----------------|------------------|---------------------|----------------|----------------|----------------|---------------|-----|-----|
| | 유 | 무 | 합계 | 기후 등 국내채종 불가능 | 채종 비용 절감 | 품종 유출 방지 | 채종 인력 부족 | 채종농가 확보 곤란 | 기타 | |
| 전체 | 73 (6.0) | 1,134 (94.0) | 1,207 (100.0) | 29.8 | 45.4 | 2.0 | 12.7 | 5.9 | 4.3 | |
| 판매액 규모별 | 소규모 | 30 (2.8) | 1,031 (97.2) | 1,061 (100.0) | 33.8 | 38.8 | 2.5 | 11.3 | 6.3 | 7.4 |
| | 중소규모 | 18 (20.5) | 70 (79.5) | 88 (100.0) | 34.0 | 39.6 | 3.8 | 15.1 | 5.7 | 1.9 |
| | 중규모 | 16 (39.0) | 25 (61.0) | 41 (100.0) | 23.4 | 51.1 | 0.0 | 19.1 | 4.3 | 2.1 |
| | 대규모 | 9 (52.9) | 8 (47.1) | 17 (100.0) | 20.0 | 68.0 | 0.0 | 0.0 | 8.0 | 4.0 |

자료 : 국립종자원 보도자료(2017. 4. 12)

주 : 1. 2015년 기준 종자업 등록업체는 1,699개소이나 폐업 등을 제외하고 1,405업체 대상으로 조사하였으며, 적격률 86.1% 무응답 3개소를 제외한 1,207개소 업체를 대상으로 한 조사결과임.

2. ()안은 해외채종 경험 여부 비중임.

○ 종자업체의 종자 판매처로는 93.8%의 업체가 국내에만 판매하는 것으로 나타났고, 국내 판매 및 해외수출을 병행하는 업체는 5.4%로 나타남. 해외 수출만 하는 업체는 8개소로 매우 적은 것으로 나타남.

○ 판매액 규모별 판매처를 살펴보면 규모가 클수록 해외판매 비중이 높은 것으로 나타남.

표 15. 종자업체의 종자 유통지역(2015년 기준)

단위 : 개소, %

| 구분 | | 국내(판매)만 | 해외(수출)만 | 국내(판매)와 해외(수출) 병행 | 합계 |
|------------|------|-----------|---------|-------------------|--------------|
| 전체 | | 974(93.8) | 8(0.8) | 56 (5.4) | 1,038(100.0) |
| 판매액 규모별 | 소규모 | 861(96.5) | 6(0.7) | 25 (2.8) | 892(100.0) |
| | 중소규모 | 75(85.2) | 2(2.3) | 11(12.5) | 88(100.0) |
| | 중규모 | 28(68.3) | 0(0.0) | 13(31.7) | 41(100.0) |
| | 대규모 | 10(58.8) | 0(0.0) | 7(41.2) | 17(100.0) |

자료 : 국립종자원 보도자료(2017. 4. 12)

주 : 1. 2015년 기준 종자업 등록업체는 1,699개소이나 폐업 등을 제외하고 1,405업체 대상으로 조사하였으며, 적격률 86.1% 무응답 3개소를 제외한 1,207개소 업체를 대상으로 한 조사결과임.

2. ()안은 합계에서 차지하는 비중임.

표 16. 종자업체의 해외채종 현황(2015년 기준)

단위 : 억 원

| 구분 | | 종자 판매액 | | | 계 |
|------------|------|-------------|--------------|-------------------------|--------------|
| | | 국내 판매 | 해외 판매(수출) | 해외 생산 후 해외 판매(수출) | |
| 전체 | | 4,414(88.1) | 564(11.3) | 29(0.6) | 5,008(100.0) |
| 판매액 규모별 | 소규모 | 584(97.0) | 11(1.8) | 6(1.0) | 602(100.0) |
| | 중소규모 | 674(94.1) | 39(5.4) | 3(0.4) | 716(100.0) |
| | 중규모 | 829(93.1) | 47(5.3) | 14(1.6) | 890(100.0) |
| | 대규모 | 2,327(83.1) | 468(16.7) | 6(0.2) | 2,800(100.0) |

주1. 판매액 규모별로 소규모는 5억 원 미만, 중소규모는 5억 원~15억 원 미만, 중규모는 15억 원~40억 원 미만, 대규모는 40억 원 이상 기준임.

주2. ()안은 비중임.

자료 : 2016 종자업 실태조사(2015년 기준). 농림축산식품부. 2017.

(2) 수출입 현황

○ 2016년 기준 채소종자 수출량은 약 579톤으로 이는 전년도의 616톤보다 약 6.1% 감소한 수치임. 채소종자의 수출은 매년 증감을 반복하고 있고 큰 변화를 보이지 않고 있지만 수출액은 매년 증가추세에 있음. 2016년 수출액은 50,521천 달러로서 이는 전년도 보다 약 11.0% 증가한 수치임. 수출량이 증감을 반복하고 있음에도 수출액이 증가하는 이유는 우리나라의 채소종자가 해외에서 품질로서 인정을 받고 있으며 해외 경쟁력 또한 이전과 비교하여 상승하고 있음을 의미함.

○ 2016년 기준 채소종자 수출량 중 가장 많은 비중을 차지하고 있는 것은 무로서, 약 44.5%의 비중을 차지하고 있다. 수출액 또한 24.8%로 상당한 비중을 차지하고 있음. 고추의

경우 수출량은 3.4%가량밖에 점유하고 있지 않지만 수출액 면에서 23.4%로 높은 비중을 차지하고 있어, 수출단가가 매우 높은 것을 알 수 있음.

표 17. 채소종자 수출현황

단위 : Kg, 천 \$

| | 2012 | | 2013 | | 2014 | | 2015 | | 2016 | |
|-----|--------------------|-------------------|--------------------|-------------------|--------------------|-------------------|--------------------|-------------------|--------------------|-------------------|
| | 수량 | 금액 | 수량 | 금액 | 수량 | 금액 | 수량 | 금액 | 수량 | 금액 |
| 채소류 | 544,606 (100.0) | 40,080 (100.0) | 638,275 (100.0) | 40,507 (100.0) | 567,931 (100.0) | 39,816 (100.0) | 616,428 (100.0) | 45,510 (100.0) | 578,904 (100.0) | 50,521 (100.0) |
| 고추 | 19,999 (3.7) | 9,934 (24.8) | 15,050 (2.4) | 9,667 (23.9) | 12,404 (2.2) | 8,513 (21.4) | 12,304 (2.0) | 9,419 (20.7) | 19,740 (3.4) | 11,831 (23.4) |
| 배추 | 60,104 (11.0) | 3,535 (8.8) | 65,380 (10.2) | 4,102 (10.1) | 43,335 (7.6) | 2,793 (7.0) | 60,866 (9.9) | 3,605 (7.9) | 98,065 (16.9) | 4,475 (8.9) |
| 무 | 168,557 (31.0) | 6,924 (17.3) | 291,974 (45.7) | 8,871 (21.9) | 290,841 (51.2) | 8,898 (22.3) | 357,625 (58.0) | 13,644 (30.0) | 257,562 (44.5) | 12,521 (24.8) |
| 수박 | 6,527 (1.2) | 679 (1.7) | 4,273 (0.7) | 384 (0.9) | 744 (0.1) | 230 (0.6) | 645 (0.1) | 91 (0.2) | 2,398 (0.4) | 634 (1.3) |
| 기타 | 289,419 (53.1) | 19,008 (47.4) | 261,598 (41.0) | 17,483 (43.2) | 220,607 (38.8) | 19,382 (48.7) | 184,988 (30.0) | 18,751 (41.2) | 201,139 (34.7) | 21,060 (41.7) |

주1. (사)한국종자협회 회원사 제출 자료를 기초로 함.

2. ()안은 채소류에서 해당 품목이 차지하는 비중임.

자료 : (사)한국종자협회 홈페이지(www.kosaseed.or.kr 접속일 : 17.03.13)

○ 2016년 우리나라의 채소종자 수입은 약 1,632 톤으로 이는 전년도에 2,061톤보다 약 20.8% 감소한 수치임. 최근 GSP사업 등의 영향으로 우수한 국내산 신품종 개발이 늘어남에 따라 채소종자의 수입이 감소추세에 있음. 특히 배추와 고추의 경우 수입량이 전년도보다 각각 30.9%, 24.4% 감소하는 등 매년 그 감소폭이 커지고 있음.

○ 수입량에서 가장 많은 비중을 차지하는 품목은 무로서 2016년 기준 전체 수입량의 약 59.3%를 차지하여 가장 많은 비중을 기록하고 있음. 하지만 수입단가가 평균보다 낮기 때문에 수입액은 전체의 약 15.1%를 차지하고 있음.

표 18. 채소종자 수입현황

단위 : Kg, 천 \$

| | 2012 | | 2013 | | 2014 | | 2015 | | 2016 | |
|------|----------------------|-------------------|----------------------|-------------------|----------------------|-------------------|----------------------|-------------------|----------------------|-------------------|
| | 수량 | 금액 | 수량 | 금액 | 수량 | 금액 | 수량 | 금액 | 수량 | 금액 |
| 채소류 | 2,010,527 (100.0) | 61,137 (100.0) | 2,308,139 (100.0) | 72,073 (100.0) | 2,061,527 (100.0) | 71,537 (100.0) | 2,061,339 (100.0) | 68,770 (100.0) | 1,632,227 (100.0) | 61,817 (100.0) |
| 고추 | 45,566 (2.3) | 4,497 (7.4) | 41,172 (1.8) | 6,502 (9.0) | 37,936 (1.8) | 6,257 (8.7) | 46,504 (2.3) | 7,273 (10.6) | 35,172 (2.2) | 5,869 (9.5) |
| 파프리카 | 184 (0.0) | 3,596 (5.9) | 393 (0.0) | 5,641 (7.8) | 293 (0.0) | 5,438 (7.6) | 240 (0.0) | 5,202 (7.6) | 221 (0.0) | 4,409 (7.1) |
| 배추 | 151,834 (7.6) | 1,565 (2.6) | 108,103 (4.7) | 962 (1.3) | 123,959 (6.0) | 1,125 (1.6) | 101,912 (4.9) | 1,359 (2.0) | 70,390 (4.3) | 995 (1.6) |
| 무 | 928,280 (46.2) | 9,022 (14.8) | 1,223,273 (53.0) | 13,033 (18.1) | 1,149,349 (55.8) | 13,793 (19.3) | 1,144,005 (55.5) | 9,604 (14.0) | 968,448 (59.3) | 9,316 (15.1) |
| 수박 | 12,434 (0.6) | 1,146 (1.9) | 14,060 (0.6) | 1,435 (2.0) | 10,053 (0.5) | 1,320 (1.8) | 13,039 (0.6) | 1,545 (2.2) | 9,351 (0.6) | 1,520 (2.5) |
| 기타 | 872,229 (43.4) | 41,311 (67.6) | 921,138 (39.9) | 44,500 (61.7) | 739,937 (35.9) | 43,604 (61.0) | 755,659 (36.7) | 43,787 (63.7) | 548,645 (33.6) | 39,708 (64.2) |

주1. 판매용 종자의 수입요건확인 신청내역을 기초로 함.

2. 해외채종과 순수입을 합한 총수입 기준임.

3. 수출전용 품종의 종자 수입내역은 해외채종에 포함하여 집계함.

4. (사)한국종자협회 회원사 제출 자료를 기초로 함.

자료 : (사) 한국종자협회 홈페이지(www.kosaseed.or.kr 접속일 : 17.03.13)

3. 종자관련 협회 통합을 위한 종자 산업 경쟁력 강화 및 통합 협회 운영방안 도출을 위한 기초 조사

가. 종자산업의 문제점

(1) 생산 부문

○ 1950년대 태동한 우리나라의 채소종자산업은 정부의 정책과 민간육종가의 노력으로 발전을 거듭하고 있으나, 최근 경작지 감소 등의 이유로 생산량이 줄어들고 매출액이 정체되고 있음.

○ 국내 종자시장 규모가 줄어들고 있는 문제점과 함께 국내 종자시장 자체는 이미 포화 상태.

- 2015년 기준 1,699 개소가 종자업 등록을 하였으며, 채소종자의 경우 277 개소임. 종자업은 10년간 평균 10.0%의 증가추세를 보이고 있으며 채소종자는 6.3%의 증가추세를 보이고 있음. 시장규모는 줄어들고 있는 반면 종자업체는 꾸준히 증가하고 있음.

- 또한 세계적인 종자기업이 국내 진출을 하면서 국내기업 뿐만이 아닌 세계기업과의 경쟁 역시 심화되고 있음. 금융위기 이후 세계적 종자기업이 국내에 본격적으로 들어옴에 따라 구조조정 등으로 퇴사한 육종가들이 소규모 영세 개인 육종가로 변화한 상황.

○ 2016년 종자업 실태조사 결과 종자업체의 대부분은 영세, 소규모 업체.

- 2016년 종자업실태조사(2015년 기준)결과 1,207개의 종자업체 중 5억 원 미만 매출규모 업체는 1,061개소로 87.9%의 비중을 차지하고 있음. 이어 5~15억 원 미만이 7.3%, 15~40억 원이 3.4%, 40억 원 이상 업체가 1.4%로 대부분의 업체가 상대적으로 소규모.

- 또한 소규모 업체수가 전체에서 차지하는 비중은 87.9%인데 규모별 판매액을 살펴보면 소규모 업체의 종자 판매액 비중은 12.0%로 낮은 상황. 이들 소규모 업체 중에서는 판매 실적이 없는 경우도 15.9% 가량 존재. 업체 수 대비 매우 낮은 판매액이 현실.

(2) 수출 부문

○ 농업부분 경쟁력 강화를 위해 정부의 강한 수출정책 시행 결과, 연간 농림축수산물 수출은 2016년 기준 약 86억 달러로 최근 10년 간 평균 10% 이상의 높은 성장세를 거듭하고 있음.

○ 농림축수산물의 수출은 꾸준히 증가하고 있지만 종자업의 수출은 매우 적은 상황.

- 종자업체의 종자 유통지역을 살펴보면 총 1,038개의 종자업체 중 93.8%가 국내 판매만 하고 있음. 특히 규모가 작을수록 국내 판매 비중이 높음.

- 판매액 기준 역시 2015년도 총 종자 판매액 5,008억 원에서 수출액은 593억 원으로 약 11.9% 수준으로 매우 낮은 비중. 종자유통지역에서 살펴봤듯이 소규모 업체로 갈수록 국내 판매비중이 압도적으로 높음. 판매액 5억 원 미만의 업체의 경우 국내 판매액이 97%.

○ 전 세계적으로 농식품 교역 규모가 확대되고 FTA체결 등으로 인하여 관세장벽이 낮아지고 있지만, 전 세계적으로 자국의 농업을 보호하기 위하여 수입과 관련된 검역, 라벨링, 통관 등 각종 제도의 강화와 더불어 비관세 장벽 등 수출입 정책은 자주 변화하고 있는 상황.

○ 이에 정부, 해외 대사관 및 수출 관련기관, 개별기업에서는 수출 대상국의 농림축수산물 수입정책, 통관정책, 검역제도 등이 우리나라 농림축수산물의 수출에 미치는 영향을 인지하고 지속적인 모니터링을 시행하며 홈페이지, 책자 발급 등을 통해 홍보하고 있음.

- 한국농수산물유통공사(KATI)에서는 주요 국가별 수출입 절차를 제공하고, 주요국 수입 제도를 모니터링 하고 있음. 또한 나라별 통관 검역, 농약·첨가물 규정, 수출검역제도, 식품안전관련 법률 등을 모니터링 하고 있음.
- 대한무역투자진흥공사(KOTRA)에서도 통합무역정보서비스를 통해 정책, 법률, 규제 등에 대한 정보를 제공하고 있고 통관거부사례 및 통관거부 지침 등의 운영을 통해 우리나라 수출기업의 수출을 지원하고 있음.
- 또한 검역 규정 등은 해당 수출 대상국의 대사관을 통해 확인할 수 있는 등 다양한 방법으로 수출 대상국의 정보를 파악할 수 있음.

○ 수출 대상국의 정책 및 제도 정보는 수출에 있어서 좋은 종자의 생산과 함께 가장 중요한 항목이라고 할 수 있음. 정부 및 유관기관에서 다양한 정보를 제공하고 있지만, 자주 변경되는 규정을 즉각 전달하기엔 한계가 있음.

- 변경된 정책 및 제도를 수출업체가 전달 받고 즉각적인 대응을 할 수 있는 시스템이 필요하지만 아직까지 구축되지 않은 상황.
- 수출 대상국의 정책 및 제도에 대한 정보는 수출이 많이 이루어지고 있는 EU, 아시아국가, 미국 등 일부 국가에만 제공되고 있어 새로운 수출국 개척을 통한 수출 다양화에 어려움이 있음.
- 또한 농업 분야 정보의 대부분은 농산물, 농식품에 대한 정보가 많고 종자에 특화된 정책 및 제도에 대한 정보는 극소수에 해당됨.

○ 정부, 해외 대사관, 수출 관련기관에서 모니터링을 시행하고 책자발급, 홈페이지 게시 등을 통해 정보가 필요한 곳에 알리고는 있지만 발간주기가 변경주기보다 적기 때문에 수요를 충족하기엔 충분하지 않은 상황. 또한 정보가 필요한 수출업체에서 활용할 수 있는 경로가 제한적임.

- 한국농수산물유통공사에서 일부 수출 대상국의 변경된 정책 제도 등을 수출업체에 직접 제공하긴 하지만 충분하다고 볼 수 없음.

○ 또한 수출업체 입장에서는 담당직원이 따로 없다면 변경된 정책을 매번 찾아보기 어렵고, 즉각적인 대응이 힘들.

- 일부 대기업에서는 자체조사를 통해 대응이 가능하지만 영업비밀 상의 이유로 외부로의 정보 공유를 꺼리고 있음. 종자업체 대부분이 영세한 상황에서 소수의 직원으로 수출시장의 정책 및 제도를 파악하기는 어려운 상황.

○ 농림수산물식품기술기획평가원(2014)의 연구에 따르면 종자업체들은 수출을 함에 있어 '해외시장 정보부족', '바이어의 접촉', '통관법규 지식의 부족', '목표시장의 법과 제도적 제한', '수출 목표 국가의 검역', '수출 목표 국가의 종자 반출입 규제' 등을 주요한 애로사항으로 선정하였음.

(3) 종자관련 협회의 난립

○ 종자에 대한 연구개발 등은 활발히 이루어지고 있지만, 종자산업 자체가 위축되고 있는 상황에서 민간 중심의 종자산업 육성을 위해 관련된 협회의 기능 및 역할은 점차 중요해지고 있는 상황.

○ 현재 식량, 종축, 수산부문을 제외한 원예 분야에서 종자와 관련된 협회는 한국종자협회 등 약 10여개의 협회가 운영 중에 있음. 각 협회는 대부분 회원사의 회비를 통해 운영되고 있으며, 종자의 생산, 유통, 연구, 교육, 정책의 발굴 등 종자와 관련된 다양한 역할을 수행하고 있음.

○ 협회에 소속된 종자기업의 이익증진과 더불어 나아가 종자산업의 발전을 위해 설립된 종자 관련 협회는 대부분 소규모로 운영되고, 예산상의 제약으로 인하여 각 협회가 수행하여야 할 본연의 기능을 수행하기에는 그 규모와 역량이 충분하지 못하다는 문제가 제기되고 있음.

표 19. 종자 관련 주요 협회

| 구분 | 주요 사업 | 홈페이지 |
|-----------|---|--------------------|
| 한국종자협회 | <ul style="list-style-type: none"> ○ 우량종묘의 육종, 생산, 유통 등 공동 이익증진 도모 ○ 종자수급조절을 위한 판매 및 수출입의 알선 ○ 종자에 관한 조사연구 및 진흥책의 조장 ○ 종자에 관한 회지, 간행물 등의 발간 및 홍보활동 | www.kosaseed.or.kr |
| 한국과수종묘협회 | <ul style="list-style-type: none"> ○ 과수종묘생산에 필요한 기술 및 정보교환 ○ 우량종묘생산을 위한 재배법의 표준화 및 종묘 규격화 지도 ○ 무허가 종묘생산 근절을 위한 지도 계몽 ○ 과수종묘의 품종생산판매신고에 관한 지도 및 협조 | www.kkja.org |
| 한국육묘산업연합회 | <ul style="list-style-type: none"> ○ 육묘산업의 발전을 위한 조사연구 ○ 육묘산업의 품질향상 및 전문기술 발전을 위한 기술교육사업 ○ 외원의 지도, 안내, 알선과 각종 분쟁의 중재조정 ○ 육묘산업의 발전을 위한 자료정보의 수집 및 연구기관과의 주기적인 정보교환 | www.hkyugmyo.or.kr |

표 20. 종자 관련 주요 협회(계속)

| 구분 | 주요 사업 | 홈페이지 |
|-------------|--|---------------|
| 한국종자연구회 | <ul style="list-style-type: none"> ○ 정기적인 연구발표회, 세미나 또는 심포지엄 개최 ○ 연구결과 및 국내외 관련 정보를 실은 연구회지 발간 ○ 종자산업 발전을 위한 연구사업 및 정책개발 ○ 민간육종가 교육훈련 | www.kosid.net |
| 한국종자포럼 | <ul style="list-style-type: none"> ○ 종자관련 정책·제도 연구사업 ○ 종자관련 기술의 컨설팅 및 교육사업 ○ 식량안보 및 종자유권 확보를 위한 국내외 종자 신자원 탐색·이용 연구 및 국제 협력 사업 ○ 종자생명산업 발전을 위한 정부, 지방자치단체 및 산업체와의 협력 사업 ○ 연구개발결과의 확산을 위한 정보지 발간, 전시회 및 연구발표회 등 개최 | www.koseed.kr |
| 한국종근생산협회 | <ul style="list-style-type: none"> ○ 종근의 질적 향상을 위한 기술지도, 교육 ○ 종근의 원활한 수급을 위한 종근판매 및 시장개척 사업 ○ 우량종근의 개발육성 보급 ○ 종근의 공정거래 질서 확립과 홍보 | |
| 한국화훼종자협회 | <ul style="list-style-type: none"> ○ 종자에 관한 조사, 연구 및 진흥책의 조장 ○ 관련 단체와의 화훼 관련 사업의 교류 ○ 화훼 수출입 업무의 지원 ○ 회원사간 정보 및 자료 교류를 위한 간행물 발간 및 홍보 | |
| 한국과수민간육종가협회 | <ul style="list-style-type: none"> ○ 육종기술의 상호교류를 통한 신품종육성 활성화 ○ 신품종 육성 및 보급을 위한 교육 ○ 홈페이지 개설 및 언론 광고 등을 통한 우수품종 홍보 ○ 기타 법인의 목적달성에 필요한 사업 | |
| 한국난과식물육종가협회 | <ul style="list-style-type: none"> ○ 난과식물의 기내육종(배양)기술 개발 및 보급 ○ 신품종육성 및 보급 ○ 워크샵을 통한 조기 기술 전수 ○ 기술서적발간 및 홈페이지 운영을 통한 신기술 홍보 | |
| 한국종자판매협회 | <ul style="list-style-type: none"> ○ 종자판매상의 육성발전과 회원 공동이익 증진을 위한 사업 ○ 종자판매상의 유통질서 확립을 위한 교육사업 ○ 종자유통 채널 구축 및 출판, 회보 발행 등 홍보사업 ○ 정부가 위임하는 사업의 수행 | |

자료 : 한국과수종묘협회 홈페이지(<http://www.kkja.org>, 접속일 : 2017. 08. 01), 한국육묘산업연합회 홈페이지(<http://www.hkyugmyo.or.kr>, 접속일 2017. 08. 01), 한국종자연구회 홈페이지(<http://www.kosid.net>, 접속일 2017. 08. 01),

한국종자포럼 홈페이지(<http://www.koseed.kr>, 접속일 2017. 08. 01), 농림축산식품부 내부자료.

○ 또한 일부 품목의 경우 2개 이상의 협회가 운영 중에 있을 만큼 종자시장의 규모에 비해 종자와 관련된 협회가 난립해 있다는 지적이 있음.

○ 이에 학계, 산업계 등에서는 종자협회의 통합을 통해 종자산업의 발전을 위한 논의가 이루어지고 있지만, 각 협회별로 수입구조, 규모 등이 매우 다르고 협회 간 통합에 대한 의견차이로 현실적으로 통합에 어려움이 있는 상황임.

○ 종자관련 협회가 난립하며 생기는 문제는 다음과 같음.

▶ (첫째) 정책 모니터링 및 정책수요 취합의 한계 : 정책 모니터링은 종자 관련 정책의 입안을 뒷받침 할 수 있는 중요한 근거가 될 수 있으며, 정책의 실행과정에서 발생하는 건의사항, 문제점 등을 즉각 반영할 수 있고, 사후 평가를 통해 개선방안을 반영할 수 있는 장점이 있음.

- 하지만 협회가 너무 많아 농림축산식품부 등 정책을 집행하는 정부기관에서 정확한 모니

터링이 어려운 상황.

- 또한 개별 종자협회별로 정책수요가 상이하어 종자산업 정책의 방향 설정에 문제가 발생함.
- ▶ (둘째) 종자 관련 통계자료의 부재 : 현재 종자와 관련된 통계는 크게 농림축산식품부와 국립종자원, 한국종자협회에서 제공되고 있음.
 - 농림축산식품부는 2017년부터 ‘종자업 실태조사’를 통해 종자업에 관한 통계를 제공함. ‘종자업 실태조사’에서는 종자업의 규모, 품목부류, 생산, 유통 등의 측면에서 유용한 통계자료가 될 수 있지만, 품목별 부류가 되어있지 않아 이용에 한계가 있음.
 - 국립종자원의 경우 ‘종자업 등록현황’, ‘종자관리사 등록현황’, ‘종자유통조사 실적 및 조치사항’, ‘품종보호출원 및 등록현황’ 등을 제공하고 있음. 이는 우리나라의 종자 업체 수를 파악할 수 있는 자료이긴 하지만, 정책 반영에 있어 기초적 자료 수준에 불과함.
 - 한국종자협회는 1년 단위의 채소종자 생산량, 수출입 현황, 매출액 등을 발표하고 있음. 품목별로 자세한 자료를 제공하기 때문에 가장 유용한 자료로 많은 정책의 근거로 사용되고 있지만, 이 역시 채소분야에 국한되어 있고, 회원사의 제출 자료로 이루어진 통계이기 때문에 대표성 측면에서 아쉬운 점이 있음.
- ▶ (셋째) 인력부족으로 인한 연구 효율성 저하 : 종자 관련 협회의 여러 가지 사업 중 가장 중요한 사업은 연구기능. 우량종자의 생산을 위해 육종, 육묘등의 연구와 마케팅 등의 분야에 대한 연구는 꼭 필요 하지만 연구인력 부재로 많은 연구가 이루어지지 못하고 있음. 일부에서 소수 연구가 이루어지고 있기도 하지만 대부분 외부 위탁 형태로 이루어지고 있음.
- ▶ (넷째) 수출실무 및 현지조사 업무
 - 대부분의 종자생산 업체는 영세하고, 포화상태인 국내 종자시장에서 종자업체가 경쟁력을 갖추고 나아가 종자산업이 발전하기 위한 방법 중 한 가지는 수출을 통한 새로운 판로를 개척하는 것임.
 - 수출 대상국의 정책 및 제도 정보는 수출에 있어서 좋은 종자의 생산과 함께 가장 중요한 항목이라고 할 수 있음. 정부 및 유관기관에서 다양한 정보를 제공하고 있지만, 자주 변경되는 규정을 즉각 전달하기엔 한계가 있음.
 - 또한 수출 대상국의 소비자 선호도 조사가 이루어지지 않고 있음. 국내에서 선호도가 높은 종자가 해외시장에서 선호도가 높다고 할 수 없기 때문에 수출 대상국의 선호도는 육종 단계부터 고려되어야 될 중요한 문제임.
 - 일부 대기업에서는 수출 대상국의 정책 및 제도, 소비자 선호도 등 수출 시장 정보를 자체조사를 통해 대응이 가능하지만 영업비밀 상의 이유로 외부로의 정보 공유를 꺼리고 있음. 종자업체 및 종자 관련 협회가 대부분 영세한 상황에서 소수의 직원으로 수출시장

의 정책 및 제도를 파악하기는 어려운 상황.

- ▶ (다섯째) 동일 업무의 중복 : 각 협회에서는 교육, 회원사의 관리, 연구, 수급조절, 생산, 수출, 홍보, 단속(지도), 판매 등 종자 산업과 관련하여 많은 업무를 담당하고 있음.
- 많은 업무를 담당하고 있음에도 개별 협회는 예산상의 제약으로 소규모로 운영되고 있어 인력부족, 역량부족 등의 이유로 협회 고유의 업무를 모두 수행하고 있지 못함.
- 또한 중복된 일을 서로 다른 협회에서 각각 진행하고 있어 운영의 효율성 측면에서 많은 문제점이 나타나고 있고, 사업의 효과가 저하되고 있음.

나. 정책제언

(1) 종자기업의 품목 전문화를 통한 경쟁력 강화

- 종자는 특성에 따라 식량, 채소, 과수 등 몇 가지로 나뉨. 식량 및 과수의 경우 국가기관의 주도로 육종이 이루어지지만 채소종자의 경우 민간 종자기업의 주도로 산업을 이끌어 나가고 있음.
- 종자산업은 전 세계적으로 국가기관 주도에서 민간 종자기업 중심으로 변환되고 있는 추세에 있음. 채소종자의 경우 민간역량과 경쟁력이 매우 중요한 상황임.
- 영세한 소규모 민간이 주도하고 있는 채소종자 산업은 정체되고 있고, 포화상태이며, 경쟁력이 많이 저하된 상황.

○ 이에 민간 위주의 채소종자사업에 대한 지원 강화와 더불어 종자기업의 품목 전문화를 통해 소규모 영세종자기업의 시장경쟁력을 강화하고, 시장 맞춤형 수출 주력품목 양성할 필요성이 있음.

(2) 종자 수출 증가를 위한 수출대상국 종자 수입 정책 및 제도 모니터링 시스템 구축

- 농업은 식량 확보의 중요한 자원이며, 특히 종자산업은 모든 농업의 기초로서 식량주권을 확보함에 있어 가장 중요한 산업이라 할 수 있음. 특히 우리나라 종자산업은 많은 부분 수입에 의존하고 있으며, 종자의 수입은 다른 품목과는 달리 로열티 지급 등 수입 이후에도 많은 비용이 소요되고 있음. 특히 ‘종자주권’이라는 말이 있을 정도로 농업에 있어 종자는 매우 중요한 산업.
- 특히 채소종자산업은 종자업 판매액 중 1위를 기록할 정도로 우리나라의 종자산업에서 가장 많은 비중을 차지하고 있는 품목.

○ 채소종자는 수출보다 수입이 많은 수입의존 품목.

- 최근 종자소비 감소 및 GSP사업을 통한 우수종자의 개발 등으로 수입량이 다소 감소하고 있는 추세이지만, 여전히 수출량 보다 수입량이 매우 많은 수입 의존형 산업임. 물론 세계적으로 품질을 인정받고 있는 고추, 배추, 무 등의 종자는 수입보다 수출이 많은 상황이지만 채소 품목의 대부분은 수입이 수출보다 많은 상황.

○ 위와 같은 이유로 인하여 제한되고 포화상태인 종자산업 특히 채소종자산업의 경쟁력 강화를 위해선 해외로 시장을 확대해 나가는 것이 필요함. 아무리 수출시장 현지에 맞는 우수한 종자를 육종하더라도 정책 및 제도가 수출에 걸림돌이 되어서는 안 될 것임. 이에 수출 대상국의 정책 및 제도에 대하여 시스템 구축을 통해 수출업체의 대응력 제고와 원활한 수출을 도모하여 경쟁력을 강화할 필요성이 있음.

○ 시시각각 변화하는 수출 대상국의 종자관련 정책 및 제도의 효과적 DB 구축 및 배포 통한 종자 수출업체의 원활한 수출 도모하고 신규 진입 가능시장의 수출입 정책 및 제도 분석 통한 수출시장 다변화가 필요한 상황임.

○ 앞서 나열한 여러 가지 문제점 등을 해결하고 종자산업 발전을 위해 종자관련 협회의 통합이 필요한 상황이지만, 여러 가지 이유로 인하여 통합에 어려움이 있으며, 가장 큰 문제점은 통합 후 협회의 운영 이라고 할 수 있음.

- 통합 후 협회의 운영은 새로운 형태의 사업개발과 더불어 정부의 지원사업, 위탁사업, 교육사업, 기타사업 등을 통한 재원마련이 필요한 상황임.
- 통합 종자협회의 운영은 1차적으로 해당 회원사의 회비로 운영되지만, 인력의 운용, 연구사업 등의 수행을 하기 위해선 정부의 지원금 없이 운영이 힘든 상황. 이에 유사한 사례를 벤치마킹할 필요성이 있음.

(3) 관련 협회 사례

○ 한국바이오협회는 2008년 한국생명공학연구조합, 한국바이오산업협회, 한국바이오벤처협회가 통합하여 출범함으로써 기존 난립하고 있던 바이오 관련 협회의 기능을 통합함. 한국바이오협회는 국제협력사업, 정책기반사업, 성장지원 및 인력양성사업, 조사연구 및 홍보사업 등을 시행하고 있음. 특히 산업통상자원부, 지식경제부 등 정부 위탁사업을 통해 자생력을 높여가고 있음.

표 21. 한국 바이오협회의 정부지원 및 위탁 사업 주요 내역

| 주요 사업 |
|---|
| <ul style="list-style-type: none"> ○ 정책기반사업(정부 R&D과제 수행) <ul style="list-style-type: none"> - 산업통상자원부 지원과제 “항체 특성분석 기반기술 및 분자재설계를 통한 항체의약품 개발”외 8건 - 지식경제부 지원과제 “바이오기반 체외진단기기 사업화 지원” ○ 정책기반사업(정부지원사업 추진) <ul style="list-style-type: none"> - 생물무기금지협약(BWC)의 국내 이행사업 위탁기관업무 수행 : 국제회의, 생물작용제 제조, 보유시설 정기검사, 생물무기 모니터링, 실무협의회 운영 등 - 지식경제부 산업융합원천기술개발 기반조성사업의 일환으로 ‘바이오의약품 비임상, 임상시험을 위한 국내 CRO 육성기반 구축사업’ 진행 중 - 지식경제부 산업원천기술개발사업 중 바이오분야사업 지원기관 선정 - 지식경제부 ‘바이오화학 2.0 사업’의 예비타당성 조성사업 - ‘바이오특성대학원 운영사업’의 주관기관 - ‘미래산업 핵심인재 10만 명 양성사업 계획’의 일환으로 시행 중인 ‘바이오 전문인력 양성사업’ 주관기관 - ‘바이오 인턴십’프로그램 주관기관 - 지식경제부의 ‘바이오산업 인적자원개발협의체’ 사무국 설치 및 운영. - 지역연고산업 육성사업 - 지식경제부 ‘미래 유망 신서비스발굴 과제기획사업’ 참여 ○ 국제협력사업 지원(‘Korea 바이오허브사업’ 주관기관 선정) ○ 성장지원 및 국제교류사업 <ul style="list-style-type: none"> - ‘코리아 바이오파크’건설사업 : 2011년 완공 후 벤처기업 입주 완료. - 국내 바이오산업 통계조사 : 트렌드 조사(상반기), 실태조사(하반기) - 대구광역시, 전라북도 바이오화학 연구용역 과제 위탁 - 바이오헬스기업 글로벌 진출을 위한 특허기반 정보시스템 구축사업 참여기관. |

자료 : 한국바이오협회(2012). 한국바이오협회 30년의 발자취.

- 한국건강기능식품협회의 경우 식품의약품 안전처의 위탁사업으로 건강기능식품의 표시 및 광고심의, 건강기능식품 법정교육 등을 실시하고 있으며, 한국바이오의약품협회는 제주수입관리자교육사업 등을 정부 위탁으로 시행하고 있음. 또한 한국제약바이오협회는 정부 위탁으로 GMP교육, 제약 마케팅 교육, 인증교육 및 시험 등을 실시하고, 광고심의 사업을 시행하고 있음.

표 22. 관련 협회별 정부지원 및 위탁 사업 주요 내역

| 협회명 | 주요 사업 |
|-------------|--|
| 한국건강기능식품협회 | <ul style="list-style-type: none"> ○ 정부위탁사업(식품의약품 안전처) <ul style="list-style-type: none"> - 건강기능식품의 표시·광고의 심의 - 건강기능식품 법정교육 : KHSA 건강기능식품 교육센터 운영 |
| 한국바이오회의약품협회 | <ul style="list-style-type: none"> ○ 정부위탁교육사업 <ul style="list-style-type: none"> - 제조수입관리자교육 : 의약품 등의 안전성 확보 및 국민 건강상의 위해방지를 위해 한국바이오회의약품협회가 교육 진행 |
| 한국제약바이오협회 | <ul style="list-style-type: none"> ○ 정부위탁교육사업 <ul style="list-style-type: none"> - GMP교육, 제약 마케팅 교육, MR인증교육 및 시험 ○ 정부위탁 광고심의 사업(식품의약품 안전처) <ul style="list-style-type: none"> - 의약품 광고가 일반인에게 노출 전 광고의 진위여부 확인 통한 소비자 보호 |

자료 : 한국제약바이오협회 홈페이지(www.kpbma.or.kr, 접속일 2017. 09. 12.), 한국건강기능식품협회 홈페이지(www.hfood.or.kr, 접속일 2017. 09. 12.), 한국바이오회의약품협회 홈페이지(www.kobia.kr, 접속일 2017. 09. 12.).

(4) 통합 종자협회 운영방안

○ 통합 종자협회는 품목별로 기존 협회가 하고 있는 고유한 업무는 각 협회별로 진행하되, 통합 사무국을 두어 앞서 제시한 사례를 바탕으로 다음과 같이 정부 위탁사업, 교육사업 등의 시행을 통한 운영이 필요함.

(5) 정부 위탁사업

○ 종자산업과 분야는 상이하지만, 앞서 소개한 바이오협회 등은 정부의 위탁사업을 통해 협회를 운영하고 있음. 통합 종자협회 역시 정부의 다양한 위탁사업을 통해 운영을 할 필요성이 있음. <표 10>은 종자와 관련된 기관인 농림축산식품부 및 국립종자원의 2017년도 종자관련 주요사업임.

- 농림축산식품부는 종자산업기반구축 사업을 바탕으로 종자와 관련하여 총 4가지 사업을 시행 중에 있고, 국립종자원 역시 수출산업화 및 육모산업 기반 조성 등 종자와 관련된 다양한 사업이 진행되고 있음.
- 현재 정부에서 시행하고 있는 종자관련 다양한 사업 중, 통합 종자협회의 위탁사업을 진행할 필요성이 있음.

표 23. 농림축산식품부 및 국립종자원 종자관련 주요 사업 내역

| 사업명(주관기관) | 주요 내용 |
|------------------------------------|---|
| 종자산업기반구축 (농림축산식품부) | - 식량·원예·특용작물 등의 우수한 종묘(종자)를 농업인에게 효율적으로 증식·보급할 수 있는 기반조성(시설·장비) 지원 |
| 농식품산업해외진출지원 (농림축산식품부) | - 우리 농산업을의 외연확대 및 비상시 해외곡물 공급기반 저변 확대를 위해 민간기업의 해외진출 지원 |
| 농업자금이차보전 (농림축산식품부) | - 대출취급기관에서 사업타당성을 갖춘 농업경영체 등을 선정·지원하고 지도금융 및 사후관리를 통해 농업정책자금 지원의 효율성 도모 |
| 농림축산식품연구개발 (농림축산식품부) | - Golden seed 프로젝트 : 국가전략형 수출·수입대체 종자 개발을 통한 글로벌 종자시장의 선점으로 글로벌 종자강국 실현 및 민간분야의 종자산업 성장기반 구축 |
| 수출산업화 및 육묘산업 기반 조성 (국립종자원) | - 해외 현지 품종전시포 확대 및 맞춤형 전시포 설치 및 운영 - 종자수출협의회운영 및 공동마케팅 지원 - 육묘업 관련제도 정비 및 등록제 조기 정착 |
| 종자시장 유통질서 확립 (국립종자원) | - 작물별 불법·불량종자의 수시·정기 조사 - 1품종이(興)명칭·품종진위 검사 통한 건전한 종자 유통기반 구축 - 종자업 등록, 품질표시 등 적법 유통질서 홍보 |
| 종자수출 활성화 지원 (국립종자원) | - 국제기구 및 국가 간 업무협력 확대 - ODA사업 추진과정에 수출기업의 참여 활성화 - 품종보호제도 운영 20주년 기념 국제심포지엄 개최 |
| 품종보호 기반 확충 및 권리거래 기반 확대 (국립종자원) | - 아열대 신품종 출원 증가에 따른 심사기반 마련(시험포 등) - 종자거래 활성화를 위한 종자거래소 구축(품종가치평가 기반) |

표 24. 농림축산식품부 및 국립종자원 종자관련 주요 사업 내역(계속)

| 사업명(주관기관) | 주요 내용 |
|-------------------------------------|---|
| 육종가 권리보호 대응체제 구축 (국립종자원) | - 품종보호권 침해분쟁 대응체제 확대 운영 |
| 식량종자산업 민간참여 활성화 등 보급종 업무 혁신 (국립종자원) | - 민간육성 식량종자의 보급종 도입 - 계획적 보급종 공급 통한 고품질 농산물의 생산 - 채종농가 경영안정 |
| 종자보증 신뢰 제고를 위한 제도개선 (국립종자원) | - 국가보증의 객관성 제고 및 신뢰도 향상 - 종자검정의 신뢰향상 - 종자검사원 및 종자관리사 역량 강화 |
| 민간종자 대상 종자검정 서비스 실시 (국립종자원) | - 민간의 종자검정 신청에 대한 신속한 검정 서비스 - 민간 재위탁을 통한 일자리 창출 |

자료 : 농림축산식품부(2017). 2017년도 농림축산식품사업안내서, 국립종자원(2017). 2017년도 업무계획.

○ 종자관련 연구과제 위탁 : 신품종 육종, 종자관련 통계구축, 종자산업의 전반적 정책제언

○ 종자 검정 및 분석사업의 위탁 : 분자표지 분석, 기능성 성분 분석 등

- 해외시장조사사업 통한 국산종자 수출판로 확보
- 국산종자 홍보사업 및 불량종자 유통 단속
- 국제 네트워크 형성-국제종자박람회, 해외 마케팅 사업

(6) 교육사업

- 정부 위탁사업과 더불어 교육사업을 통한 통합 종자협회 운영이 요구됨. 현재 종자와 관련된 정부의 정책은 R&D를 기반으로 종자 인프라 확대에 집중되어있고 종자와 관련된 전문인력 양성은 상대적으로 미흡한 상황.
 - 일부 자자체와 국립종자원에서 교육을 시행하고는 있지만, 종자 전문 교육훈련기관이 없는 상황임.
- 과수와 식량작물을 제외하고는 대부분 민간에 의해 주도되고 있는 종자산업이기 때문에 개인 육종가 양성 교육이 필요하며, 종자산업의 전반적 경쟁력 강화를 위하여 종자관련 전문 교육훈련이 필요한 상황.
 - 종자의 품종개발을 위한 사전 조사에서 종자의 생산, 판매까지 종자 생산의 전 과정에 대한 체계적 교육이 필요함.
 - 특히 종자관련 미래인력 양성을 위해 농업관련 전공 고등학교, 대학교 위주의 교육훈련 프로그램과 실무자교육 등 다양한 교육프로그램을 통해 협회 차원에서 종자산업 경쟁력 강화방안을 마련할 필요성이 있음.
 - 이와 더불어 종자기능사 등의 자격 취득교육 등 국가인증 자격증 교육 역시 함께 이루어져 종자 생산농가의 전문성을 높일 필요가 있음.

4. 국제협력사업과의 연계를 통한 채소 종자 수출 기반 확대

가. 종자산업과 Golden Seed 프로젝트 수출 전략

세계 종자시장 규모는 2015년 기준 372억 달러로 과거 10년간 1.5배 성장하였으며, 종자교역 규모도 동기간에 2배 이상 증가하였으나, 국내 종자시장은 5,793억 원(육묘업 포함 시 9,674억 원)으로 세계시장의 1% 수준에 머물러 있는 상황이다.

- 세계시장규모 : ('05) 247억 달러 → ('15) 372억 달러
- 종자교역규모 : ('05) 48억 달러 → ('15) 107억 달러

2015년 기준 우리나라 종자업 등록업체수는 1,207개이나 연 매출액이 5억 미만인 업체가 전체의 87.9%인 1,061개로 산업구조가 영세한 상황이며, 해외수출 실적이 있는 업체 비중은 7% 미만으로 나타나 글로벌 종자시장 선점이라는 측면에서도 어려움이 존재한다.

- 매출액 40억 원 이상(대규모)은 17개, 15~40억 원(중규모)은 41개, 5~15억 원(중소규모) 88개, 5억 원 미만(소규모)은 1,061개 업체임.
- 2016년 기준 판매실적이 있는 종자업체의 93.7%는 국내 시장에서만 판매가 이루어 졌으며, 국내 판매 및 해외 수출을 병행하는 업체는 5.9%, 해외 수출만 하는 업체는 5개소임.

표25. 판매실적 존재 종자업체의 종자 유통지역(2016년 기준)

단위 : 업체수, %

| 구분 | 국내(판매)만 | 해외(수출)만 | 국내(판매)와 해외(수출) 병행 | 계 | |
|------------|-----------------|---------------|----------------------|------------------|----------------|
| 전체 | 1,084 (93.7) | 5 (0.4) | 68 (5.9) | 1,157 (100.0) | |
| 판매액 규모별 | 소규모 | 963 (96.8) | 3 (0.3) | 29 (2.9) | 995 (100.0) |
| | 중소규모 | 78 (80.4) | 2 (2.1) | 17 (17.5) | 97 (100.0) |
| | 중규모 | 35 (76.1) | 0 (0.0) | 11 (23.9) | 46 (100.0) |
| | 대규모 | 8 (42.1) | 0 (0.0) | 11 (57.9) | 19 (100.0) |

자료 : 2017 종자업 실태조사(2016년 기준).

주 1) 2016년 판매 실적이 있는 종자업체만을 대상으로 산정한 수치임.

2) () 내는 계에 대한 구성비임.

Golden Seed 프로젝트(이하 GSP)는 글로벌 종자 강국으로 도약을 위한 수출전략 종자 개발 및 산업기반 구축을 위한 4개 부·청 협력 프로젝트로 미래 농업환경 변화에 따라 새롭게 전개되고 있는 글로벌 종자시장을 선점함으로써 글로벌 종자강국으로의 실현을 목적으로 하고 있다. 현재 수립한 목표로는 1,000만 달러 수준의 국가 전략형 수출종자 19개 이상 개발로 2020년 종자수출 2억 달러 달성 및 2030년 종자수출 50억 달러 달성과 같이 종자 수출액 증가에 초점이 맞추어져 있다. GSP의 추진전략도 수출시장 성장모형에 따라 글로벌 시장개척, 품종보호 전략종자 개발, 그리고 관련기술 개발 추진 등으로 설정되어 있다.

- 글로벌 시장개척 종자 : 현재 보유하고 있는 강점기술을 기반으로 수출시장 개척용 종자 개발
- 품종보호 전략종자 : 품종보호 기반 확보 및 장기적 수출시장 진입용 종자 개발

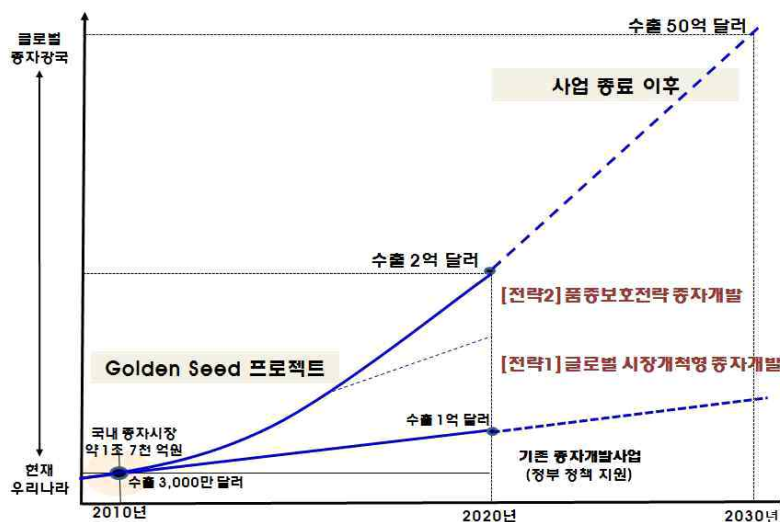


그림 1. 국내 종자산업 수출시장 성장모형

자료 : 예비타당성조사 대응 신규사업기획 연구-Golden Seed 프로젝트-

GSP 사업의 목표 달성을 위해 그간 정부에서는 국가적 역량결집, 민간 수출역량 강화, 목표시장 개척을 위한 수출형 R&DB 통합지원을 시행해 왔으나, 2017년 수출액은 58백만 달러로 목표치인 120백만 달러에 절반 수준에도 미치지 못한 상황이다. 목표치 미달성의 원인으로는 글로벌 종자로 개발된 품종의 현지화 지원 등 해외 수출 정책 미흡과 현장 맞춤형 인력양성 프로그램개발 및 총괄 컨트롤타워 부재 및 기 구축된 인프라의 효율적 활용 미흡 등이 원인으로 거론되고 있다. 정부는 제 2차 종자산업 육성 5개년 계획에서 주요 개선 방향으로 GSP 2단계 사업에서 연구개발 방향을 로열티 대응사업(로열티 수출 품종개발까지 확대), 차세대 신육종 기술 활용 연구개발 사업, 미래 환경 대응 우수품종 개발로 수정하였으며, 인프라의 경우 해외진출 및 수출 확대 대응기반 조성을 주요 방향으로 설정한 상황이다.

표26 제2차 종합계획 수출확대 및 글로벌 경쟁력 제고 정책 방향

| 전략 | 중점 추진 과제 |
|--|---|
| GSP 2단계 성공적 마무리로 종자수출 2억불 달성, 수출확대 지원 강화 | <ul style="list-style-type: none"> • GSP 성과달성을 위해 업체 중심지원 확대 및 성과관리 점검 강화 (1단계 개발품종의 수출시장개척·마케팅 강화를 통해 국내외 시장 점유율 확대) • 종자수출 촉진을 위해 국제종자박람회 확대 개최(매년) • 해외바이어 초청·국제박람회 참가지원, 해외영업인력 육성 교육비 일부지원 • 해외종자시장 정보 수집·제공, 수출국 종자분석 증명서 발급 서비스 제공 |
| 종자산업 인프라 연계 시설 확충 및 협력체계 구축 | <ul style="list-style-type: none"> • 종자산업 거점지역과 연계한 기업 공동활용형 종자종합지원시설 구축 • 종자삼각벨트와 종자연구소 등 밀집지역과 연계한 협력네트워크 구축으로 종자기업 첨단 기술지원 서비스 제공 |
| 식량종자 해외진출 기반 조성 | <ul style="list-style-type: none"> • 식량종자(벼, 감자, 옥수수) 해외틈새 시장 진출 전략 추진 (GSP 참여기업: 팜한농, 오리온, 농우바이오, 아시아종묘) • 정부의 종자정선 대행, 보관시설 민간 임대 시범 추진 (종자업체의 기반구축시 까지(3년이내) 종자정선 대행, 정선담당자에 대해 기술 연수) • 민간부분 식량작물 육종·생산기술 지원 및 종자산업기반구축사업을 통한 식량종자생산기업에 대한 시설, 기자재 구축 지원 |

자료 : 종자산업 육성 및 지원을 위한 종합계획 수립 방안 연구.

나. 국제협력사업 관계기관과의 협력을 통한 채소 종자 수출 확대 가능성

우리나라는 2009년 11월에 OECD 개발원조위원회(Development Assistance Committee: DAC) 가입 이후 공적개발원조(Official Development Assistance: ODA) 규모는 지속적으로 증대되어 왔으나 2016년 총 국민소득대비 ODA 비율(ODA/GNI)은 0.16%로 DAC 회원국 평균인 0.31%와 UN의 권고치인 0.7%에는 크게 못 미치는 수준으로 지속적인 증가가 필요한 상황이다. (ODA는 한 국가의 중앙 혹은 지방정부, 공공기관이 개발도상국의 경제발전과 복지증진을 위해 개발도상국이나 국제기구에 제공하는 자금의 흐름을 말함.) 우리나라 정부는 ODA 규모를 우리의 경제수준에 부합하는 수준으로 지속적으로 확대하기 위해 2015년까지 ODA를 GNI 대비 0.25% 규모까지 확대한다는 목표를 수립한 바 있으나 실적은 이에 미치지 못하고 있다. ODA를 포함한 우리나라의 대외무상원조실적은 2002년 777억 원에서 2016년 6,405억 원으로 8배 증가하였으며, 이중 농림수산분야 지원실적은 동기 간에 80억 원에서 961억 원으로 12배 가까이 증가한 것으로 나타났다.

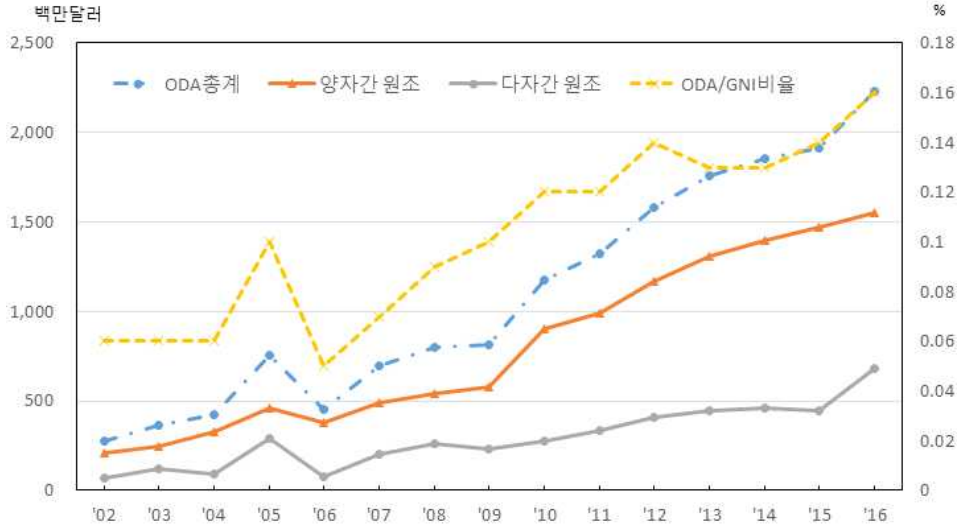


그림 2. 한국 ODA 실적 추이

자료 : 2016 KOICA 대외무상원조실적 통계.

주 : ODA총계, 양자간 원조, 다자간 원조는 왼쪽 축, ODA/GNI비율은 오른쪽 축 기준임.

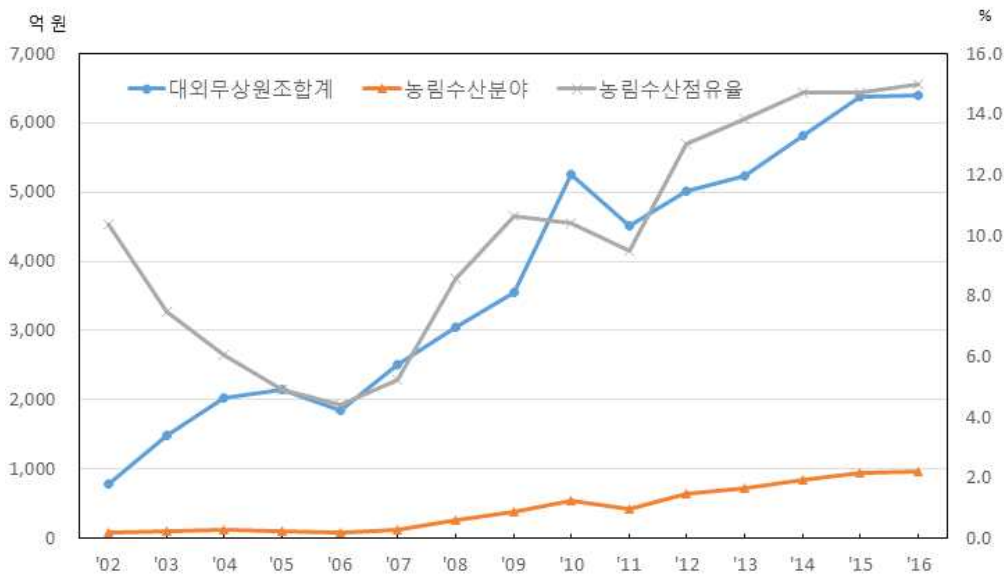


그림 3. 한국 대외무상원조실적합계와 농림수산분야 원조실적

자료 : 2016 KOICA 대외무상원조실적 통계.

주 : 대외무상원조합계와 농림수산분야는 왼쪽 축, 농림수산점유율은 오른쪽 축 기준임.

2016년 대외무상원조실적을 사업유형별로 구분하면, 프로젝트가 2,567억 원으로 비중이 가장 높고, 다음으로는 봉사단과전 1,181억 원, 개발컨설팅 707억 원, 연수생 초청 477억 원, 국제기구협력사업 380억 원 등의 순이며, 최근 비중이 증가추세인 민관협력은 120억 원 규모이다. 농림수산분야의 사업유형별 지원실적은 프로젝트 485억 원(49건), 개발컨설팅 132억 원(7건), 연수생초청 72억 원(30개 과정/1,407명), 봉사단과전 102억 원(316명), 민관협력 17억 원(6건), 소규모무상원조 3억 원(3건), 국제기구지원실적 150억 원(4건) 등이다. 농림수산분야의 ODA를 포함한 대외무상원조실적액 중 증가와 관련된 지원 규모를 파악하기가 어렵고, 또한 그 규모도 그리 크지 않을 수도 있으나 향후 ODA

사업규모가 지속적으로 증가할 것이라는 점과 ODA 사업이 해당 국가나 지역의 농업발전의 토대를 조성함으로써 향후 농업협력의 거점지역으로 활용할 수 있다는 점을 고려하면 GSP 사업단과 한국 국제협력단(이하 KOICA), 농진청 등 국제협력사업 관련 기관과의 협력체계 구축은 우리나라의 채소종자 수출의 잠재적 시장 확대에 기여할 수 있을 것으로 예상된다.

GSP 사업단 중 식량종자사업단과 수산종자사업단은 ODA 사업의 후속효과로 해당 사업단의 수출 확대 기반을 고려한 바가 있다. 식량종자사업단의 경우 씨감자 종자와 생산시스템 2가지 모두를 수출하는 전략을 구상한 바 있으며, ODA 지원사업 등을 통한 해외 식량 원조 사업을 추진하여 대규모 씨감자 생산시설이 해당국에 보급된다면 사업단의 수출용 씨감자 대량 생산 과일롯 시스템 개발로 생산된 씨감자의 수출확대로 이어질 수 있을 것으로 전망된다. 수산종자사업단은 새로운 수출 대상국 개척 노력의 결과로 페루 시장을 새로 개척하였으며, 페루가 ODA 대상국가에 포함되어 있는 점에 착안하여 페루의 양식업을 활성화시켜 줌으로서 향후 우리나라 수산종자의 수출기반 확대 기회로 활용할 수 있을 것으로 기대된다. 또한 KOICA가 주관하는 농림수산 분야 ODA 사업을 통해 채소종자 수출확대 기반을 강화하기 위해서는 사업의 발굴과 기획, 수행, 평가 등 각 단계에서 구체적인 협력방안을 모색할 수 있도록 GSP 사업단과 KOICA 간의 업무협조약정(MOU) 체결 등 실질적인 협력체계를 구축할 필요가 있다.

채소 종자와 관련된 사업 추진에 있어서 GSP 사업단의 역할과 기능, 그리고 참여 방식 등을 구체화할 필요가 있다. 이를 위해 GSP 사업단 내 국제협력 분야만을 전담하는 조직의 구축 및 운영방안을 모색할 필요가 있으며, 국제협력에 관심이 있는 GSP 참여 사업단 전체를 대표할 수 있는 조직을 지향하는 것이 운영 효율성 측면에서 바람직할 것으로 사료된다.

또한 KOICA 이외에 농촌진흥청에서 주관하는 해외농업기술개발사업(Korea Program on International Agriculture: 이하 KOPIA)도 채소 종자 수출 확대를 위한 협력기반으로 활용할 필요가 있다. KOPIA는 개도국에 대한 맞춤형 농업기술지원과 자원의 공동개발을 통한 협력 대상국의 농업 생산성 향상을 유도하여 농업발전에 기여하는데 목적을 두고 있다. 농촌진흥청은 KOPIA의 운영을 위해 세계 20개국⁴⁾에 KOPIA 센터를 운영하고 있으며, 해당 국가와의 관계 기관과 양해각서 체결 및 ODA 사업을 통해 생산자 소득 증대 및 농업 발전기반을 조성하고 있다. 또한 농촌진흥청은 최근 KOPIA를 통해 베트남에서 농업분야 기술 협력을 통해 한국 채소 4개 작물 6개 품종이 베트남 국가장려품종으로 선정되는 성과를 거두었으며, 이를 토대로 한국 기업이 종자를 수출할 수 있는 기반을 마련한 바 있다.

대표적인 예로 2018년 4월 베트남 농업농촌개발부 전문농업과학 심의회에서 한국 채소 4개 작물 6개 품종을 베트남 국가장려품종으로 선정된 사례를 들 수 있다. 선정된 국가장려품종들은 참외 ‘슈퍼007꿀’, ‘참사랑꿀’ 품종과 주키니호박 ‘스타올’, ‘아주라’ 품종, 양배추 ‘CT-17’, 상추 ‘하청’ 등으로 이는 농촌진흥청과 국내 종묘회사가 육성한 일대잠종들이며, 시험 결과 재배에 알맞은 곳으로는 베트남 홍강델타와 북부 산악 지역이 권장되고 있다. 품종별 수출실적(’17)은 슈퍼007꿀 3.28kg(2,380달러), 스타올 500kg(50,000달러), CT-17 200kg(28,000달러)으로 알려져 있다. 이와 같이 베트남 현지에서도 국내에서 개발한 우수 품종의 특성이 널리 알려지면서 국내 업체를 통한 종자 수출도 진행되고 있어 이는 수출 확대로 이어질 것으로 기대된다.

이와 같은 베트남의 성공 사례는 농촌진흥청, 국립종자원, 사단법인 한국종자협회 간에 2015년에 이루어진 업무협력협약에서 비롯되었다. 그로 인해 국립종자원과 한국종합협회가 KOPIA 베트남 센터에 우수 품종을 추천하고 지원하면서 관계 부처와 기관 간의 성공적인 협업 모델로 평가받고 있

4) 아시아 9개국(베트남, 우즈베키스탄, 미얀마, 캄보디아, 필리핀, 스리랑카, 태국, 몽골, 라오스), 아프리카 6개국(케냐, 알제리, 에티오피아, 우간다, 세네갈, 짐바브웨), 중남미 5개국(파라과이, 볼리비아, 에콰도르, 도미니카공화국, 니카라과) 등임.

다. 2018년 3월에는 농촌진흥청과 KOICA 간에 농업분야 ODA 관련 상호협력을 강화하기 위한 MOU를 체결한 바 있으며, GSP 사업단도 이들 기관과의 협력체계 구축을 통해 개발도상국을 위한 국제개발협력사업에 참여함으로써 채소 종자 수출의 교두보를 마련해 나가는 방안을 모색할 필요가 있다.

제8절 국제협력사업 관계 기관들과의 협력체 구축을 통한 채소 종자 수출 활성화 기반 조성

1. 수출 확대 및 글로벌 경쟁력 제고와 관련된 정부의 종합계획 조사 및 구체적 방안 제시

○ 종자산업은 미래의 성장동력으로 발전가능성이 매우 큰 분야이다. 종자업의 경우, 세계 시장은 10년간 1.5배이상 성장했으며, 종자교역 규모 또한 2배 이상 증가한 추세이다. [*세계시장규모 : ('05) 247억불 → ('15) 372억불, 종자교역규모 : ('05) 48억불 → ('15) 107억불]. 그리고 육묘업의 경우 농업의 분업화 추세를 보이고 있으며 양질묘 수요 증가에 따라 육묘시장은 지속적으로 성장하고 있는 추세이다. [*국내시장규모: ('10) 1,870억원 → ('15) 3,881억원].

○ 작년에 한국종자기업은 456억 원 정도의 종자수출을 달성했고(한국종자협회. 2019), 현재, 2021년 GSP 수출목표인 2,000억 원 달성을 위해 순조롭게 나아가고 있다. 한국종자협회에 따르면 약 277업체가 육종, 생산, 판매 등에 관여하여 채소종자를 종자대리점 또는 시, 군, 면의 단위조합(농협)을 통해 보급하고 있다. 현재 채소종자시장은 약 2,279억원의 규모를 가지고 있다고 추정되며, 주요 작물은 고추, 무, 양파, 토마토, 배추, 수박 등으로 구성되어 있다. 근본적으로 시장이 포화되어 있어서 종자업계의 환경은 그다지 좋은 편이 아니다. 중국 등의 해외로부터 농산물 수입이 매년 증가하고 있고, 국내에서도 산업화와 인구 고령화가 심각한 수준에 이르러 작물의 경지면적이 지속적으로 줄어드는 추세이다. 또한 최근 지구온난화에 따른 이상 기후 현상 때문에 작물재배의 패턴이 변하고 이에 따라 재배 농민이 요구하는 품종의 성능수준도 높아지고 있다. 경쟁이 심하고 포화된 시장이라서 총액으로는 미미하다고 볼수 있다. 그러나 신품종에 부여되는 품종보호권이라는 배타적인 상업적 독점권의 중요성이 크게 부각되면서, 한국 정부는 종자산업이 미래의 국가경쟁력을 좌우할 중요한 분야로 인식, 2012년부터 10년간 걸친 GSP 등의 다양한 육성책을 수립하여 추진하고 있다. 그리고 농업실용화재단, 종자산업진흥센터 등이 선두에 서서 조직 배양 및 유전자원을 활용한 분자표지분석 등 생명공학기법을 이용한 종자육종 신기술을 도입하고 있다. 최근 전통적인 육종방법으로 해결하기 어렵거나 도입이 불가능한 형질을 도입해야 할 때 유전자 재조합 기술을 이용함으로써 다양한 형질을 가진 새로운 품종 즉, 유전자변형작물(GMO)과 형질전환물질, 그리고 유전자가위 기술(gene editing)에 초점을 맞춘 식물 육종 혁신(plant breeding innovation) 등의 연구가 정부의 연구소 및 대학에서 제초제 저항성, 내충성, 내병성, 내재해성 등의 도입과 다수확 고품질 품종의 육종을 목표로 진행되고 있다.

○ 품종보호제도 도입 면에서 주요농작물종자법과 종묘관리법을 통합하여 모든 식물종자를 대상으로 하는 종자산업법이 1995년 12월 6일 제정되고, 1997년 12월 31일부로 시행되었으며 국제신품종보호연맹(UPOV)에는 2002년 1월 7일, 50번째 가입국으로 등록하였다. 또한, 2017년 6월 28일부로 종자산업법이 육묘산업과 관련하여 육묘업 등록제, 품질표시 의무화 등의 내용을 포함하여 개정되어 2017년 6월 28일부로 종자산업법이 육묘산업과 관련하여 육묘업 등록제, 품질표시 의무화 등의 내용을 포함하여 개정되어 2018년 12월 28일부로 시행되고 있다.

☞ 전략별 중점추진과제는 다음과 같다.

- 첫째, 수출확대 및 글로벌 경쟁력 제고
 - ① GSP 2단계 성공적 마무리로 종자수출 2억불 달성, 수출확대 지원 강화
 - ② 종자산업 인프라 연계 시설 확충 및 협력체계 구축
 - ③ 식량종자 해외진출 기반 조성
- 둘째, 성과중심 연구개발 체계 구축
 - ① 미래환경 대응 우수품종 개발
 - ② 공공분야 품종개발기술 민간공유 확대
 - ③ 최신 육종기술 활용 가능성 고품질 품종 육성
- 셋째, 선진적 생산·유통 체제구축 및 제도개선
 - ① 육묘산업의 활성화
 - ② 국내 종자생산확대와 선진 유통질서 확립
 - ③ 보급종 종자 공급체계 개선
 - ④ 품종보호제도 전문성 및 효율성 강화
- 넷째, 전문인력양성 및 중소기업체 역량 강화
 - ① 종자산업 전문인력 양성 및 일자리 창출
 - ② 규모화된 중소기업체 주도형으로 체질 개선
 - ③ 업체 맞춤형 지원체계 전황으로 중견기업 육성

☞ 종자산업을 육성하기 위해서는 해결해야 할 몇 가지 우선 과제는 다음과 같다.

- 첫째, 종자산업의 국내외 여건 변화에 대응하여 채소종자 뿐만 아니라 모든 종자산업 구조를 정부주도에서 민간주도로 전환하여 민관이 유기적으로 역할을 분담하고 협력함으로써 시너지 효과를 극대화 할 수 있는 방안을 모색해야 한다.
- 둘째, 품종보호제도의 운영 활성화 측면에서, 국제경쟁력이 있는 새로운 품종이 나오기 위해서는 육성자의 권리보호가 철저히 이루어져야 한다.
- 셋째, 2006년 11월에 준공된 증장기 각각 50만점의 유전자원을 보존할 수 있는 세계적 규모의 유전자원 저장시설인 농촌진흥청 농업유전자원센터를 통하여, 체계적이고 효율적인 유전자원 종합관리 및 활용이 추진되어야 한다.
- 넷째, 국가연구기관은 기초연구를 담당하고 기업체는 실용화 품종연구를 분담하는 체제를 갖추어야 하며, 국가연구기관의 연구결과가 실질적으로 품종개발에 적극 활용될 수 있도록 협의, 조정되어야 한다.
- 다섯째, 민간 종자산업이 역동적인 산업구조로 되기 위해서는 민간 종자회사가 품종육성, 종자생산, 종자판매 등을 모두 하는 백화점식 경영에서 벗어나 회사별로 특정분야에 경쟁력을 가질 수 있도록 전문화 되어야 한다. 자본과 기술을 모두 갖춘 회사는 품종육성, 종자생산 및 판매 등을 종합적으로 취급하는 회사로 성장할 수 있지만, 자본과 기술이 부족한 회사는 자기 능력에 적합한 분야 또는 품목별로 전문화하여야 할 것이다.
- 여섯째, 글로벌 리더십을 가진 통찰력과 창조적 재능이 구비된 기업가 정신이 있어야 한다. 장기간의 투자가 필요한 종자산업은 재무적인 물적 자본뿐만 아니라 인적 자본, 그리고 신뢰, 규범, 신념, 참여, 제도와 규칙, 잘 형성된 사회적 연결망 등의 사회적 자본이 반드시 형성되어 있어야 할 것이다.

2. KOICA, 농촌진흥청(KOPIA) 등 국제협력사업 관련 기관들의 농업분야 ODA 사업 효과 및 채소 종자 수출 기반 조성 방향 제시

가. 국제협력사업 관련 기관들의 농업분야 ODA 사업 효과 및 채소 종자 수출 기반 조성 방향

(1) 종자업체들의 수출 현황

○ Golden Seed 프로젝트(이하 GSP)는 글로벌 종자 강국으로 도약을 위한 수출전략 종자 개발 및 산업기반 구축을 위한 4개 부·청 협력 프로젝트로 미래 농업환경 변화에 따라 새롭게 전개되고 있는 글로벌 종자시장을 선점함으로써 글로벌 종자강국으로의 실현을 목적으로 하고 있음.



○ GSP의 추진전략⁵⁾도 수출시장 성장모형에 따라 글로벌 시장개척, 품종보호 전략종자 개발, 그리고 관련기술개발 추진 등으로 설정됨.

- 글로벌 시장개척 종자 : 현재 보유하고 있는 강점기술을 기반으로 수출시장 개척용 종자 개발
- 품종보호전략종자 : 품종보호 기반 확보 및 장기적 수출시장 진입용 종자 개발

○ GSP 사업의 목표 달성을 위해 그간 정부에서는 국가적 역량결집, 민간 수출역량 강화, 목표시장 개척을 위한 수출형 R&DB 통합지원을 시행해 왔으나, 2018년 수출액은 52백만 달러로 목표치인 120백만 달러의 절반 수준에도 미치지 못하고 있음(국립종자원).

- 2018년 종자 수출량은 2017년 1,336톤에서 11.4% 증가한 1,488톤이었으나, 수출액은 2017년 58,542천\$보다 6,245천\$(10.7%) 감소한 52,297천\$이었음.
- 채소종자 수출량은 동기간에 616톤에서 569톤으로 47톤(7.6%) 감소하였으며, 수출액도 57,773천\$에서 50,766천\$로 7,007천\$(12.1%) 감소함.
- 전체 수출종자 중 채소종자가 차지하는 비율은 물량기준으로는 38.2%, 금액기준으로는 97.1%로 전체 종자 수출에서 대부분을 차지한 것으로 나타남.
- 목표치 미달성의 원인으로는 글로벌 종자로 개발된 품종의 현지화 지원 등 해외 수출 정책 미흡과 총괄 컨트롤타워 부재, 기 구축된 인프라의 효율적 활용 미흡 등이 원인으로 거론됨.

5) 정부는 제2차 종자산업 육성 5개년 계획에서 주요 개선 방향으로 GSP 2단계 사업에서 연구개발 방향을 로열티 대응사업(로열티 수출 품종개발까지 확대), 차세대 신육종 기술 활용 연구개발 사업, 미래 환경 대응 우수품종 개발로 수정하였으며, 인프라의 경우 해외진출 및 수출 확대 대응기반 조성을 주요 방향으로 설정한 바 있음.

표 1. 우리나라 종자 수출입 현황(2017년, 2018년)

단위 : 톤, 천\$

| | | 수출(A) | | 수입(B) | | 무역수지(A-B) | |
|--------------|-------------|---------------|------------------|-----------------|------------------|-------------------|-------------------|
| | | 물량 | 금액 | 물량 | 금액 | 물량 | 금액 |
| 2018 (a) | 전체 | 1,488 | 52,297 | 17,858 | 126,746 | △16,370 | △74,449 |
| | 채소종자 (%) | 569 (38.2) | 50,766 (97.1) | 2,091 (11.7) | 75,834 (59.8) | △1,522 (267.5) | △25,068 (49.4) |
| 2017 (b) | 전체 | 1,336 | 58,542 | 16,243 | 115,892 | △14,907 | △57,350 |
| | 채소종자 (%) | 616 (46.1) | 57,773 (98.7) | 2,172 (13.4) | 73,479 (63.4) | △1,556 (10.4) | △15,706 (27.4) |
| 증△감 (a-b) | 전체 (%) | 152 (11.4) | △6,245 (10.7) | 1,615 (9.9) | 10,854 (9.4) | △1,463 (9.8) | △17,099 (29.8) |
| | 채소종자 (%) | △47 (7.6) | △7,007 (12.1) | △81 (3.7) | 2,355 (3.2) | △34 (2.2) | △9,362 (59.6) |

자료 : 국립종자원(2018.12), 『종자 수출입 물량 현황』.

○ 2017년 기준 우리나라 종자업 등록업체 중 판매실적이 있는 업체수는 1,152개로 연 매출액이 5억 미만인 업체가 전체의 87.0%인 1,002개로 산업구조가 영세한 상황이며, 해외수출 실적이 있는 업체 비중은 7%이었음.

- 매출액 40억 원 이상(대규모)은 29개, 15~40억 원(중규모)은 44개, 5~15억 원(중소규모) 77개, 5억 원 미만(소규모)은 1,002개 업체로 2016년에 비해 다소 양극화 된 것으로 나타남.
- 2017년 기준 판매실적이 있는 종자업체의 93.0%는 국내 시장에서만 판매가 이루어졌으며, 국내 판매 및 해외 수출을 병행하는 업체는 4.9%, 해외 수출만 하는 업체는 2.2%(25개)임.

표 2. 판매실적 존재 종자업체의 종자 유통지역(2016년, 2017년 기준) 단위 : 업체수, %

| 구분 | 국내(판매)만 | | 해외(수출)만 | | 국내(판매)와 해외(수출) 병행 | | 계 | | |
|------------|-----------------|-----------------|---------------|-------------|-------------------|--------------|------------------|------------------|------------------|
| | 2016년 | 2017년 | 2016년 | 2017년 | 2016년 | 2017년 | 2016년 | 2017년 | |
| 전체 | 1,084 (93.7) | 1,071 (93.0) | 5 (0.4) | 25 (2.2) | 68 (5.9) | 56 (4.9) | 1,157 (100.0) | 1,152 (100.0) | |
| 판매액 규모별 | 소규모 | 963 (96.8) | 967 (96.5) | 3 (0.3) | 16 (1.6) | 29 (2.9) | 19 (1.9) | 995 (100.0) | 1,002 (100.0) |
| | 중소규모 | 78 (80.4) | 59 (76.6) | 2 (2.1) | 5 (6.5) | 17 (17.5) | 13 (16.9) | 97 (100.0) | 77 (100.0) |
| | 중규모 | 35 (76.1) | 29 (65.9) | 0 (0.0) | 3 (6.8) | 11 (23.9) | 12 (27.3) | 46 (100.0) | 44 (100.0) |
| | 대규모 | 8 (42.1) | 16 (55.2) | 0 (0.0) | 1 (3.4) | 11 (57.9) | 12 (41.1) | 19 (100.0) | 29 (100.0) |

자료 : 2017 종자업 실태조사(2016년 기준), 2018 종자업 실태조사(2017년 기준).

주 1) 판매 실적이 있는 종자업체만을 대상으로 산정한 수치임.

2) () 내는 계에 대한 구성비임.

표 3. 판매실적 존재 종자업체의 연간 판매액(2016년, 2017년 기준) 단위 : 억 원

| 구분 | | 종자관련 판매액 | | 종자 외 판매액 | | 총 판매액 | |
|----|----------------------|----------|---------|----------|---------|-------|----------|
| | | 2016년 | 2017년 | 2016년 | 2017년 | 2016년 | 2017년 |
| 전체 | 국내 | 4,722 | 5,208.8 | 3,918 | 4,753.1 | 8,641 | 9,961.8 |
| | 해외(수출) | 655 | 687.5 | 143 | 95.8 | 798 | 783.3 |
| | 해외 생산 후 해외 판매(수출) | 31 | 22.4 | 0 | 0 | 31 | 22.4 |
| | 계 | 5,408 | 5,918.7 | 4,061 | 4,848.9 | 9,469 | 10,767.5 |

자료 : 2017 종자업 실태조사(2016년 기준), 2018 종자업 실태조사(2017년 기준).

주 1) 판매 실적이 있는 종자업체만을 대상으로 산정한 수치임.

- 2017년 기준 종자 판매실적이 있는 종자업체(1,152개)의 총 판매액은 1조 767억 원으로 2016년 종자업체(1,157개)의 총 판매액 9,469억 원보다 13.7% 증가하였음.
- 종자판매액은 2017년 5,919억 원으로 2016년 5,408억 원보다 9.4% 증가하였고, 종자 외 판매액은 2016년 4,061억 원에서 2017년 4,849억 원으로 19.4% 증가하였음.

- 종자의 해외 수출액도 2016년 655억 원에서 2017년 688억 원으로 4.7% 증가하여 정부의 민간 수출역량 강화, 목표시장 개척을 위한 수출형 R&DB 통합지원 등과 같은 사업 등이 2017년에는 일정 수준 긍정적인 효과로 이어졌으나, 2018년 종자 수출액이 2017년보다 10.7% 감소한 것으로 나타나 지속적인 종자 수출 확대를 위한 방안을 모색할 필요가 있음.

나. 우리나라 국제농업협력사업

- 우리나라는 2009년 11월에 OECD 개발원조위원회(Development Assistance Committee: DAC) 가입 이후 공적개발원조(Official Development Assistance: ODA⁶⁾) 규모는 지속적으로 증대되어 왔으나 2017년 총 국민소득대비 ODA 비율(ODA/GNI)은 0.14%로 UN의 권고치인 0.7%에 비해 크게 낮은 상황임.

- ODA를 포함한 우리나라의 대외무상원조실적은 2013년 1조 9,214억 원에서 2017년 2조 4,930억 원으로 30% 가량 증가하였으며, 2019년 ODA 예산은 3조 4,922억 원으로 2017년 실적 대비 40.1% 증가하였음(국제개발협력위원회, 2018).
- 2019년 양자간 원조는 2조 7,049억 원으로 전체 ODA 예산의 77.5%를 차지하고 있으며, 이 중 무상원조는 1조 5,437억 원, 유상원조는 1조 1,612억 원임.

6) ODA는 한 국가의 중앙 혹은 지방정부, 공공기관이 개발도상국의 경제발전과 복지증진을 위해 개발도상국이나 국제기구에 제공하는 자금의 흐름을 말함.



그림 1. 국제농업협력사업 사업비와 사업수 현황

자료 : 농림축산식품부(www.mafra.go.kr).

주 : 2019년은 예산액 기준임.

- 우리나라는 개도국에 우리 농업분야 성공경험을 전수하고, 농촌 개발 및 빈곤 퇴치 지원을 목적으로 농림축산식품부를 중심으로 국제농업협력사업을 추진하고 있음.
 - 2006년부터 국제농업협력사업을 추진하였으며, 2011년부터는 농업 인프라 지원 유형의 중장기·대규모 사업을 확대하였고, 2018년부터는 국내산 쌀을 활용한 식량원조를 추진하고 있음.
- 최근 10년간 국제농업협력사업에 투입된 사업비와 사업수는 지속적으로 증가하는 추이를 보이고 있음.
 - 사업비는 2010년 49억 원에서 2018년 651억 원으로 13배가량 증가하였으며, 사업수는 동기간에 17개에서 30개로 두 배 정도 증가함⁷⁾.
 - 2019년에는 아시아 6개국, 아프리카 3개국, CIS(Commonwealth of Independent States: 독립국가연합) 및 중남미 1개국 등 12개국과, 4개 국제기구(FAO, IFAD, ADB, WFP) 대상으로 28개 사업을 추진하고 있음.

다. 사업 유형

- 사업 유형은 기획협력사업, 개도국 농정 컨설팅, 다자성 양자사업 등 3가지로 구분되며, 각 유형별로 시행기관과 예산 규모에 차이가 있음.

(1) 우리나라와 수원국 간 직접 협의를 거쳐 추진하는 사업(기획협력사업)

- 기획협력사업은 3~5년에 걸쳐 농업 인프라를 제공하고, 운영교육과 컨설팅 등 소프트웨어를 함께 지원하는 사업으로 농촌종합개발, 영농기술 전수, 기계화 영농, 농업 관개 등 다양한 분야를 지원함.

7) 2018년부터 시작한 국내산 쌀 식량원조 사업비 460억 원을 제외하면 4배가량 증가함.

○ 시행기관은 한국농어촌공사, 한국농촌경제연구원, 농림수산식품교육문화정보원, 축산물 품질평가원 등 4개 기관이며, 2019년 기준 사업비 규모는 167억 원으로 12개국에서 20개 사업을 추진하고 있음.

(2) 개도국 농정 컨설팅(KAPEX: Korean Agricultural Policy Experience for Food Security)

○ 본 사업은 연간 3개 개도국의 수요에 따라 분야를 선정하여 공동조사, 초청연수, 워크숍 등을 시행하는 사업임.

○ 시행기관은 한국농촌경제연구원이며, 2019년에는 12억 원의 사업비로 미얀마, 인도네시아, 필리핀 등 3개국에 대한 농정 컨설팅을 추진하고 있음.

(3) 국제기구를 통한 사업(다자성 양자사업)

○ 본 사업은 FAO, IFAD, ADB 등을 통해 개도국 농업기술 지도, 유통구조 개선 등을 지원하고, WFP를 통해 개도국에 식량 원조를 하는 사업임.

○ 시행기관은 FAO, IFAD, ADB, WFP 등이며, 2019년 예산은 486억 원으로 5개 사업을 추진하고 있음.

라. 사업추진체계

○ 사업추진은 사업 기획, 사업 시행, 종료 후 평가 및 관리 등 3단계로 이루어져 있음.

○ 사업 기획은 사업시행 2~3년 전부터 시작하게 되며, 사전 공동조사, 컨설팅을 통해 개도국의 현지 수요를 직접적으로 제안을 받아 사업을 발굴하게 되고, 제안된 사업 중 일부를 선정하여 수원국과 협의 이후 전문가 그룹의 사업 심의와 타당성 조사를 거쳐 신규 예산의 반영을 추진하게 됨.

- ODA 예산은 외교부 무상원조 적합 심사 및 국무조정실 유무상 원조 종합심사 이후 기재부에서 반영하게 됨.

○ 사업 시행은 한국농어촌공사 등 기관에서 관리하게 되며, 수원국 현지에서의 사업 수행은 외부 용역사(Project Management Consultancy)에 위탁함.

- 외부 용역사는 주로 엔지니어링 회사, 대학의 산학협력단 등 국내 기관이며, PMC 업체 직원을 사업관리자(PM)로 파견하여 현지에서 설계업체, 건설업체 등을 고용함.

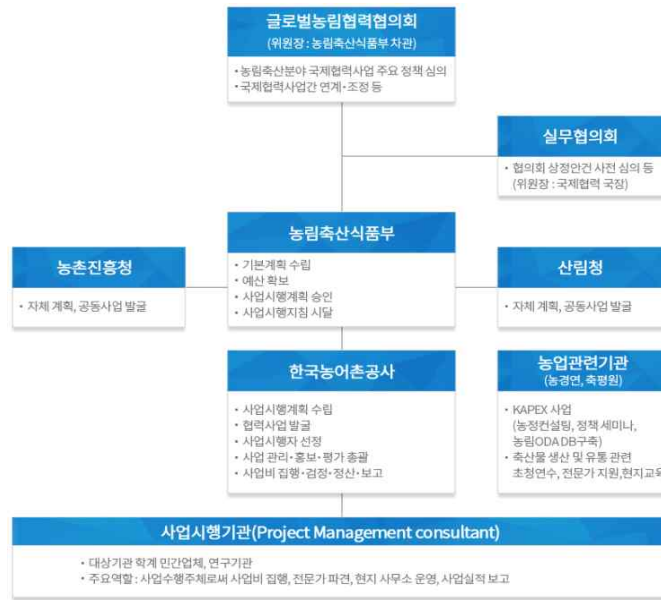


그림 2. 국제농업협력사업 사업추진체계도

자료 : 농림축산식품부(www.mafra.go.kr).

- 종료 후 평가 및 관리는 통상적으로 사업 종료 시점으로부터 2~3년이 경과된 후에 시행하며, 사후평가를 통한 성과 및 문제점을 도출하여 향후 사업 추진 시 반영하게 됨.
 - 사업 종료 후에도 별도 예산을 편성하여 문제 사업에 대한 시설을 보수하거나, 현지 주민들의 재교육 등 지속적인 관리가 이루어짐.

마. 유관기관의 국제농업협력사업

- 농림축산식품부 이외에도 소속기관, 농촌진흥청, 산림청 등은 해당 기관의 전문 분야별로 예산을 확보하여 농림분야 ODA 사업을 추진하고 있음.
 - 농촌진흥청은 2019년 기준 211억 원의 예산으로 해외농업기술개발사업(Korea Program on International Agriculture: KOPIA)을 추진하고 있으며, 산림청은 조림사업 등에 129억 원의 예산을 투입하고 있음.
 - 이들 기관 이외에도 농림축산검역본부(2.3억 원), 국립종자원(0.9억 원), 농산물품질관리원(7천만 원) 등도 국제농업협력사업을 추진하고 있음.
- 농림분야 유관기관들의 ODA 사업을 함께 논의하기 위해 2010년부터 『글로벌 농림협력협의회』를 농림축산식품부 차관 주재로 연 1~2회 운영하고 있음.
 - 협력위원회에는 농촌진흥청, 산림청, 농림축산식품부 소속기관, 한국농어촌공사, 한국농촌경제연구원, 농림수산물교육문화정보원, 축산물품질평가원, 해외농업자원개발협회 등이 참석함.

3. 국제농업협력사업 관계기관과의 협력을 통한 채소 종자 수출 기반 조성 방향

- 향후 국제농업협력사업(ODA)의 규모는 지속적으로 증가하고, ODA 사업이 해당 국가나 지역의 농업발전 토대를 조성함으로써 향후 농업협력의 거점지역으로 활용할 수 있음.
 - GSP 사업단의 종자 수출을 활성화하기 위해서는 한국국제협력단이나 농림축산식품부 산하 국제농업협력사업 추진 관계기관들과의 긴밀한 협력체계 구축은 우리나라의 채소 종자 수출의 잠재적 시장 확대에 기여할 수 있을 것임.

- 최근 GSP 사업단 내 식량종자사업단과 수산종자사업단은 ODA 사업 후속효과로 해당 사업단의 수출확대 기반을 고려한 바 있으며, 농촌진흥청, 국립종자원, 사단법인 한국종자협회 간에 2015년에 이루어진 업무협력협약을 통한 종자 수출 기반 조성 등의 사례는 우리나라의 종자 수출 확대 측면에서 중요한 시사점을 제공한다고 할 수 있음⁸⁾.
- 2021년 농림축산식품부 국제농업협력사업 사전 심의 대상 사업들 중 『베트남 채소 가치사슬 확대를 통한 소농의 지속가능한 생계향상 프로젝트』에서 GSP 종자 보급이 언급됨.
 - 베트남 프로젝트의 경우 기자재 지원 사업에서 ‘고추와 배추의 GSP 종자 보급 및 판매를 통한 한국 종자 산업 해외 진출 기반 마련’이라는 내용이 포함되어 있으며, 캄보디아 프로젝트에서는 채소 및 씨감자 보급이 주요 내용으로 포함되어 있음⁹⁾.

- GSP 사업단 또는 GSP 사업을 관리하고 있는 농림식품기술기획평가원 내에 종자 관련 국제농업협력사업을 관리하고, 농림축산식품부 및 관계기관들의 협력을 담당하는 조직을 신설·운영할 필요가 있음.
 - 해당 조직은 국제농업협력사업 발굴 및 기획 단계인 사전 공동조사, 컨설팅 등에 참여함으로써 GSP 사업단 내 종자 기업들의 수원국 진출과 수출 기반 조성에 기여할 수 있도록 함.
 - 이를 위해 해당 조직이 농림식품기술기획평가원이나 GSP를 대표하여 『글로벌 농림 협력협의회』에 참여할 수 있도록 하고, 종자 부문 국제농업협력사업 발굴을 위해 각 GSP 사업단에 참여하고 있는 민간 종자 기업들과의 협의회를 정기적으로 운영할 필요가 있음.

- 『글로벌 농림 협력협의회』 참여 이외에도 『농식품산업 해외진출 지원사업』 대상 기업들과의 종자 부문의 협력을 강화하도록 함.
 - 『농식품산업 해외진출 지원사업』은 민간의 해외농업 진출 및 정착을 지원하여 우리나라 농식품산업의 저변확대와 국제경쟁력을 확보하고, 미래 해외 식량 확보 기반을 마련하는 사업임.
 - 본 사업은 농산물 생산 및 유통에 필요한 농기계구입, 부대시설, 건조·저장·가공 시설 설치에 필요한 비용 및 영농비에 대해 융자금을 제공하고 있음.

8) 해당 사례들의 구체적인 내용은 『2018년 채소종자사업단 연차실적보고서』 참조.

9) 베트남과 캄보디아 프로젝트는 한국농촌경제연구원이 KAPEX 사업을 통해 발굴·기획된 사업이며, 해당 사업 책임자는 사업 기획 단계에 GSP 사업단이 참여할 경우 시너지 효과를 얻을 것으로 기대하고 있음.

- 동 사업을 통해 해외에 진출한 기업들 중 현지에 적절한 종자 선정 및 확보에 어려움을 겪을 수 있어 본 지원사업의 시행주체인 한국농어촌공사, 해외농업자원개발협회와 협력 체계를 구축함으로써 종자 수출 기반을 확대하는 효과를 기대할 수 있음.

4. 해외종자관련 특허현황 및 흐름 파악을 통한 해외 종자시장 현황 및 경향 분석

○ 분석범위

GSP채소종자사업단의 주요 품목인 고추, 배, 무, 수박, 파프리카를 대상으로 국내 및 해외 출원·공개된 특허기술을 대상으로 함. 품종특허, 육종관련 기술(재배기술, 조직배양기술, 수분·수정기술, 옹성불임 등), 분자표지(품종구별, 주요형질선별 등), 농업형질(색상, 과실 경도, 수량 증대 등) 관련 기술을 포함함 본 동향조사를 통해 주요 품목인 고추, 배, 무, 수박, 파프리카에 대한 국·내외 기술동향을 알아보하고자 함

○ 검색범위

특허검색 DB는 KIPRIS(한국특허정보원)와 WIPS를 주로 사용하고 그 외 USTPO(미국특허청), EPO(유럽특허청), JPO(일본특허청) 등을 사용하여 특허 검색을 수행함
특허검색구간은 2008.01.01.-2019. 05.로 하며, 정량분석 구간은 2017. 12. 31. 공개된 특허를 대상으로 함

특허검색 범위는 아래와 같음

한국 : 특허공개 및 특허공고/ 등록된 전체문서

미국 : 특허공개 및 특허 등록된 전체문서

일본 : 특허공개, 특허공개(공표), 특허공개(재공표) 및 특허공고/ 등록된 전체 문서

중국 : 특허공개 및 등록된 전체문서

유럽 : EP-A(Applications) 및 EP-B(Granted) 전체 문서

○ 기술 분류

GSP채소종자사업단의 주요 품목인 고추, 배, 무, 수박, 파프리카를 중분류로 구분하였음
각 중분류는 품종, 육종기술, 분자표지, 농업형질로 구분하였음

표 4. 분석대상 기술 분류

| 대분류 | 중분류 | 소분류 |
|--------------------|------|------|
| 주요 채소작물 육종기술 현황 | 고추 | 품종 |
| | | 육종기술 |
| | | 분자표지 |
| | | 농업형질 |
| | 배추 | 품종 |
| | | 육종기술 |
| | | 분자표지 |
| | | 농업형질 |
| | 무 | 품종 |
| | | 육종기술 |
| | | 분자표지 |
| | | 농업형질 |
| | 수박 | 품종 |
| | | 육종기술 |
| | | 분자표지 |
| | | 농업형질 |
| 파프리카 | 품종 | |
| | 육종기술 | |
| | 분자표지 | |
| | 농업형질 | |

○ 유효특허

주요 채소작물 육종기술 현황의 총 유효 데이터는 총 1,055건이 선별되었으며, 정량분석 대상 특허(2008.01.01.~2017.12.31.)는 총 932건이 선별됨
중분류간 중복건을 허용하였음

표 5. 국가별 특허 검색결과

| | 한국 | 미국 | 일본 | 중국 | 유럽 | 국제 | 전체 |
|--------|----|-----|----|-----|----|----|-------|
| 유효특허 | 87 | 291 | 15 | 565 | 53 | 44 | 1,055 |
| 정량분석대상 | 87 | 272 | 15 | 463 | 53 | 42 | 932 |

표 6. 기술 분류별 정량분석 대상 건수

| 중분류 | 소분류 | 한국 | 미국 | 일본 | 중국 | 유럽 | 국제 | 전체 |
|------|------|----|-----|----|-----|----|----|-----|
| 고추 | 품종 | 4 | 141 | 3 | 0 | 0 | 0 | 148 |
| | 육종기술 | 5 | 3 | 1 | 47 | 1 | 1 | 58 |
| | 분자표지 | 4 | 12 | 0 | 18 | 11 | 8 | 53 |
| | 농업형질 | 14 | 19 | 5 | 23 | 13 | 12 | 86 |
| 배추 | 품종 | 2 | 0 | 0 | 0 | 0 | 0 | 2 |
| | 육종기술 | 16 | 4 | 0 | 141 | 2 | 0 | 163 |
| | 분자표지 | 2 | 6 | 0 | 48 | 4 | 3 | 63 |
| | 농업형질 | 12 | 10 | 0 | 55 | 7 | 6 | 90 |
| 무 | 품종 | 0 | 2 | 0 | 2 | 0 | 0 | 4 |
| | 육종기술 | 4 | 2 | 0 | 38 | 1 | 0 | 45 |
| | 분자표지 | 3 | 0 | 1 | 14 | 1 | 0 | 19 |
| | 농업형질 | 9 | 1 | 1 | 15 | 2 | 1 | 29 |
| 수박 | 품종 | 2 | 37 | 2 | 1 | 1 | 1 | 44 |
| | 육종기술 | 8 | 10 | 2 | 37 | 3 | 1 | 61 |
| | 분자표지 | 1 | 4 | 0 | 13 | 7 | 6 | 31 |
| | 농업형질 | 3 | 12 | 0 | 8 | 2 | 3 | 28 |
| 파프리카 | 품종 | 2 | 12 | 0 | 0 | 0 | 0 | 14 |
| | 육종기술 | 1 | 0 | 0 | 6 | 0 | 0 | 7 |
| | 분자표지 | 0 | 1 | 0 | 1 | 1 | 1 | 4 |
| | 농업형질 | 0 | 1 | 0 | 4 | 1 | 0 | 6 |

○ 주요 채소작물 육종기술 현황

출원일 기준 유효구간¹⁰⁾ 중 주요 채소작물 육종기술 분야의 특허 출원 동향을 분석한 결과, 작물별 출원 점유율은 고추 36%, 배추 34%, 수박 17%, 무 10%, 파프리카 3% 순으로 나타났으며, 전체적으로 고추 관련 육종기술은 지속적으로 증가하는 추세를 보이는 반면 배추 관련 육종기술은 증가와 감소를 반복하는 경향을 보이고 있음

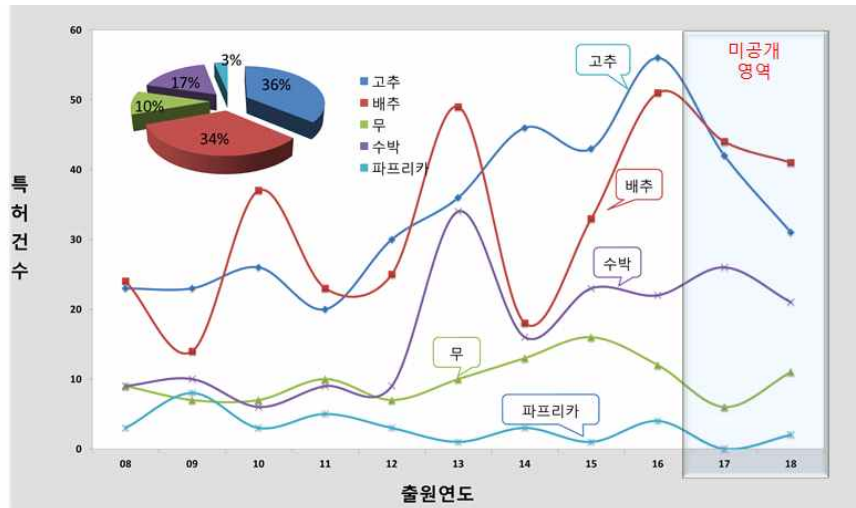


그림 3. 주요 채소작물 육종기술 분야 연도별 기술동향

○ 특허출원 구간별(2008-2009, 2010-2011, 2012-2013, 2014-2015, 2016-2017, 2년 단위로 5개 구간) 점유율 변화를 살펴보면, 고추, 배추, 수박 관련 기술은 구간별로 지속적으로 특허건수가 증가하고 있음

○ 고추는 2008-2009에 35%점유율에서 2016-2017에 37% 점유율로 2% 증가하였으며, 배추는 2008-2009에 39%에서 2016-2017에 36%로 7% 증가를 보여주고 있으며, 수박은 2008-2009에 15%에서 2016-2017에 18%로 3% 증가를 보여준 반면, 무와 파프리카는 특허건수 및 점유율이 감소하였음

10) 한국, 미국, 중국, 일본, 유럽 및 국제(PCT)특허: 출원일 기준으로 분석하며, 일반적으로 특허 출원 후 18개월이 경과된 때에 출원 관련정보를 대중에게 공개하고 있음. 따라서 아직 미공개 상태의 데이터가 존재하는 2018~2019년 자료는 유효하지 않으므로 정량분석은 ~2017년까지 유효데이터로 분석함.

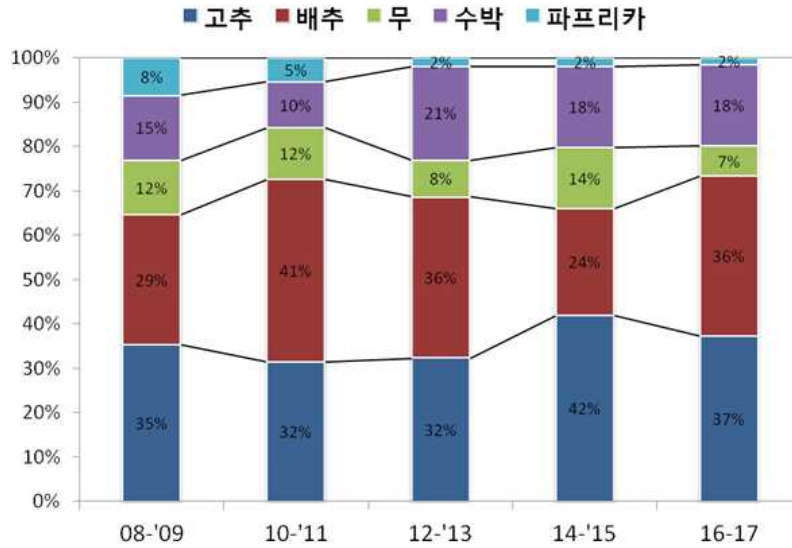


그림 4. 주요 채소작물 육종기술 분야 특허출원 점유율

○ 주요시장국별 기술 점유율을 살펴보면, 한국특허청(KIPO)은 배추 35%, 고추 29%, 무 18%, 수박 15%, 파프리카 3%순으로 나타났으며, 일본특허청(JPO)은 고추 60%, 수박 27%, 무 13%로 나타났으며, 중국특허청(SIPO)은 배추 52%, 고추 19%, 무 15%, 수박 12%, 파프리카 2%순으로 나타났으며, 미국특허청(USPTO)은 고추 63%, 수박 23%, 배추 7%, 파프리카 5%, 무 2%순으로 나타났으며, 유럽특허청(EPO)은 고추 44%, 배추와 수박이 각각 23%, 무 7%, 파프리카 3% 순으로 나타남

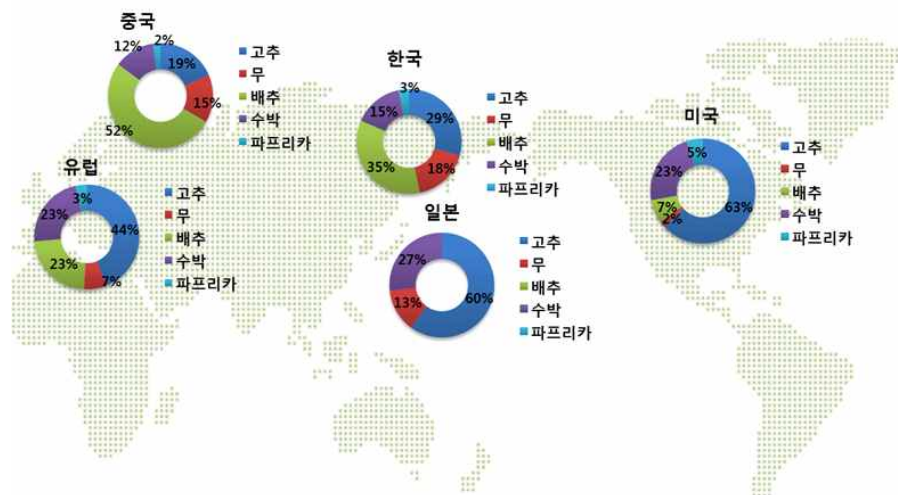


그림 5. 주요 채소작물 육종기술 분야 주요시장국 기술 점유율

5. 베트남 등 해외에서의 육종특화 시험포 사업 지원시스템 구축을 통한 수출 활성화 기반 조성

○ 베트남 채소 품종은 농업생산의 중요한 요인이며 우수 품종은 채소의 생산성을 15~20%로 늘릴 수 있다. 현재 종자생산 과학기술은 발전하는 중이므로 이는 곧 우수한 신품종이 많이 육종됨을 의미한다. 베트남의 다양한 자연조건은 다양한 채소류의 재배를 가능하게 한다. 농민의 채소 재배 경험과 첨단 과학기술이 적용되면 채소는 벼보다 2배의 수익성을 가질 수 있다. 최근에는 채소 시비량 및 가공량이 증가하는 추세이다. 이는 채소 재배 면적 확대와 우수 품종에 대한 요구 증가를 의미한다.

○ 1997년에서 2010년간 채소 재배 면적은 37.7만ha에서 70만ha로 증가했다(85.67%). 이에 베트남 농업은 재배면적 대비 생산량이 저조하여 내병성, 내충성, 이상기후에 잘 적응하는 우수 종자의 필요성을 가지게 되었다. 채소의 주요 재배 지역은 메콩델타(221,800ha)지역과 홍강델타(166,200ha)지역이다. 그리고 채소를 더욱 전문적으로 재배하는 지역은 도시근교이다. 이는 도시민들에게 다양한 채소를 공급(80품종, 주요 15품종 포함)하기 위해서이다. 농업농촌개발부 발전계획에 따라 2020년까지 총 채소재배면적이 120만ha까지 증가할 것으로 예상된다.

○ 2009년의 베트남 채소종자 소비량은 8,000톤이었다. 이 중 약 4,160톤(50% 이상)이 수입 품종이고, 약 3,280톤(41%)은 자가 채종, 나머지 약 560톤(7%)은 베트남 국내 품종이다. 수입품종은 병해충 저항성과 생산성이 높다. 품질이 좋지만 가격이 높고 베트남 기후에 대한 적응성이 낮다. 그리고 자가채종은 병해충 저항성과 생산성이 낮으며, 종자의 품질이 안 좋다. 베트남국내 품종은 다양하지 않고 약 70% 이상이 F1품종이다. 주요 품종은 토마토, 오이, 여주, 고추, 유채이다. 아직은 우수품종이 별로 없고, 병해충 저항성이 높은 품종도 없다.

○ 2011년과 2012년도 베트남은 채소류 종자 45품종 이상을 수입했다. 2011년도 채소 종자 총 수입액 약 21천만불(수입량 : 1,731톤)이고, 2012년도 채소 종자 총 수입액은 약 25천만불 (수입량 : 2,616톤)이었다. 그 중 가장 많이 수입되는 채소종자는 박과, 겨자과, 가지과 (총 수입액의 76.84%) 등이다. 2012년 종자를 수입하는 회사는 총 59개가 있다. 북부지방에 20개, 남부지방에 39개가 있다. 하지만 2012년에 수입을 하지 못한 회사가 10개였다.(북부 4개, 남부 6개). 상위 20여개의 회사의 수입액이 전체 회사 수입액의 90%정도를 차지하고 있다.



자료 : <http://www.farmnmarket.com/news/article.html?no=2613>

제9절 종자산업 발전 및 마케팅 전략 구축을 위한 기반연구

1. 지속가능하고, 안전하며, 건강한 채소종자 보급을 위한 유통망 구축

○ 글로벌화 되어가는 시대의 변화에 맞게 채소종자시장 역시 그 영향을 받고 있다. 농산물 수입개방에 따른 국제화에 대응하여 우리나라 농산물의 유통방식 또한 그에 맞게 변화되어야 한다. 이러한 상황에 있어 유통구조의 변화와 새로운 정책대안은 필수이며, 농산물의 원활한 유통과 소비자의 니즈를 충족시키기 위한 노력이 절실히 요구된다.

○ 농산물 유통에 있어서는 농산물의 정직한 가격을 받는 것이 요구되며, 공정거래를 통한 가격의 불균형을 해소할 수 있는 방안이 강구되어야 할 것이다. 또한, 농산물의 출하선택의 폭을 넓혀 농산물의 안정적인 상품공급과 적절한 수요와 공급망을 유지하도록 노력해야 한다. 그리고 투명한 유통과정을 유지하기 위해 효율적인 유통정책과 이를 집행하기 위한 정확한 유통정보와 통계가 필요하다.

○ 농산물 유통구조 개선의 궁극적인 목적은 생산자부터 소비자에게까지 전달되는 일련의 과정에 있어 생산자에게는 농산물의 적정가격보장과 생산의 효율성과 극대화를 위한 시설개선이 되어야 할 것이며, 소비자에게는 니즈에 맞는 적정가격의 최상품질의 농산물을 제공받는 것이 될 것이다. 단순히 경영 효율성을 높혀 유통비용이나 마진을 낮추는 방식의 단편적인 방식이 아닌 실질적이며 단계적인 과정에 입각한 유통구조의 변화가 요구되며 이를 반영할 시, 최상의 품질을 최적의 가격에 유통시킬수 있는 유통망이 형성될 수 있을 것으로 기대된다.

○ 농산물의 특성상 계절 간 특성으로 인해 가격의 불규칙성이 크며, 유통마진또한 과도한 편이다. 게다가 값싼 수입 농산물들에 밀려 소비자 시장에서 가격경쟁을 비롯한 질적 측면에 있어서도 새로운 전략이 필요한 상황이다. 우리나라 농산물의 경쟁력 향상을 위해서는 단순히 생산과정에만 신경을 써야할 문제가 아니다. 왜냐하면 농산물은 다른 공산품들과 달리 생산과정이 유통과정에 종속된 구조이기 때문이다. 이는 1차산물이 가진 특성으로 계절성, 저장성, 부패성이 이에 기인한다고 볼 수 있다. 또한, 농업인구가 소규모 단위로 지역마다 다르게 나뉘어 있고 소매점과 도매시장의 비중또한 무시할 수 없는 처사이기 때문이다.

○ 이를 해결하기 위해서는 도매시장의 유통구조개선이 시급하며, 직거래형태 또는 인터넷 또는 어플 사용을 통한 유통경로의 변화 역시 지켜볼 필요가 있다. 그리고 지속적인 마케팅 및 홍보를 통하여 국민들의 적극적인 관심을 이끌어내는 노력이 강구된다.

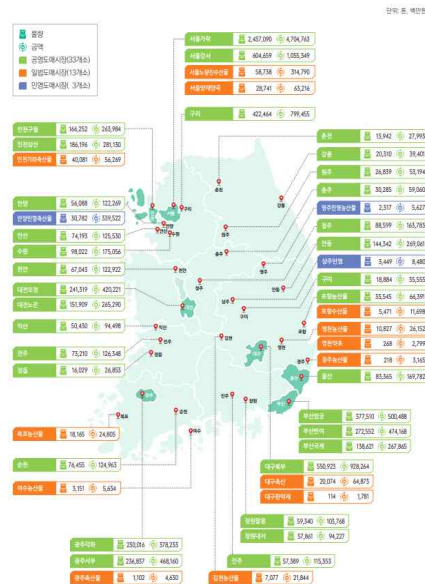


그림1.2018년도 도매시장 거래실적

자료: 농림축산부, 2018농수산물도매시장통계연보

2. 질적 성장을 통한 차별화 및 품질 경쟁력강화를 기반으로 한 우리나라 종자산업의 세계화 방안 모색

○ 종자는 모든 산물의 근본이다. 작물을 생산하기에 있어 시작점이며, 끝으로 볼 수 있다. 종자에는 생산요소인 토지, 노동, 자본이 들어가 있으며 기술의 영역까지 포함된다. 종자산업은 과거의 단순 교배육종 접근에서 벗어나 최근에는 의약 및 재료산업과의 융복합산업화 등 부가가치가 높고 첨단과학기술과의 접목이 용이한 분야로 변화하고 있다. 그렇기 때문에 선진국들은 이미 종자산업을 한 나라의 기반을 구축하는 데에 없어서는 안 될 필수불가결의 수단으로 활용하고 있다. 또한, 국제기업들은 기업간 인수 및 합병 또는 업무 제휴 등 다양한 방식으로 종자산업을 이끌어 나가고 있다. 세계 10대 글로벌 종자기업의 시장집중도는 2011년 69.6%에서 2015년 72.0%, 2016년에는 73.1%로 종자시장의 집중화는 더욱 심화되고 있다.

○ 세계 종자 판매액은 2010년 307억 달러에서 연평균 약 7.2%의 증가율로 2014년에는 약 405억 달러까지 확대되었으나 그 이후에는 잠시 성장세가 둔화되어 2016년에는 369억 달러 수준으로 감소하였다. 하지만 2017년에는 성장세가 회복되어 시장규모가 411억 달러 수준까지 증가하였으며, 2018년은 재배면적의 증가로 인해 약 417억 달러 수준으로 성장하였다.

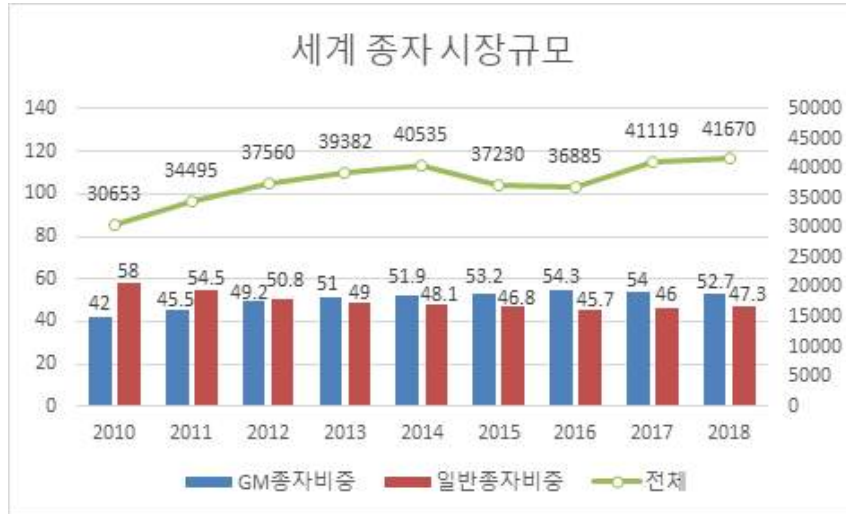


그림2. 세계 종자 시장규모

자료:Phillips McDougall Seed Service 2019

○ 2017년과 2018년에는 글로벌 종자기업간 초대형 인수 및 합병이 이루어지면서 세계종자시장은 새로운 전환기를 맞이하였는데, 글로벌 종자기업의 경쟁구도가 6강 체제에서 4강 체제로 재편된 점이다. 세계 종자시장의 새로운 독과점 체제를 주도할 기업은 ChemChina (중국), Corteva(미국), Bayer(독일), BASF(독일)이다. Chemchina의 경우 2017년 6월에 Syngenta를 430억 달러에 인수하였으며, Corteva의 경우 2015년 12월에 Dow와 DuPont의 합병이 1300억 달러에 이루어졌다. 그리고 Bayer의 경우에는 2018년 6월에 Monsanto를 660억 달러에 인수했으며 BASF의 경우, Bayer CropScience의 자산 및 사 소위 'Big 4' 라고 불리는 글로벌 종자기업은 세계 종자 매출액의 2/3이상을 점유하고 있다. 향후 글로벌 거대 종자기업은 더 많은 시장 접근을 위해 종자 관련 R&D 투자의 새로운 포트폴리오를 구축하거나 기존 종자업체의 유력한 자산매입 또는 인수합병을 통해 규모를 키울것으로 예측된다. 세계종자시장에서 거대종자기업이 가지는 영향력은 상대적으로 시장점유율이 낮은 종자기업에게 큰 부담으로 작용하고 있다. 하지만 이러한 구조변화 속에도 소규모 종자기업들은 value chain에 걸친 유통 효율성 제고, 거래처와의 파트너십 강화를 통한 상호이익증대 등으로 시장에서의 입지를 견고히 하고 있다.

표1. 글로벌 종자기업의 시장점유율 현황(2018년 기준)

| 구분 | 시장점유율 | | | |
|----|---------|---------------------------|------------------------------|---|
| | 20% 이상 | 4~20% 미만 | 1~4% 미만 | 0.6~1% 미만 |
| 유럽 | Bayer | Syngenta, BASF | Vilmorin,KWS,DLF, Rijk Zwaan | Enza Zaden, |
| | | | | Barenbrug,Saaten-Union., RAGT Semences, Florimond Desprez, Union InVivo,Euralis |
| 미국 | Corteva | AgReliant, Genetics | | |
| 일본 | | Sakata,Takii, LPHT,Kaneko | | Advanta,Nuziveedu |

자료:Phillips McDougall Seed Service 2019

○ 우리나라의 경우에는 농림축산식품부 산하 국립종자원을 설립하여 미래 블루오션사업으로 종자산업에 투자하고 있다. 한국 최대 종자기업은 고추 신품종에 강점을 가진 농우바이오가 있으며 2017년 매출액이 1044억원에 이른다. 각종 잎채소와 기능성 채소종자에 특화된 아시아종묘의 경우에는 184억으로 2위의 기업이다. 한국의 전체 종자기업 매출을 다 합쳐도 5408억원(2016년 기준)에 불과하여 영세한 종자회사들이 많이 있는 타입이므로, 이에 대책이 요구되고 있는 실정이다.



그림 3. 국내 종자기업 등록업체 현황

자료 : 국립종자원(2019)

3. 맞춤형 품종개발을 통한 변화된 소비자의 니즈 충족

○ 시대가 변화함에 따라 소비자의 니즈도 많은 방향으로 발전되어 왔다. 기술의 첨단화와 최신화에 따라 과거에는 기아극복에 초점이 맞춰져 이에 따른 육종기술도 수량성에 목표를 두었으나 이제는 맛에 초점이 맞추어져 그러한 목표에 맞춘 작물이 나오게 되었다. 앞으로는 성분과 기능에 초점이 맞춰진 작물연구가 진행되고 또한 건강적 가치에 목표를 둔 새로운 패러다임으로 전환이 진행될 것이다.

○ 변화하는 트렌드인 건강에 초점을 맞춰, 인간게놈분석에 따른 개인별 건강상태에 기반한 맞춤 방식의 농작물 개발 및 생산과 그것을 기반으로 한 식품 및 의약의 제조를 통해 치료의 방식에서, 예방의 방식으로 변화하여 건강한 삶을 살 수 있고 이를 활용하여 융복합적 패러다임으로 진입이 가능하게 개발하는 것을 목표로 한다.

○ 농업의 가치는 인류의 생존을 위한 생산성에 초점을 둔 1세대 농업에서, 맛과 영양 등 품질지향적 시장가치에 초점을 둔 제2세대농업으로 변화하고, 3세대부터는 식품의 안전성과 환경에 중심을 둔 친환경농업으로 발전하고 있고 앞으로는 새로운 방향의 패러다임을 창출할 필요가 있다.

○ 미래의 농업은 이러한 개인별 맞춤형 능동적 먹거리를 제공할 수 있는 농업으로 변화할 것이며 이를 ‘친인간농업’의 개념으로 정의할 수 있다. ‘친인간농업’이란 인간을 위한 농업으로 친환경 개념을 넘어선 차세대 농업으로 정의할 수 있다. 기존의 형성된 시장(농업, 의료, 식품, 의학등)에 신개념의 패러다임을 제공하여 새로운 부가가치 창출이 가능할 것으로 전

망할 수 있다. 이를 위해서는 각 분야간 융합을 통한 피드백이 구성되어 상호간 시너지를 제공하여 이를 통한 부가가치를 창출하여 새로운 시장을 구축하고자 한다.

○ 농업의 종자산업은 국가에서 Golden Seed Project(9년간 약 4000억원 투자)를 통해 많은 발전과 개선을 이끌어왔으며 타 산업과의 연계로 더 많은 고부가가치 이윤을 창출할 것으로 기대된다.

○ 인간게놈의 이해

- 인간게놈 분석 및 빅데이터의 활용(국가적으로 지원예정)한 전인류의 병력 DB화 모델 정립(국가연구소의 인간게놈 연구프로젝트들과 연계하여 모델을 정립함)
- 인간게놈의 유전정보 분석 상품화 (현재 일부 유전체분석관련 회사에서 이미 진행되고 있으나 목적성에 맞추어 공동 연구수행)
- U-health와 게놈의 연계를 위한 기반구축 (삼성 기술진 등과 연계된 공동 시스템 구축)

○ 친인간농업

- 모든 작물의 유전자원수집, 성분분석기술 확립 (1단계에서는 고추, 배추, 무 등 특정 채소에 국한해 시스템 구축 및 실용화 전략 수립) 및 종자개발 (Golden Seed Project와 연계) 수행 모델 정립
- 식물게놈 분석 및 이해를 통한 유전체육종연구
- 종자개발을 약품개발과 연계모델창출 (Lutein, Kaempferol, Querceti, Gluconasturtiin 등 주요 물질 대량생산프로젝트 가능)
- 종자회사와 제약회사의 융합 등을 통한 연계성 및 시스템 구축 검토, 등



그림4. 친인간농업 Flow

자료 : 충남대 임용표 교수

○ 고부가가치 작물의 정밀재배

- 정밀농업기술을 위한 전자, 통신, 기계, 로봇의 융합시스템 구축
- 식물공장의 실용화를 위한 모델 개발
- 작물 재배의 정밀화를 위한 인공위성프로젝트 가능성 탐색



그림5. 4차산업혁명과 농업

자료 : 국립종자원

○ 식품개발 및 영양학적 접근

- 인간 건강에 초점을 맞춘 맞춤형종을 이용한 맞춤형식품의 현실화 모델 확립
- 각 개인의 영양요구성에 기초한 식품개발
- 각종 식품의 영양분 및 식단의 영양정보에 대한 앱 개발을 통한 소비자 구매성 창출

○ 의료기술을 치료에서 예방으로 전환

- 환자식을 포함한 식품을 이용한 처방전 개발 가능성 연구
- 개인 건강 맞춤형 병원, 개인 건강 맞춤 농산품의 판매 시스템 설립 가능성 연구
- 개인별 맞춤형 작물을 이용한 맞춤형 식품제공을 통한 치료 및 예방적 차원의 관리가 가능한 종합병원을 설립하고 하나의 시스템으로 운영할 수 있도록 기반구축 (예: 송도에 국제적인 U-health 종합화 병원의 설립 및 활용)



그림6. 맞춤의학개념변화와 친인간농업의 접목

자료:충남대 임용표 교수

4. SWOT기법을 통한 우리나라 종자산업의 향후 미래 발전 방향 모색

○ 우리나라의 종자산업의 지표를 파악하기 위해서 SWOT분석을 실시하였다. SWOT분석의 가장 큰 장점은 기업의 내부나 외부환경의 변화를 동시에 파악할 수 있다는 것이다. 기업의 내부환경을 분석하여 강점과 약점을 찾아내며, 외부환경 분석을 통해서는 기회와 위협을 찾아낼 수 있다.

표2. 우리나라 작물육종 SWOT 분석

| | 집중요소 | | 보완요소 | |
|-------------|--|---|--|--|
| | 강점(S) | | 약점(W) | |
| 내부역량 | <ul style="list-style-type: none"> - 세계적인 전통육종 기술력 보유 - 풍부한 유전자원 및 효율적 관리체계 - 정부의 종자산업 및 농생명산업 R&D 투자 의지(종자산업진흥센터, 차세대BG21, GSP Project) - 종자산업법(품종보호권) 강화 - 국제경쟁력 있는 채소종자기업 보유 | | <ul style="list-style-type: none"> - 식량 및 주요작물 세계시장 경쟁력 낮음 - 첨단육종 기반 취약 및 사회적 수용성 미흡 - 종자산업 및 농생명산업 기업규모 영세성 및 품종침해사례 다발 - 단기 성과 위주의 종자산업 R&D 투자 관리 - 육종 전문인력 양성 미흡 | |
| | 우선수행(S-O) | 위협전략(S-T) | 우선보완(W-O) | 장기보완(W-T) |
| | <ul style="list-style-type: none"> - 기존 품종들의 결점을 분자육종기술로 보완하여 국제경쟁력 보유한 스타품종 개발 - 종자산업을 수출주도형 산업으로 집중육성 - 종자산업 기반기술발전을 위한 R&D강화 | <ul style="list-style-type: none"> - 세계적인 첨단육종 연구진 및 수출대상국과 국제공동연구개발 추진 - 주요작물 첨단 육종 경쟁력 강화 R&D 투자 확대 - 다국적 기업을 인수하거나 또는 틈새 시장확보 | <ul style="list-style-type: none"> - 농산물 및 종자의 국가경쟁력 향상과 수출을 위한 종자기업의 규모화 - 시장규모가 큰 식량 및 사료작물 육종에 민간 참여 확대 - 품종보호법률강화로 시장질서 확립 | <ul style="list-style-type: none"> - 장기적인 식물분자육종 연구개발 투자와 인프라구축 및 인력 양성 - 국제경쟁력 취약 작물의 육종 R&D 투자 - 첨단육종기술의 수용과 종자산업기반기술 확보를 위한 장기적 R&D 투자 |
| | 기회(O) | | 위협(T) | |
| 외부환경 | <ul style="list-style-type: none"> - UPOV 체제하의 식물품종보호 강화 - FTA/WTO 자유무역환경 조성 - 탄산염 대비 높은 종자산업 및 식물분자육종 관련 산업 성장률 - 국가별/지역별 다양한 신품종 요구도 증가 | | <ul style="list-style-type: none"> - 선진국과 다국적기업의 첨단화된 분자육종 기술발전 및 R&D 투자 확대 - 다국적 기업의 세계 종자시장 점유율 상승 - 다국적 기업의 첨단육종기술 및 소재특점 심화 - 주요 농업국의 종자 산업보호 장벽 강화 | |

자료 : 서울대 고희종 교수

○ 종자산업의 발전을 위해서는 첫째, 육종목표의 다각화를 들 수 있다. 국내시장의 규모는 세계 규모에 비해 월등히 낮을 수 밖에 없으므로, 대상을 해외로 돌려 수출지향적인 품종을 육성해야 하는 것은 어쩌면 당연한 것이 아닐 수 없다. 수출 대상지역에서 요구하는 품종을 명확히 정의하고, 현지 환경조건 또는 기호에 맞는 육종전략을 수립하여 장기적이고 체계적인 플랜이 필요하다.

○ 두 번째로는 육종기술 첨단화가 있다. 우리나라의 벼나 고추, 배추등의 몇몇 작물의 전통육종기술은 세계수준으로 인정받고 있다. 그러거나 생명공학기술을 기반으로 하는 분자육종기술수준은 그에 비해 상당수준 미흡하다. 다국적 종자기업들 및 선진국과 세계시장에서 종자시장 경쟁을 해야 하는 상황에서 첨단 육종기술 분야에 대한 연구개발투자는 시급한 과제이며, 미래의 종자산업 시장경쟁력은 이러한 육종기술과 속도의 관점에서 결정될 것으로 전망된다.

○ 세 번째는 종자회사의 규모화이다. 세계 상업종자시장이 급성장하고, 무역규모 또한 119억달러(ISF,2019) 이상으로 급증하게 되면서 종자는 이제 그 하나로써 중요한 경제적 자원이 되었다. 세계시장에서 수출로 경쟁하기 위해서는 영세규모의 회사로써는 어렵다. 이미 2017년 몬산토의 연 종자 매출액은 109억 달러를 돌파하였으며, 매출 2억달러 이상의 기업

이 20개였다.(Agropages, 2018). 반면에 우리나라 종자회사는 대부분 영세하여 소수의 업체를 제외하고는 수출 판로를 개척하기도 힘든 실정이다. 이에, 종자산업진흥센터나 국립종자원 등에서 관련 지원을 하고 있지만 본격적인 시장경쟁을 위해서는 기업체 자체에서 수출 노하우를 축적하고 체계적으로 영업망을 구축할 수 있어야 한다. 물론, 수출 대상국가에 우수한 품종들을 지속적으로 육성하여 공급할 수 있는 인프라가 구축되어야 한다. 이를 위해서는 종자기업의 규모화는 필수적이며 정부의 종자산업 지원 또한 개인육종가나 영세기업을 지원하는 방법과 중견기업이상의 규모가 있는 회사들을 육성하는 것 또한 필요하다.

○ 네 번째는 민간기업의 글로벌작물 육종강화이다. 국내 상업종자시장은 채소작물 중심으로 편성되어 있다. 그러나 채소종자시장은 물가상승분을 감안하면 수년째 답보상태에 있으며 수출 또한 증가하지 못하고 있다. 세계 종자무역시장이 가파르게 증가하는 것을 감안해 볼 때, 우리나라는 종자수출전략을 재점검해야 할 것이다. 그 중, 가장 먼저 해야 할 것이 작물의 다변화이다. 세계시장에서는 곡물이나 사료작물의 종자시장이 80%이상을 차지하는데 우리는 15% 정도인 채소종자시장에만 집중하고 있어 문제가 있다. 최근 일부 기업에서 옥수수품종 개발과 시장개척에 나서고 있는데 이러한 움직임을 적극적으로 지원해야 할 것이다. 세계시장에서 점유율이 높은 글로벌작물들에 대해 우수한 품종을 육성하여 시장경쟁력을 높일 때 비로소 우리나라 종자산업이 수출주도형 미래산업으로 평가받을 수 있을 것이다.

○ 다섯 번째로는 육종가 및 종자산업전문인력 양성이다. 종자산업은 크게 종자개발 및 연구개발, 종자생산 및 품질관리, 상품화 및 마케팅으로 이루어진다. 그 중에서 종자개발은 육종가가 담당하는 것인데 종자산업의 성패를 결정하는 가장 중요한 분야이며, 이를 위해서는 대체로 고학력의 연구진이 필요하다. 그리고 종자생산 및 품질관리 분야는 품종 및 종자에 대한 이해와 비교적 실무에 숙련된 인력이 필요하다. 상품화 및 마케팅분야에는 종자산업에 특화된 전문 인력이 요구된다. 우리나라는 현재 종자산업을 위해 특성화된 대학 또는 대학원이 없으며, 한시적으로 단기 교육훈련 프로그램을 운영하여 단기적으로 인력을 양성하고 있는 실정이다. 향후에는 연구개발은 물론 세계시장에서 국제적으로 경쟁할 수 있는 전문인력을 양성하는 것은 종자산업 발전을 위한 가장 기본적인 조건이라 할 수 있다.

○ 마지막으로 종자산업에 대한 국가 지원체계가 민간기업 중심으로 강화되어야 한다는 것이다. 정부의 종자산업지원은 바이오그린사업 연구개발(농촌진흥청), 골든시드프로젝트를 통한 품종개발, 민간육종연구단지와 종자산업진흥센터를 통한 기업 서비스 지원, 국립종자원을 통한 개인육종가 지원 및 해외 시범포 설치 등의 사업이 진행중이다. 종자산업은 결국 민간기업이 중심이 되어 세계시장에 진출하며 성장해야 하는 것이기 때문에 이러한 지원사업은 민간기업 중심으로 이루어져야 한다. 특히 제도와 법률 등이 이러한 기업에 도움이 될 수 있도록 다듬어져야 할 것이다. 식량작물의 종자산업에도 민간이 참여할 수 있도록 국립종자원을 통한 식량종자 공급시 종자가격을 일정 부분 현실화해야 하며, 가급적 종자증식 및 공급사업도 일부를 민간이 담당할 수 있도록 제도적 장치를 마련해야 한다. 특히 연구개발 사업의 경우 단기기간의 성과에 급급해 할 것이 아니라 장기적인 안목으로 투자해야 세계적으로 우수한 품종들이 육성될 수 있음을 명심해야 한다.

제10절 국내 종자산업 현황과 국제경쟁력 제고 방안 연구

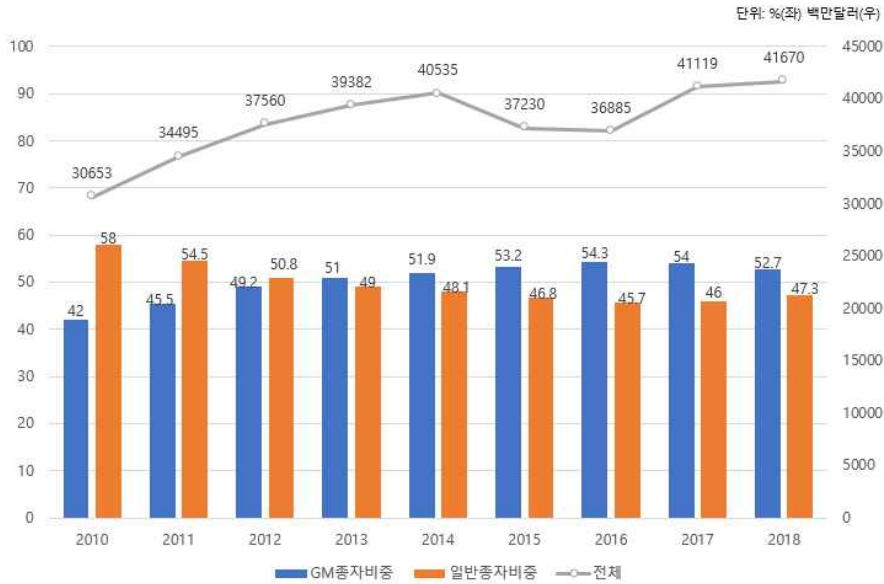
1. 해외 종자산업 현황과 분석

가. 해외종자산업 현황

(1) 세계 종자시장 현황

○ 세계 종자 판매액은 지속적으로 증가하고 있다. 2016년을 제외하고 2010년 307억 달러, 2014년 405억 달러, 2017년 411억 달러, 2018년은 417억 달러로 증가하였다 (그림 1).

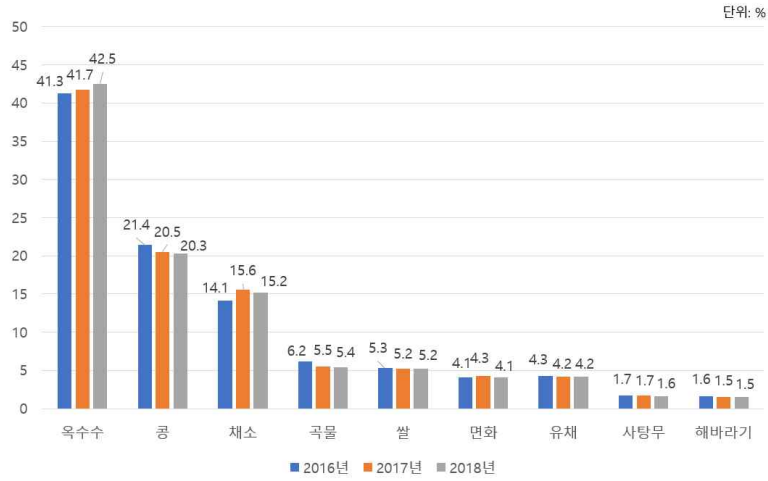
그림01. 세계 종자시장 규모



출처: Phillips McDougall Seed Service 2019

○ 세계 3대 주요 종자 작목인 콩, 옥수수, 채소류의 매출은 전체 종자 매출의 2018년 기준 78%에 달한다. 이는 324억 8,300만 달러에 이르는 금액이다. 그중 옥수수의 종자 판매액이 전체 약 42.5%로 가장 큰 비주를 차지하며 약 177억 100만 달러이며, 다음으로 콩 종자가 20.3%로 약 84억 4,300만 달러이고 채소 종자가 15.2%로 약 63억 3,900만 달러 수준이다 (그림02).

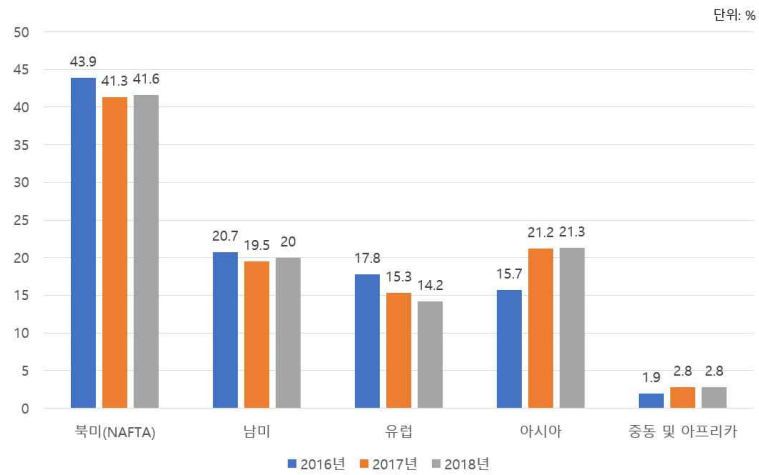
그림02. 작목별 세계 종자시장 규모 비율



출처: Phillips McDougall Seed Service 2019

○ 권역별 종자 시장규모를 보면, 아시아 종자 시장은 지속적인 성장세를 보이는 반면, 유럽은 2016년 이후로 지속적으로 판매액이 감소하고 있다. 북미(NAFTA)와 남미지역의 종자 시장규모는 전년보다 소폭 확대되었다(그림03).

그림03. 권역별 세계 종자시장 규모 비율



출처: Phillips McDougall Seed Service 2019

(2) 세계 종자시장 점유율 분석

○ 2017년과 2018년에 대형종자기업간의 인수 및 합병이 이루어지면서, ChemChina(중국), Bayer(독일), Corteva(미국), BASF(독일)의 4개의 글로벌 종자 기업이 세계 종자시장을 주도하게 되었다(그림4).

○ 2015년 12월에는 Dow와 Dupont의 합병이 1,300억 달러에 이루어졌으며, 2017년 6월 ChemChina의 Syngenta 인수가 430억 달러에 이루어졌다. 또한 2018년 Bayer의 Monsanto 인수는 660억 달러에, BASF의 Bayer CropScience를 인수하였다. 이러한 초대형 기업의 M&M는 세계 종자시장 4강 체제를 만들었다. 이러한 'Big4'로 불리는 글로벌 종자기업은 세계 종자 매출액의 2/3 이상을 점유하고 있다. 위 기업은 지속적인 시장규모 우세 경쟁을 위해 종자 관련 R&D 투자 지속하고 있으며, 종자업체의 인수 및 합병을 통해 주도력을 넓힐 것으로 예상된다. 이로 인해 국내 종자 기업의 글로벌 종자기업으로 성장하는데 진입장벽을 더욱 높이는 계기가 되고 있다.

그림04. 세계 종자시장의 인수 및 합병



○ 상위 10개 글로벌 종자기업의 매출액은 2018년 279억 100만 달러로 전체 416억 7,000만 달러의 67%이다. Bayer로 2018년 103억 3,600만 달러로 글로벌 종자기업 중에서 종자 판매액이 가장 많은 기업으로 세계 종자 매출액의 24%을 차지하고 있다. Corteva는 두 번째로 판매액이 많은 기업으로 종자 매출액 비중이 18.8% (8억 4,200만 달러)로 집계된다. 세 번째로 판매액이 많은 기업은 Syngenta로 30억 400만 달러로 전체의 7.2%를 차지하고 있다. 이외에 BASF가 15억 3,500만 달러로 3.7%이며, Vilmorin은 14억 9,400만 달러로 3.6%, KWS는 13억 3,900만 달러로 3.2% 등의 순으로 종자 판매액을 기록하였으며 매출액을 통해 상위 10개 기업의 비중을 파악해 보았을 때 시장점유율은 약 67%에 이르며 이는 곧 과점(Oligopoly)의 양상으로 볼 수 있다(그림 05).

그림05. 주요 글로벌 종자기업의 종자매출액 점유율

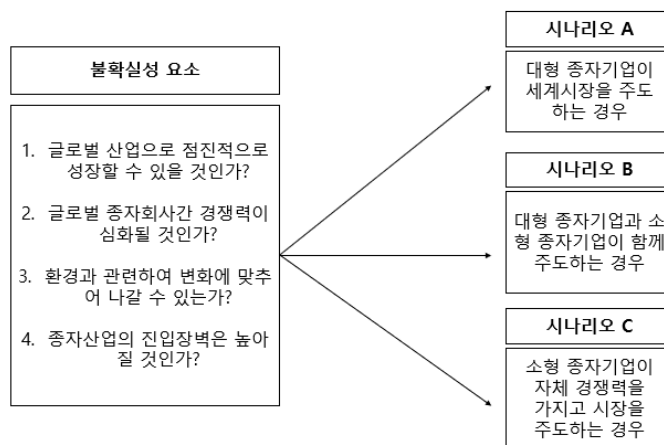


출처: Phillips McDougall Seed Service 2019

(3)세계종자기업 시나리오 분석(Scenario Planning)

○ 세계종자기업의 변화의 요소와 불확실성 요인을 조합하여 발생 가능한 여러 가지 미래의 모습을 설정하여 시나리오 분석(Scenario Planning)를 진행하였다(그림 6). 글로벌 산업으로 점진적인 성장성, 종자회사간 경쟁력 심화, 환경 변화에 따른 지속개발 가능성, 종자산업의 진입장벽을 불확실성 요소로 정하여 3개의 시나리오(A,B,C)를 도출하였다. 시나리오 A는 대형 종자기업이 세계시장을 주도하는 경우, 시나리오 B는 대형 종자기업과 소형 종자기업이 함께 주도하는 경우, 시나리오 C는 소형 종자기업이 자체 경쟁력을 가지고 시장을 주도하는 경우이다.

그림06. 글로벌 종자산업의 시나리오 도출



출처: P. Schoemaker, "Scenario Planning : A Tool for Strategic Thinking"

○ 대형 종자기업의 5년간 인수 및 합병 건수는 총 127건으로 평균 연간 21건이다. 이는 대형 종자 기업의 경쟁 구도는 8강에서 4강으로 좁혀지는 양상을 보인다. 이는 차후 3강 또는 독점으로 좁혀질 것으로 분석할 수 있다(표 1).

표01. 글로벌 종자기업의 시장점유율 현황('18년)

| 구분 | 시장점유율 | | | |
|----|---------|----------------|--------------------------------|--|
| | 20%이상 | 4~20%미만 | 1~4%미만 | 0.6~1% 미만 |
| 유럽 | Bayer | Syngenta, BASF | Vilmorin, KWS, DLF, Rijk Zwaan | Enza Zaden, Barenbrug, Saatan-Union, RAGT Semences, Florimond Desprez, Union InVivo, Euralis |
| 미국 | Corteva | | AgReliant Genetics | |
| 일본 | | | Sakata, Takii, LPHT, Kaneko | Advanta, Nuziveedu |

출처: Phillips McDougall Seed Service2019

○ 다음은 대형 종자기업과 소형 종자기업이 함께 주도하는 경우이다. 종자 시장에서 활동하는 기업들은 크게 에너지, 화학, 제약 등으로 구분되며, Bayer, BASF기업을 대표로 농약, 비료, 농자재, 종자로 확대하고 있다. 농약과 비료 등 농자재 기반 글로벌 기업으로는 Syngenta, Corteva, Vilmorin, KWS, UPL, Nufarm 등이 있다. 위 기업들은 GM작물을 기반으로 세계 곡물 시장을 주도하고 있다. 마지막으로 종자개발부터 순수 종자기업으로 지역이나 틈새시장을 기반으로 DLF, Sakata, Takii, LPHT 등이 있다.

2. 국내 종자산업의 국제경쟁력 결정요인

가. 국내 종자산업 개요

○ 우리나라는 '97년 11월 IMF(International Monetary Fund, 국제통화기금) 구제금융 사태 이후 대부분의 종자주권을 상실하였다. 국내 상위 5대 종자회사 중 4개 회사가 외국기업에 인수되었다. 이로 인해 중소, 개인 육종가가 늘어나면서 국내에 소규모 업체 비중이 증가하였다.

○ '국제신품종보호동맹(The International Union for the Protection of New Varieties of Plants, UPOV) 가입(2002년)과 '종자산업법' 개정, '식물신품종보호법' 개정, '지식재산기본법' 개정을 통하여 개인육종가, 소규모 종자산업에 대한 지원으로 품종보호출원, 해외채종 국내전환등을 지원하였다. 이후, 이후로 국내 종자산업 발전을 위해 골든시드프로젝트

(Golden Seed Project, GSP), 채소육종연구센터, 민간육종연구단지 등 대규모 집중 투자를 진행하였다.

나. 국내 종자시장 현황과 수출입 동향

○ 국내 종자기업의 수는 2020년 기준 3,315개(채소:15.9%, 과수 26.2%, 화훼 12.6%)이며 재배품목과 품목별 품종의 다양화와 정부의 세제지원 혜택 등으로 매년 지속적으로 증가하고 있다. 이로 인해 내수시장 성장의 과포화 상태의 문제를 겪고 있다.

○ 소규모 업체(5억 미만)의 수는 88%로 지속적으로 증가하고 있다. 기업의 수는 지속적으로 증가하고 있으나, 품종 육성부터 가공 및 유통 판매까지 이루는 기술력과 경쟁력을 가진 업체는 많지 않다(표2).

표02. 종자업체의 종자판매액 규모별 분포

| | 소규모 (5억 미만) | 중소규모 (5억 이상~15억 미만) | 중규모 (15억이상~40억 미만) | 대규모 (40억 이상) | 계 |
|------|----------------|---------------------------|--------------------------|-----------------|-------------|
| 2015 | 1,061(87.9%) | 88(7.3%) | 41(3.4%) | 17(1.4%) | 1,207(100%) |
| 2016 | 1,175(87.9%) | 97(7.3%) | 46(3.4%) | 19(1.4%) | 1,337(100%) |
| 2017 | 1,165(88.6%) | 77(5.9%) | 44(3.3%) | 29(2.2%) | 1,315(100%) |

출처: 국립종자원

○ 국내 민간 종자업체의 R&D 투입액은 '17년도 기준 약 821억 원으로 자체연구개발비 비중은 매년 커지는 추세에 있다. 매출액 10억원 미만의 기업은 약 39%이며, 51%의 기업이 종업원이 10명 미만이다. 8%의 기업만이 100명이 넘는 인원을 가지고 있을 뿐이다. 종자산업의 전 영역(품종 개발, 가공, 유통)을 아우르는 회사로는 농우바이오, 코레곤, 삼성종묘, 팜한농, 아시아종묘 등 7개사에 불과하다.

표03. 종자업체 품종개발 R&D 투입액('17년)

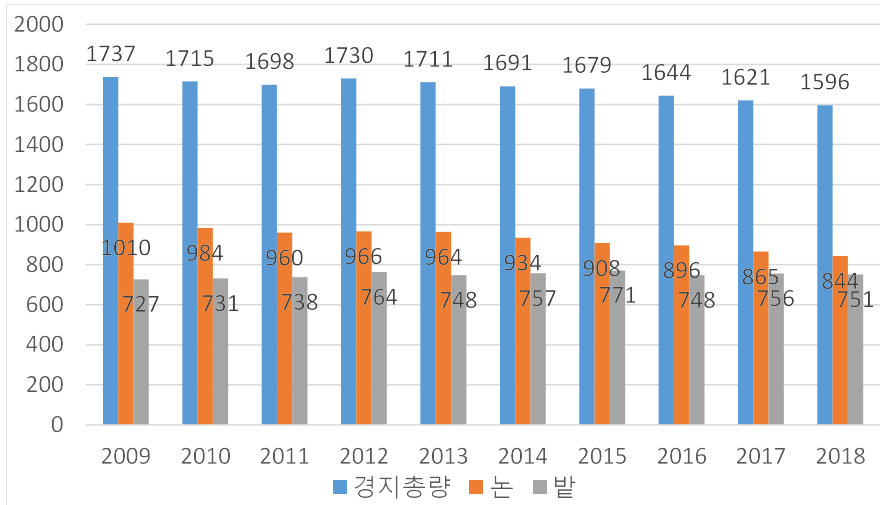
단위: 만원

| 구분 | 자체 연구개발비 | 위탁 연구개발 비 | 수탁 연구개발 비 | 계 | |
|------------|------------------|-----------------|-----------------|-----------------|-----------|
| 전체 | 7,109,942(86.6%) | 283,250(3.5%) | 814,500(9.9%) | 8,207,692(100%) | |
| 판매액 규모별 | 소 | 1,115,544 | 94,050 | 184,500 | 1,394,094 |
| | 중소 | 1,080,500 | 51,200 | 105,000 | 1,236,700 |
| | 중 | 901,500 | 98,000 | 33,000 | 1,032,500 |
| | 대 | 4,012,398 | 40,000 | 492,000 | 4,544,398 |

출처: 국립종자원 '2017년 종자업실태조사'

그림7. 연도별 경지면적 추이

단위 :ha

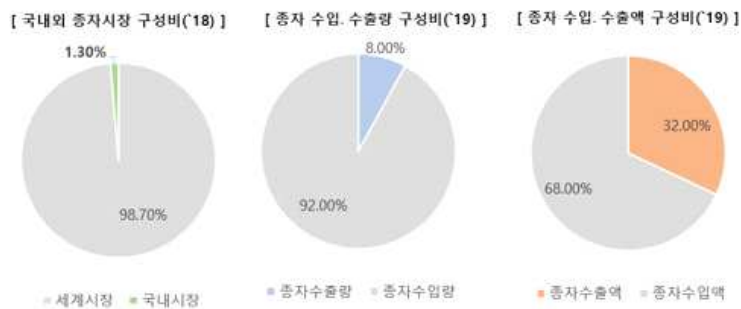


출처: 통계청 '2018년 경지면적조사 결과

○ 우리나라 종자시장 규모는 1.3% 수준으로 '18년 기준으로 약 9,662억 원이며, 채소종자 시장의 경우 '11년 약1,977억원에서 '18년에는 약 2,369억원으로 25%를 차지하고 있다. 국내 종자 수출량은 전 년도 1,488톤에서 1,408톤으로 80톤 감소한것에 반해 수출액은 5,852만 달러로 전년도의 약5,229만 달러보다 약 622만 달러로 11.9% 증가했다. 수입량으로는 약 16,662톤으로 전년도의 17,858톤에 비해 1,196톤으로 6.7% 감소했다. 종자 수입액은 약 1억 2,624만 달러이며 이는 전년도 1억 2,674만 달러보다 0.4% 감소하였다.

○ 수입과 수출을 비교하면 종자 수입량은 수출량보다 11.8배가 많고 (약 15,254톤), 수입액은 수출액보다 2.2배(약 6,772만 달러) 많다(그림 8). 세계 시장에서 우리나라가 차지하는 비중은 약 1.3%이다.

그림8. 종자시장의 수출입 구성비



출처: 국립종자원

표04. 종자 수출입 현황('17년~'19년)

| 2019년 | | | | |
|-------|--------------|--------------|---------------|---------------|
| 품 목 | 수 출 현 황 | | 수 입 현 황 | |
| | 물량(kg) | 금액(USD) | 물량(kg) | 금액(USD) |
| ❖❖합계 | 1,408,327.90 | 58,520,490.0 | 16,662,444.30 | 126,243,213.0 |
| ❖식량작물 | 766,613.20 | 433,332.0 | 1,249,143.40 | 6,772,442.0 |
| ❖사료작물 | 953.00 | 5,947.0 | 11,939,573.70 | 29,094,979.0 |
| ❖화훼작물 | 30.80 | 4,320.0 | 233,985.60 | 5,704,272.0 |
| ❖채소작물 | 633,730.00 | 57,015,034.0 | 2,063,584.20 | 77,041,307.0 |
| ❖과수작물 | 30.00 | 86.0 | 72,914.40 | 175,417.0 |
| 2018년 | | | | |
| 품 목 | 수 출 현 황 | | 수 입 현 황 | |
| | 물량(kg) | 금액(USD) | 물량(kg) | 금액(USD) |
| ❖❖합계 | 1,487,901.30 | 52,297,379.0 | 17,857,704.60 | 126,745,741.0 |
| ❖식량작물 | 913,638.50 | 508,324.00 | 2,286,280.00 | 8,368,577.00 |
| ❖사료작물 | 106.00 | 96.0 | 11,904,869.50 | 29,149,200.00 |
| ❖화훼작물 | 61.90 | 11,321.0 | 258,721.90 | 6,017,710.0 |
| ❖채소작물 | 568,569.90 | 50,766,129.0 | 2,090,956.60 | 75,834,380.0 |
| ❖과수작물 | 62.20 | 3,778.0 | 60,237.70 | 235,307.0 |
| 2017년 | | | | |
| 품 목 | 수 출 현 황 | | 수 입 현 황 | |
| | 물량(kg) | 금액(USD) | 물량(kg) | 금액(USD) |
| ❖❖합계 | 1,335,883.20 | 58,542,052 | 16,243,178.60 | 115,891,793 |
| ❖식량작물 | 717,025.30 | 526,510 | 1,963,085.60 | 6,651,801 |
| ❖사료작물 | 1,144.50 | 1,791 | 10,660,826.90 | 24,769,246 |
| ❖화훼작물 | 203.80 | 85,646 | 311,360.90 | 6,379,007 |
| ❖채소작물 | 615,882.00 | 57,772,641 | 2,172,435.40 | 73,479,108 |
| ❖과수작물 | 51.20 | 7,102 | 80,820.20 | 360,305 |

3. 국내 종자산업의 국제경쟁력 강화 방법

2020년에 국립 종자원에 등록된 국내 종자기업의 수는 3천여 개이며, 매년 지속적으로 증가하고 있다. 증가한 주요 요인으로는 품목별 품종의 다양화와 정부의 세제지원 혜택등으로 민간 종자기업의 참여 비중이 증가하였으며, 다국적 기업의 인수 합병 이후 양산된 업체의 수의 증가가 있다. 채소종자업체의 수는 480개로 전체의 약 15.9%를 차지한다. 국내 기업의 수 증가로 인해 기업 간 가격 경쟁심화가 일어나게 되어 내수시장 성장의 한계점이 생기게 되었다. 또한 재배면적 감소와 육묘산업으로 인한 종자 수요량 감소 그리고 수입농산물의 증가로 중요 요인 중 하나이다.

국내 종자업체의 규모별 분포는 <표02>와 같이, 5억 미만의 소규모 업체의 수(88% 이상)는 지속적으로 증가하고 있다. 종자관련 기업의 수는 지속적으로 증가하고 있으나 영세한 소규모 업체가 많고 육성부터 가공처리 및 유통판매까지의 모든 과정에 기술력과 경쟁력을 겸비한 전문 업체는 많지 않다. 대규모 기업의 경우에도 글로벌 종자기업과 비교할 때 규모가 작으며 연구개발 투자예산 또한 상당히 적어 국제 경쟁력도 약한 입장이다.

국내 민간 종자업체의 R&D 투입액<표04>은 '17년도 기준 약 821억 원으로 R&D투자액 중 자체연구개발비 비중은 매년 커지는 반면에 수탁연구개발비의 비중은 감소하는 추세에 있다. 한국종자협회 가입업체의 매출액 10억 원 미만의 기업(약 39%)들로 종업원 10명 미만의 기업체 수가 26개사(51%)이며 100명이 넘는 기업체는 4개사(8%)이다. 농우바이오, 코레콘, 삼성종묘, 팜한농, 아시아종묘 등 7개사만이 품종 육성부터 가공처리 및 유통까지 기술력과 경쟁력을 갖춘 업체이다. 이들 기업들조차 글로벌 종자기업들에 비해 매출 규모뿐만 아니라 연구개발비용도 적어 경쟁력이 미흡한 상황에 있다.

또한, 연도별 경지면적 추이<그림07>를 보면 2009년은 논 1,010ha, 밭 727ha 규모에서 점차적으로 감소하여 2018년에는 논 844ha, 밭 751ha 수준까지 감소한 것으로 나타났다.

국제경쟁력 강화 요인은 임금, 소비자물가지수, 수출물가지수가 종자 수출에 가장 중요한 요인이나, 국가마다 상이한 부분이 있다. R&D, 생산량은 국제경쟁력 강화 요인에 미치는 영향이 국가마다 상이한 차이가 있어 추가적인 연구가 필요하다. 자료의 제한으로 인해 다른 요인을 고려할 경우 연구 확대를 통한 통합적 연구 확대 또한 요구된다.

고추, 배추, 무의 경우 이미 세계적인 위치에 있다. 현재 아시아권 채소시장은 성장세에 있기 때문에 주력 품조를 대상으로 아시아권에 활약하는 것이 중요하다. 이를 위해 약 3천개의 영세한 종자업체의 규모화가 필요하며, 인수 및 합병 또는 대기업이 진출해야 한다. 화학, 에너지, 제약산업을 기반으로 성장한 글로벌 기업이 현재 종자산업을 주도하고 있다. 생명공학과 같은 첨단 기술을 적용하기 위해서는 융복합 산업이 필수이며, 이는 영세 농가에서는 접근하기 불가능하다. 이를 극복하기 위해 종자회사의 규모화는 필수적이다. 이를 위해서는 정부의 지원과 역할이 중요하다. 농림축산식품부는 총 9개 분야(농촌, 농업, 식량, 국제, 축산, 농생명, 유통, 식품, 방역)로 나누어 정책을 시행하고 있다. 1차와 2차 종자유성 5개년 계획을 통해 국내종자산업의 해외진출을 목표로 지원을 강화하였다. 종자산업의 특성상 장기적인 연구개발이 전제되어야 한다. 원활한 시행을 위해서는 행정기관의 컨트롤 타워 역할을 통해 생산부터 상품화 그리고 마케팅의 전과정이 연계되어 이루어져야 한다. 우리나라가 종자 선진국으로 발돋움하기 위해서는 국가간 협력과 해외시장 플랫폼 구축을 통해 민간기업의 해외 진출 진입장벽을 낮추어야 하며 이는 기업과 정부 그리고 연구기관 간의 지속적인 협조와 참여가 형성되어야 한다.

제 3 장 목표 달성도 및 관련 분야 기여도

3-1. 목표

- 채소작물의 종자개발을 위한 육종 특화 통합 DB 구축 및 운영
- 유전체 정보 통합 DB 구축으로 종자 개발에 있어 분자유종의 효율화를 지원
- 각 프로젝트별(채소별)로 생산되는 통합해 제공함으로써 프로젝트간 시너지 효과 유발
- 종자 개발에 필요한 유전체, 유전분석을 수행하고 이를 통해 목표 형질 개체선발 및 계통선발, 목표 형질 접근 방법에 효율화

3-2. 목표 달성여부

1. 기술적 측면

- 육종가의 경험적인 역량으로 선발과 교배가 이루어져 왔던 전통육종에 NGS(차세대 염기서열분석)를 통해 발굴된 대량 분자마커를 활용하여 현대 육종의 진화를 도모함.
- 분자유종에 전문화된 데이터베이스를 구축하여 작물 육종에서 필요로 하는 형질에 관련된 유전자의 분석된 정보를 제공함으로써, 분자마커 개발에 소요되는 시간을 단축하여 효율적이고 차원 높은 분자유종 시스템을 구축하는데 기여할 수 있음.
- 육종 기술에 대한 사용자 중심의 데이터베이스 및 분석 지원 시스템을 통하여 검증된 분자마커의 정보를 국내 육종공동체로의 보급하여 신품종 육종과정에서 체계적으로 활용이 가능함.
- 5대 작물에 대한 분자마커 육종지원시스템 개발을 통해 분자마커를 이용하는 육종기술의 체계화 및 타 작물에 대한 유사 시스템 기술지원이 가능함.
- 작물 육종 활동에 대한 GWAS 분석 방식의 접목은 다양한 표현형 중 특정 형질에 대한 신규 분자마커의 개발을 집단 작성 및 전개에 소요되는 시간과 비용을 크게 절감할 수 있음.

2. 경제적·산업적 측면

- 복잡 형질의 유전 현상을 시스템적으로 이해하고 예측 가능한 육종 계획을 세움으로 육종 기간을 단축하고 맞춤형육종(breeding-by-design)을 구현할 수 있는 기반을 제공함.
- 공통된 작물을 육종하는 여러 육종가에게 동일한 분자마커의 사용을 유도함으로써 육종 연한의 감소를 이룩함.
- 다양한 작물 종을 대상으로 연구하는 분자유종가에게 공통의 분자마커 사용을 촉진함으로써, 신품종 육성 가속화로 인한 육종 기간 단축과 분자마커 개발 및 파생 기술을 이용한 산업화를 통하여 다가오는 바이오경제의 경쟁력을 확보할 수 있을 뿐만 아니라 관련 민간기업의 활성화에 기여할 수 있음.
- GWAS 분석 시스템의 구축은 관심 형질에 대한 candidate gene 상의 유전자SNP를 신속하게 분자마커화 할 수 있기 때문에 이를 활용한 육종과 신품종의 생산 및 시장 진출을 유도하는 기반 연구의 초석을 마련함.
- MAB와 MAS를 통해 신품종 육성에 대한 최적의 교배 조합과 필요 개체수의 예측으로 육종 효율성을 극대화시켜 우리 종자의 세계시장에서의 경쟁력을 확보하고 종자 수출량의 증대를 기대.
- 시장 동향 파악을 통해 빠르게 변화하는 세계 종자시장의 수요에 대한 맞춤형 품종 개발의 효율성을 극대화 할 것으로 기대.

3-3. 목표 미달성 시 원인(사유) 및 차후대책(후속연구의 필요성 등)

제 4 장 연구결과의 활용 계획 등

○ 채소작물의 육종에 활용할 수 있는 성과로서 29편의 논문의 게재(SCI급 28편, 비SCI 1편)와 12건의 특허 실적(등록: 7건, 출원: 5건)을 달성하였다. 해당 성과들은 과제 수행을 통해 구축한 데이터베이스를 통해 연구자 및 육종가들의 접근이 가능하도록 서비스하여 국내의 채소작물의 분자유종의 활성화를 실현할 수 있도록 할 것이다. 또한 과제 수행 중의 연구에서 개발한 분자마커 및 관련정보들을 분자유종 기술이 부재한 종자회사에게 마커 검정 서비스를 제공할 수 있을 것으로 기대된다.

가. 육종 특화 데이터베이스 구축을 위한 기반 정보 수집 및 업데이트

채소작물의 종자개발을 위한 육종 특화 통합 데이터베이스는 5대 작물에 대한 기본 플랫폼이 완성되었다. 수박 분자 육종 활성화를 위해 데이터베이스 구축 및 업데이트로는 수박의 유전자원 염기서열 수집과 SNP 데이터 구성과 전사체 데이터 분석을 통한 옹성불일 카과 개발이 있다. 고추의 경우 SRA 데이터 수집, 유전체 데이터 확보 및 데이터화 유전자 주석, 고추 RIL 집단의 genotype 데이터의 구성, 변이 분석과 같은 유전데이터를 확보하였다. 무의 경우 무의 서열 정보 확보와 6개의 계통의 변이 데이터 생성 그리고 무의 유용형질에 대한 유전분석을 통하여 무의 육종 특화 데이터 베이스를 구축하였다. 이와 같은 5대작물의 데이터 플랫폼 업데이트는 국내의 채소 분자유종 공동체에 각 작물에 대한 육종 관련 정보를 통합적으로 제공하여 신규 마커의 개발과 신품종 육성의 촉진을 기대할 수 있다.

나. 신규 분자 마커 개발을 위한 유전연구

효율적인 분자 마커 개발을 위한 전사체 정보 처리 방식을 개선 및 분석 파이프라인을 구축하였다. 작물의 유용한 특성(노화, 뿌리혹병, 잎털, 결각)에 대한 유전 분석(GWAS, QTL)을 진행하였다. 여기서 개발된 표현형 분석 방법은 육종가들의 균일한 표현형 평가 방법에 사용될 수 있을 것이며, 분석된 유전형질 또는 분자 마커는 육종가들의 형질관련 기초자료와 계통선발에 사용될 것이다. 또한 전사체 분석을 통한 네트워크 분석은 식물의 생리 및 유전형질에 대한 좀 더 깊은 이해를 하는데 도움을 줄 것이다.

다. 채소 작물의 유용 형질관련 분자마커의 범용화 및 분자마커의 high-throughput selection을 위한 마커 이전 서비스

기보고된 유용 표현형관련 분자마커가 지정하는 유전자의 위치는 현재 공개된 배추의 표준 유전체 정보를 통해 쉽게 접근이 가능하다. 하지만 수집한 유전체 정보를 육종가가 사용 및 제공하기 어렵기 때문에 이것을 가공 또는 효율적인 변이 분석 방식이 필요하다. 배추의 표준 유전체 정보를 지속적인 업데이트와 SNP 생산 matrix의 개선 및 SNP 산출 방식의 개선을 진행하였다. 이러한 데이터를 수집한 분자마커의 세부사항을 확인하였다. 개발된 다양한 마커들을 실제 육종 소재에 적용함으로써 범용성을 확인 하였다. 개발 및 검증이 종료된 범용마커는 육종공동체에 보급되어 신속한 목표 형질의 선발 및 신품종 육성에 기여할 수 있다.

라. 종자산업 발전 및 마케팅 전략 구축을 위한 기반 연구

국내외 종자 산업의 현황과 문제점을 파악하였으며, 이를 해결하기 위한 다양한 방향들을 제시하였다. 국제협력사업 관계 기관들과의 협력계 구축을 통한 채소 종자 수출 활성화 기반 조성 및 종자 산업 발전 및 마케팅 전략 구축을 위한 기반 연구로 맞춤형 품종개발을 제시하였다. 세계 종자시장기업과 국제 정제를 기반으로 서술된 시나리오들은 국내 종자기업의 국제 경쟁력 제고 방안에 기초 자료로 사용될 것이다.

참고문헌

<보고서 및 해외 문헌>

- 국립종자원. 2017. 「2017년도 업무계획」.
- 국립종자원. “종자업 실태 1차년도 조사 결과.” 2017. 4. 12. 보도자료.
- 김설중. 2014. “우리나라 채소종자산업의 수출확대 방안.” 경북대학교 대학원 석사학위논문.
- 김홍수. 2014. “파프리카 산업의 생산 및 수출활성화 방안에 관한 연구.” 경상대학교 대학원 석사학위 논문.
- 농림축산식품부. 2017. 「2016 종자업 실태조사(2015년 기준)」.
- 농림축산식품부. 2017. 「2017년도 농림축산식품사업안내서」.
- 한국바이오협회. 2012. 「한국바이오협회 30년의 발자취」.
- 한국종자협회. 2017. 「종자소식」. 2017-1호.
- EU. 2013. 「The EU seed and plant reproductive material market in perspective : a focus on companies and market shares」.
- EU. 2015. 「Policy Department Structural and Cogesion Policies-Overview of the Agricultural Inputs Sector in the EU」.
- Monsanto. 2017. 「2016 Annual report : A Limitless perspective」.
- Syngenta. 2016. 「Annual Review 2015」.
- Vilmorin & Cie. 2017. 「Vilmorin & Cie Annual Report 2015-2016」.

<홈페이지>

- 국립종자원 홈페이지 (<http://www.seed.go.kr>, 접속일 : 2017. 02. 27)
- 한국건강기능식품협회 홈페이지(<http://www.hfood.or.kr>, 접속일 2017. 09. 12.)
- 한국과수종묘협회 홈페이지(<http://www.kkja.org>, 접속일 : 2017. 08. 01)
- 한국바이오회의약품협회 홈페이지(<http://www.kobia.kr>, 접속일 2017. 09. 12.).
- 한국육묘산업연합회 홈페이지(<http://www.hkyugmyo.or.kr>, 접속일 2017. 08. 01)
- 한국종자연구회 홈페이지(<http://www.kosid.net>, 접속일 2017. 08. 01)
- 한국제약바이오협회 홈페이지(<http://www.kpbma.or.kr>, 접속일 2017. 09. 12.)
- 한국종자포럼 홈페이지(<http://www.koseed.kr>, 접속일 2017. 08. 01)
- 한국종자협회 홈페이지 (<http://kosaseed.or.kr>, 접속일 2017. 03. 13)

<논문>

1. 고희중. (2020). 우리나라 작물육종 성과와 발전 방안. Korean Journal of Breeding Science, 52
2. 이기섭. (2019). 다국적 거대 종자기업이 국내 종자산업에 미치는 영향. 한국국제농업개발학회지, 31(1), 82-89
- 3.Cai C, Wang X, Liu B, Wu J, Liang J, Cui Y, et al. *Brassica rapa* Genome 2.0: A Reference Upgrade through Sequence Re-assembly and Gene Re-annotation. Mol Plant. 2017;10(4):649-51.
- 4.Choi SR, Oh SH, Chhapekar SS, Dhandapani V, Lee CY, Rameneni JJ, Ma Y, Choi GJ, Lee SS, Lim YP (2020a) Quantitative trait locus mapping of clubroot resistance and Plasmodiophora brassicae pathotype Banglim-specific marker development in Brassica rapa. Int J Mol Sci 21(11):4157

5. Choi SR, Oh SH, Dhandapani V, Jang CS, Ahn CH, Rameneni JJ, Kim H, Jeon I, Lim YP (2020b) Development of SNP markers for marker-assisted breeding in Chinese cabbage using Fluidigm genotyping assays. *Hortic Environ Biotechnol* 61:327–338
6. Chu M, Song T, Falk KC, Zhang X, Liu X, Chang A, Lahlali R, McGregor L, Gossen BD, Yu F, Peng G (2014) Fine mapping of *Rcr1* and analyses of its effect on transcriptome patterns during infection by *Plasmodiophora brassicae*. *BMC Genomics* 15:1166
7. Engelke T, Hirsche J, Roitsch T. Anther-specific carbohydrate supply and restoration of metabolically engineered male sterility. *J Exp Bot.* 2010;61(10):2693–706.
8. Guo S, Zhang J, Sun H, Salse J, Lucas WJ, Zhang H, et al. The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. *Nat Genet.* 2013;45(1):51–8.
9. Hatakeyama K, Suwabe K, Tomita RN, Kato T, Nunome T, Fukuoka H, Matsumoto S (2013) Identification and characterization of *Crr1a*, a gene for resistance to clubroot disease (*Plasmodiophora brassicae* Woronin) in *Brassica rapa* L. *PLoS One* 8:e54745
10. Karim M, Dakouri A, Zhang Y, Chen Q, Peng G, Strelkov SE, Gossen BD, Yu F (2020) Two clubroot-resistance genes, *Rcr3* and *Rcr9wa*, mapped in *Brassica rapa* using bulk segregant RNA sequencing. *Int J Mol Sci* 21:5033
11. Kim MS, Hong S, Devaraj SP, Im S, Kim JR, Lim YP. Identification and characterization of the leaf specific networks of inner and rosette leaves in *Brassica rapa*. *Biochem Biophys Res Commun.* 2017;490(3):821–6.
12. Li H, Chang J, Zheng J, Dong Y, Liu Q, Yang X, et al. Local melatonin application induces cold tolerance in distant organs of *Citrullus lanatus* L. via long distance transport. *Sci Rep.* 2017;7:40858.
13. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics.* 2009;25(14):1754–60.
14. Rhee SJ, Seo M, Jang YJ, Cho S, Lee GP. Transcriptome profiling of differentially expressed genes in floral buds and flowers of male sterile and fertile lines in watermelon. *BMC Genomics.* 2015;16:914.

연구개발보고서 초록

| | | | | | |
|--|---|------------|------------|-----------------------|-----|
| 프로젝트명 | (국문)채소작물의 종자개발을 위한 육종 특화 통합 DB 구축 및 운영 | | | | |
| | (영문)Development and Establishment of Database System for Vegetable Breeding | | | | |
| 프로젝트 연구기관 | 충남대학교 | 프로젝트연구 책임자 | (소속)충남대학교 | | |
| 참여 기업 | | | (성명)임용표 | | |
| 총연구개발비 (5,903,000천원) | 계 | 5,903,000 | 총 연구 기간 | 2017.01.~2021.12.(5년) | |
| | 정부출연 연구개발비 | 5,153,000 | 총 참여 연구원 수 | 총 인원 | 202 |
| | 기업부담금 | | | 내부인원 | 197 |
| | 연구기관부담금 | | | 외부인원 | 5 |
| | 정부외출연금 | 750,000 | | | |
| <p>○ 연구개발 목표 및 성과</p> <ul style="list-style-type: none"> - 유전체 정보 통합 DB를 구축하여 종자 개발에 있어 분자유종의 효율화를 지원함. - 각 프로젝트에서 생산된 정보기지를 구축함으로써 연계성 및 시너지 효과 유도 - 종자 개발에 있어 목표 형질 개체선발 및 계통선발, 목표 형질 접근 방법에 효율화 <p>○ 연구내용 및 결과</p> <ol style="list-style-type: none"> 1.육종 특화 데이터베이스 구축을 위한 기반 정보 수집 및 업데이트 <ul style="list-style-type: none"> - 수박, 고추, 파프리카, 무의 육종특화 데이터베이스 구축 및 정보 수집 - 염기서열, 전사체, 변이 분석 정보 수집 및 가공 및 정보를 활용한 Indel 마커 개발 2.신규 분자 마커 개발을 위한 유전연구 <ul style="list-style-type: none"> - 전사체 정보를 활용한 마커 생산 방식 개발 및 네트워크 연구 - 유용 형질에 대한 QTL, GWAS 분석 및 신호전달경로/유전자 네트워크의 분석 기법 개발 - 저항성과 이병성 품종의 유전자 발현 프로파일, 네트워크 Pathway 기반 차이 분석 3.채소 작물의 유용 형질관련 분자마커의 범용화 <ul style="list-style-type: none"> - 수집단의 변이 정보 생산 방식 개선 및 annotation 방식의 확립 - 배추 표준 유전체 기반 정보의 생산 및 마커 개발 파이프라인 구성 및 평가 - 유용 표현형 관련 염기서열의 범용 마커로의 전환 4.분자마커의 high-throughput selection을 위한 마커 이전 서비스 <ul style="list-style-type: none"> - 마커 개발 및 high-throughput system 용 마커 개발 및 범용성 검증 - 카로티노이드 패스웨이 상의 변이 확보를 위한 파이프라인의 개발 5.종자산업 발전 및 마케팅 전략 구축을 위한 기반 연구 <ul style="list-style-type: none"> - 국제협력사업 관계 기관들과의 협력체 구축을 통한 채소 종자 수출 활성화 기반 조성 - 종자산업 발전 및 마케팅 전략 구축을 위한 기반연구 - 국내 종자산업 현황과 국제경쟁력 제고 방안 연구 <p>○ 연구성과 활용실적 및 계획</p> <ol style="list-style-type: none"> 1.기술적 측면 <ul style="list-style-type: none"> - NGS를 통한 대량 분자마커 발굴을 이용하는 육종 기술의 진화. - 데이터베이스 구축, 형질에 관련 유전자 분석, 분자마커 개발 소요 시간 단축, 효율적 분자유종 시스템 구축. - 체계화된 분자마커 지원을 위한 사용자 중심의 데이터베이스 및 분석 지원 - 분자마커 육종지원시스템 개발을 통해 육종기술의 체계화 및 타 작물에 대한 유사 시스템 기술지원. - 유전 분석을 통해 개발한 분자 마커는 집단 작성 및 전개에 소요되는 시간과 비용을 크게 절감할 수 있음 2.경제적/산업적 측면 <ul style="list-style-type: none"> - 복잡 형질의 유전 현상을 시스템적으로 이해하고 예측 가능한 육종 계획을 세움으로 육종 기간을 단축하고 맞춤형육종(breeding-by-design)을 구현할 수 있는 기반을 제공함. - 육종가에게 동일한 분자마커의 사용을 유도함으로써 육종 연한의 감소를 이룩함. - 신품종 육성 가속화로 인한 육종 기간 단축과 분자마커 개발 및 파생 기술을 이용한 산업화를 통하여 다가오는 바이오경제의 경쟁력을 확보할 수 있을 뿐만 아니라 관련 민간기업의 활성화에 기여할 수 있음. - 종자 개발의 효율성 극대화로 시장에서 경쟁력 확보 및 수출 증대 기대. - 빠르게 변화하는 시장의 소비자 요구에 부합하는 품종 개발의 효율성 증진 - 국내외 종자산업의 현황 분석과 마케팅전략 분석을 통하여 각 종자회사들의 국제경쟁력 제고 방안에 대한 효율적인 품종 개발 및 마케팅 방안 제공 | | | | | |

[별첨 2]

연구성과 활용계획서 (2017~2021)

1. 연구과제 개요

| | | | | | |
|-----------|--|-------|-----------|--------------------|-----------|
| 사업추진형태 | <input type="checkbox"/> 자유응모과제 <input checked="" type="checkbox"/> 지정공모과제 | | 분 야 | 육종특화 통합 DB 구축 및 운영 | |
| 프로젝트명 | 채소작물의 종자개발을 위한 육종 특화 통합 DB 구축 및 운영 | | | | |
| 프로젝트 연구기관 | 충남대학교 | | 프로젝트연구책임자 | | 임용표 |
| 연구개발비 | 정부출연 연구개발비 | 기업부담금 | 연구기관부담금 | 정부외출연금 | 총연구개발비 |
| | 3,279,000 | | | 750,000 | 4,029,000 |
| 연구개발기간 | 2017.01.01 ~ 2021.12.31 | | | | |
| 주요활용유형 | <input checked="" type="checkbox"/> 산업체이전 <input type="checkbox"/> 교육 및 지도 <input type="checkbox"/> 정책자료 <input checked="" type="checkbox"/> 기타(데이터 통합 및 제공) <input type="checkbox"/> 미활용 (사유:) | | | | |

2. 연구목표 대비 결과

| 당초목표 | 당초연구목표 대비 연구결과 |
|--|---|
| ①5대 채소 작물 관련 신규 가용 정보의 지속적 업데이트 | <ul style="list-style-type: none"> - 5대 채소 작물의 육종특화 데이터베이스 구축 및 정보 수집 - 염기서열, 전사체, 변이 분석 정보 수집 및 가공 - 수박 옹성불임 전사체 정보를 활용한 Indel 마커 개발 |
| ②신규 분자 마커 개발을 위한 유전연구 | <ul style="list-style-type: none"> - 전사체 정보를 활용한 마커 생산 방식 개발 및 네트워크 연구 - 유용형질에 관한 QTL, GWAS 분석 - 배추 유용형질 관련 신호전달경로/유전자 네트워크의 분석 기법 개발 - 저항성과 이병성 품종의 유전자 발현 프로파일, 네트워크 Pathway 기반 차이 분석 |
| ③채소 작물의 유용 형질관련 분자마커의 범용화 | <ul style="list-style-type: none"> - 수집단의 변이 정보 생산 방식 개선 및 annotation 방식의 확립 - 배추 표준 유전체 기반 정보의 생산 및 마커 개발 파이프라인 구성 및 평가 - 유용 표현형 관련 염기서열의 범용 마커로의 전환 |
| ④분자마커의 high-throughput selection을 위한 마커 제공 | <ul style="list-style-type: none"> - 마커 개발 및 high-throughput system 용 마커 개발 및 범용성 검증 - 카로티노이드 패스웨이 상의 변이 확보를 위한 파이프라인의 개발 |
| ⑤종자산업 발전 및 마케팅 전략 구축 | <ul style="list-style-type: none"> - 국제협력사업 관계 기관들과의 협력체 구축을 통한 채소 종자 수출 활성화 기반 조성 - 종자산업 발전 및 마케팅 전략 구축을 위한 기반연구 - 국내 종자산업 현황과 국제경쟁력 제고 방안 연구 |

* 결과에 대한 의견 첨부 가능

3. 연구비 집행실적 (2017~2021 누적)

| 구분 | 금액 | | 계획금액 | 사용액 | 잔액 | 비고 |
|------|---------------------------------------|--|--------------|---------------|-----------|----|
| | 세부프로젝트명 | | | | | |
| 단장과제 | 채소작물의 종자개발을 위한 육종 특화 통합 DB 구축 및 운영 | | 3,642,000천 원 | 3,633,600.913 | 8,399.087 | |
| | | | | | | |
| | | | | | | |
| 총계 | | | 3,642,000천 원 | 3,633,600.913 | 8,399.087 | |

4. 연구목표 대비 성과

| 구분 | 품종개발 | | 특허 | | 논문 | | 분자 마커 | DB 구축 | 유전자원 | | 국내 매출액 | 종자 수출액 | 기술 이전 | 마케팅 전략 보고서 | 인 력 양 성 |
|----------|--------|--------|--------|--------|-----|------|----------|----------|--------|--------|-----------|-----------|----------|------------------|------------------|
| | 출 원 | 등 록 | 출 원 | 등 록 | SCI | 비SCI | | | 수 집 | 등 록 | | | | | |
| 최종목표 | | | | | 3 | 1 | | 1 | | | | | | | |
| 최종실적 | | | 5 | 7 | 27 | 2 | 3 | 1 | 60 | 18 | | | | | 16 |
| 달성율(%) | | | 100 | 100 | 900 | 200 | 100 | 100 | 100 | 100 | | | | | 100 |
| 1차 년도 | 목표 | | | | | | | | | | | | | | |
| | 실적 | | | 2 | 4 | | 3 | | | | | | | | 6 |
| 2차 년도 | 목표 | | | | 1 | | | | | | | | | | |
| | 실적 | | | 2 | 1 | 1 | 1 | | 60 | | | | | | 2 |
| 3차 년도 | 목표 | | | | | 1 | | | | | | | | | |
| | 실적 | | | 1 | | 5 | 1 | | | | | | | | 2 |
| 4차 년도 | 목표 | | | | 1 | | | | | | | | | | |
| | 실적 | | | | 3 | 6 | | | | 18 | | | | | 3 |
| 5차 년도 | 목표 | | | | 1 | | | 1 | | | | | | | |
| | 실적 | | | 2 | 1 | 11 | | 1 | | | | | | | 3 |

5. 핵심기술

| 구분 | 핵심기술 명 |
|----|-----------------------------------|
| ① | 분자유종 활성화를 위한 특화된 데이터베이스 구축 및 업데이트 |
| ② | 신규 분자 마커 개발을 위한 유전연구 |
| ③ | 종자산업 발전 및 마케팅 전략 구축을 위한 기반 연구 |

6. 연구결과별 기술적 수준

| 구분 | 핵심기술 수준 | | | | | 기술의 활용유형(복수표기 가능) | | | | |
|-------|---------|-------|---------|------------|------------|-------------------|-------------|---------|-------|----|
| | 세계 최초 | 국내 최초 | 외국기술 복제 | 외국기술 소화·흡수 | 외국기술 개선·개량 | 특허 출원 | 산업체이전 (상품화) | 현장애로 해결 | 정책 자료 | 기타 |
| ①의 기술 | | v | | | | | | v | | |
| ②의 기술 | v | v | | | v | | | v | | |
| ③의 기술 | | v | | | v | | | v | | |

* 각 해당란에 v 표시

7. 각 연구결과별 구체적 활용계획

| 핵심기술 명 | 핵심기술별 연구결과활용계획 및 기대효과 |
|--------|---|
| ①의 기술 | 분자유종 기술로 생산된 데이터에 접근하기 어려운 육종가를 위하여 만들어 졌으며, 이를 통해 형질에 관련된 유전자의 분석된 정보를 제공함으로써, 분자마커 개발에 소요되는 시간을 단축하여 효율적이고 차원 높은 분자유종 시스템을 구축하는데 기여할 수 있음. |
| ②의 기술 | 작물 유용한 특성(노화, 뿌리혹병, 잎털, 결각)에 대한 유전 분석(GWAS, QTL)을 진행하였다. 여기서 개발된 표현형 분석 방법은 육종가들의 균일한 표현형 평가 방법에 사용될 수 있을 것이며, 분석된 유전형질 또는 분자 마커는 육종가들의 형질관련 기초자료와 계통선발에 사용될 것이다. |
| ③의 기술 | 국내외 종자 산업의 현황과 문제점을 파악하였으며, 이를 해결하기 위한 다양한 방향들을 제시하였다. 국제협력사업 관계 기관들과의 협력체 구축을 통한 채소 종자 수출 활성화 기반 조성과 종자 산업 발전 및 마케팅 전략 구축을 위한 기반 연구로 맞춤형 품종개발을 제시하였다. 세계 종자시장기업과 국제 정체를 기반으로 서술된 시나리오들은 국내 종자기업의 국제 경쟁력 제고 방안에 기초 자료로 사용될 것이다. |

8. 연구종류 후 성과창출 계획

| 구분 | 품종개발 | | 특허 | | 논문 | | 분 자 마 커 | 유전자원 | | 국내 매출 액 | 종 자 수 출 액 | 기 술 이 전 | 마 케 팅 전 투 입 보 고 서 | 인 력 양 성 |
|-------------------|--------|--------|--------|--------|-----|------|------------------|--------------|--------|---------------|-----------------------|------------------|---|------------------|
| | 출 원 | 등 록 | 출 원 | 등 록 | SCI | 비SCI | | DB 구 축 | 수 집 | | | | | |
| 최종목표 | | | | | 3 | 1 | | 1 | | | | | | |
| 연구기간 내 달성실적 | | | 5 | 7 | 27 | 2 | 3 | 1 | 60 | 18 | | | | 16 |
| 연구종료 후 성과창출 계획 | | | | | | | | | | | | | | |

9. 연구결과의 기술이전조건(산업체이전 및 상품화연구결과에 한함)

| | | | |
|-------------|--|---------|----|
| 핵심기술 명 | | | |
| 이전형태 | <input type="checkbox"/> 무상 <input type="checkbox"/> 유상 | 기술료 예정액 | 천원 |
| 이전방식 | <input type="checkbox"/> 소유권이전 <input type="checkbox"/> 전용실시권 <input type="checkbox"/> 통상실시권 <input type="checkbox"/> 협의결정 <input type="checkbox"/> 기타() | | |
| 이전소요기간 | | 실용화예상시기 | |
| 기술이전 시 선행조건 | | | |

* 핵심기술이 2개 이상일 경우에는 각 핵심기술별로 위의 표를 별도로 작성

** 기술이전 시 선행요건 : 기술실시계약을 체결하기 위한 제반 사전협의사항(기술지도, 설비 및 장비 등 기술이전 전에 실시기업에서 갖추어야 할 조건을 기재)

*** 실용화예상시기 : 상품화인 경우 상품의 최초 출시 시기, 공정개선인 경우 공정개선 완료시기 등

[별첨 3]

자체평가보고서

| | | | | | |
|----------|------------------------------------|------------------------|-------------------|----------------------|--|
| 사업단명 | GSP채소종자사업단 | 과제번호 | 213006-05-5-CG100 | | |
| 프로젝트명 | 채소작물의 종자개발을 위한 육종 특화 통합 DB 구축 및 운영 | | | | |
| 프로젝트연구기관 | 충남대학교 | | | | |
| 연구담당자 | 프로젝트 연구책임자 | 임용표 | | | |
| | 세부프로젝트 연구책임자 | 기관(부서) | | 성명 | |
| | | 기관(부서) | | 성명 | |
| | | 기관(부서) | | 성명 | |
| | | 기관(부서) | | 성명 | |
| 연구기간 | 총기간 | 2017.01.01.~2021.12.31 | 당해연도기간 | 2021.1.1.~2021.12.31 | |
| 연구비(천원) | 총규모 | 2,892,000 | 당해연도규모 | 592,000 | |

1. 연구는 당초계획대로 진행되었는가?

당초계획 이상으로 진행
 계획대로 진행
 계획대로 진행되지 못함

○ 계획대로 수행되지 않은 원인은?

2. 당초 예상했던 성과는 얻었는가?

예상외 성과 얻음
 어느 정도 얻음
 얻지 못함

| 구분 | 품종개발 | | 특허 | | 논문 | | 분자 마커 | 유전자원 | | 국내 매출액 | 종자 수출액 | 기술 이전 | 마케팅 전략 추진 보고서 | 인력 양성 |
|----------------|------|----|-----|-----|-----|------|----------|------|-----|-----------|-----------|----------|------------------------|----------|
| | 출원 | 등록 | 출원 | 등록 | SCI | 비SCI | | 수집 | 등록 | | | | | |
| 최종목표 | | | | | 3 | 1 | | 1 | | | | | | |
| 연구기간 내 달성실적 | | | 5 | 7 | 27 | 2 | 3 | 1 | 60 | 18 | | | | 16 |
| 달성율(%) | | | 100 | 100 | 900 | 200 | 100 | 100 | 100 | 100 | | | | 100 |

3. 연구개발 성과 세부 내용

3-1 기술적 성과

- 육종가의 경험적인 역량으로 선발과 교배가 이루어져 왔던 전통육종에 NGS(차세대 염기서열분석)를 통해 발굴된 대량 분자마커를 활용하여 현대 육종의 진화를 도모함.
- 분자육종에 전문화된 데이터베이스를 구축하여 작물 육종에서 필요로 하는 형질에 관련된 유전자의 분석된 정보를 제공함으로써, 분자마커 개발에 소요되는 시간을 단축하여 효율적이고 차원 높은 분자육종 시스템을 구축하는데 기여할 수 있음.
- 육종 기술에 대한 사용자 중심의 데이터베이스 및 분석 지원 시스템을 통하여 검증된 분자마커의 정보를 국내 육종공동체로의 보급하여 신품종 육종과정에서 체계적으로 활용이 가능함.
- 5대 작물에 대한 분자마커 육종지원시스템 개발을 통해 분자마커를 이용하는 육종기술의 체계화 및 타 작물에 대한 유사 시스템 기술지원이 가능함.
- 작물 육종 활동에 대한 GWAS 분석 방식의 접목은 다양한 표현형 중 특정 형질에 대한 신규 분자마커의 개발을 집단 작성 및 전개에 소요되는 시간과 비용을 크게 절감할 수 있음.

3-2 과학적 성과

- QTL GWAS 분석을 통하여 작물의 유용한 형질과 연관된 마커 또는 후보유전자를 탐색할 수 있었다. 배추의 노화에 관한 연구로 GWAS과 배추의 뿌리혹병 저항성에 연관노던 QTL 분석 그리고 무의 열근 현상에 대한 QTL 분석을 통하여 각각의 형질에 대한 분자유전적 기초 자료를 제공하였으며 이를 이용하여 형질 연관 마커를 개발하는 초석이 될 것이다.
- 배추의 조직 특이적인 유전자 발현 분석을 진행하였으며 이를 통하여 각각의 유전자간 네트워크 분석을 하였다. 이는 배추의 유전자의 신호전달 및 유전자 네트워크 동정을 통하여 배추를 형질을 이해하는 기초자료가 될 것으로 기대
- GWAS 분석 시스템의 구축은 관심 형질에 대한 candidate gene 상의 유전자SNP를 신속하게 분자마커화 할 수 있기 때문에 이를 활용한 육종과 신품종의 생산 및 시장 진출을 유도하는 기반 연구의 초석을 마련함.

3-3 경제적 성과

- MAB와 MAS를 통해 신품종 육성에 대한 최적의 교배 조합과 필요 개체수의 예측으로 육종 효율성을 극대화시켜 우리 종자의 세계시장에서의 경쟁력을 확보하고 종자 수출량의 증대를 기대.
- 시장 동향 파악을 통해 빠르게 변화하는 세계 종자시장의 수요에 대한 맞춤형 품종 개발의 효율성을 극대화 할 것으로 기대.

3-4 사회적 성과

- 국제 종자 시장 동향 파악을 통해 빠르게 변화하는 세계 종자시장의 수요에 대한 맞춤형 품종 개발을 도울 수 있는 다양한 경제 지표를 제공함으로써 국내 종자회사의 수출 증대로 국가 발전에 기여할 수 있다.

3-5 인프라 성과

- 복잡 형질의 유전 현상을 시스템적으로 이해하고 예측 가능한 육종 계획을 세움으로 육종 기간을 단축하고 맞춤형육종(breeding-by-design)을 구현할 수 있는 기반을 제공함.
- 공통된 작물을 육종하는 여러 육종가에게 동일한 분자마커의 사용을 유도함으로써 육종 연한의 감소를 이룩함.
- 다양한 작물 종을 대상으로 연구하는 분자육종가에게 공통의 분자마커 사용을 촉진함으로써, 신품종 육성 가속화로 인한 육종 기간 단축과 분자마커 개발 및 파생 기술을 이용한 산업화를 통하여 다가오는 바이오경제의 경쟁력을 확보할 수 있을 뿐만 아니라 관련 민간기업의 활성화에 기여할 수 있음.

4. 연구과정 및 성과가 농림어업기술의 발전·진보에 공헌했다고 보는가?

- 공헌했음 현재로서 불투명함 그렇지 않음

7. 연구개발 착수 이후 국내 다른 기관에서 유사한 기술이 개발되거나 또는 기술 도입함으로 연구의 필요성을 감소시킨 경우가 있습니까?

- 없다 약간 감소되었다 크게 감소되었다

○ 감소되었을 경우 구체적인 원인을 기술하여 주십시오.

8. 관련된 기술의 발전속도나 추세를 감안할 때 연구계획을 조정할 필요가 있다고 생각하십니까?

- 없다 약간 조정필요 전반적인 조정필요

9. 연구과정에서의 애로 및 건의사항은?

(※ 아래사항은 기업참여시 기업대표가 기록하십시오)

1. 연구개발 목표의 달성도는?

- 만족 보통 미흡

(근거 : 정량적 목표로 논문과 DB 구축이 있다. 논문의 총 29개의 논문을 출판하였으며 목표를 초과하였다. DB구축 목표 기간내에 달성하였다. 정성적 목표 또한 전부 달성하였기 때문에 연구개발 목표의 달성도에 만족하였다.)

2. 참여기업 입장에서 본 본과제의 기술성, 시장성, 경제성에 대한 의견

가. 연구 성과가 참여기업의 기술력 향상에 도움이 되었는가?

- 충분 보통 불충분

나. 연구 성과가 기업의 시장성 및 경제성에 도움이 되었는가?

- 충분 보통 불충분

3. 연구개발 계속참여여부 및 향후 추진계획은?

가. 연구수행과정은 기업의 요청을 충분히 반영하였는가?

- 충분 보통 불충분

나. 향후 계속 참여 의사는? (※중간·단계평가에 한함)

- 충분 고려 중 중단


다. 계속 참여 혹은 고려중인 경우 연구개발비의 투자규모(전년도 대비)는? (※중간·단계평가에 한함)

- 확대 동일 축소

4. 연구개발결과의 상품화(기업화) 여부는?

- 즉시 기업화 가능 수년 내 기업화 가능 기업화 불가능

5. 기업화가 불가능한 경우 그 이유는?

| 구 분 | 소 속 기 관 | 직 위 | 성 명 |
|----------|---------|-----|--|
| 프로젝트 책임자 | 충남대학교 | 교수 | 임 용 표 (인 ) |

주 의

1. 이 보고서는 농림축산식품부에서 시행한 Golden Seed 프로젝트 사업 연구개발과제 최종 보고서이다.
2. 이 연구개발내용을 대외적으로 발표할 때에는 반드시 농림축산식품부(농림식품기술기획평가원)에서 시행한 Golden Seed 프로젝트 사업의 결과임을 밝혀야 한다.
3. 국가과학기술 기밀 유지에 필요한 내용은 대외적으로 발표 또는 공개하여서는 안 된다.