Vol.2

AI DRIVES DATA BREEDING

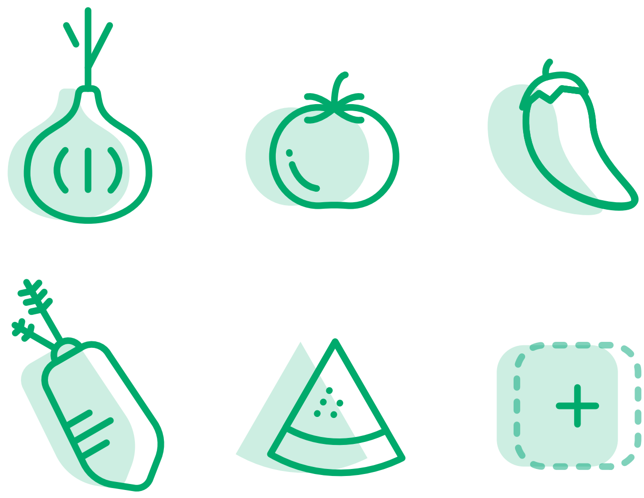
DATA BREEDING STORY

ibreeder

양파 구중 개량

고추 여교잡 육종

분자 마커 실험·검증 GUI 솔루션



㈜인실리코젠의 데이터 육종은

생물정보, 빅데이터, AI 기술을 융합하여 *in silico* 상에서 새로운 종자를 찾아 미래 바이오 산업의 근간이 될 가치를 창출합니다.

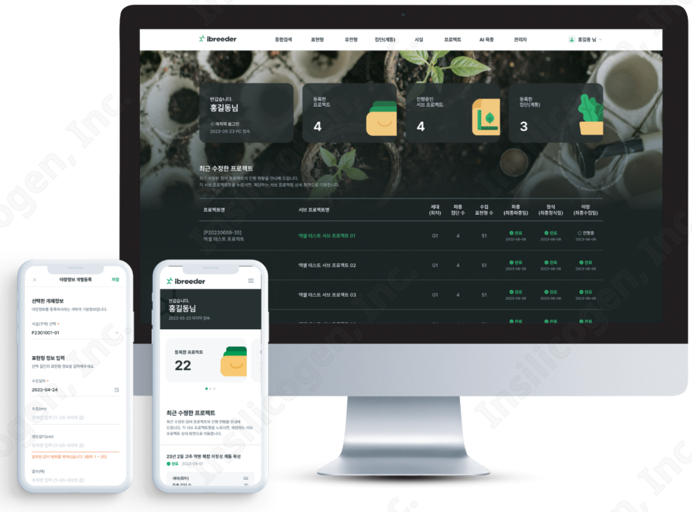
OFFICE	경기도 용인시 기흥구 흥덕1로 13 흥덕IT밸리 타워A동 2901~2904, 2906호
EMAIL	info@insilicogen.com
PHONE	031-278-0061
FAX	031-278-0062

㈜ 인실리코젠

ibreeder

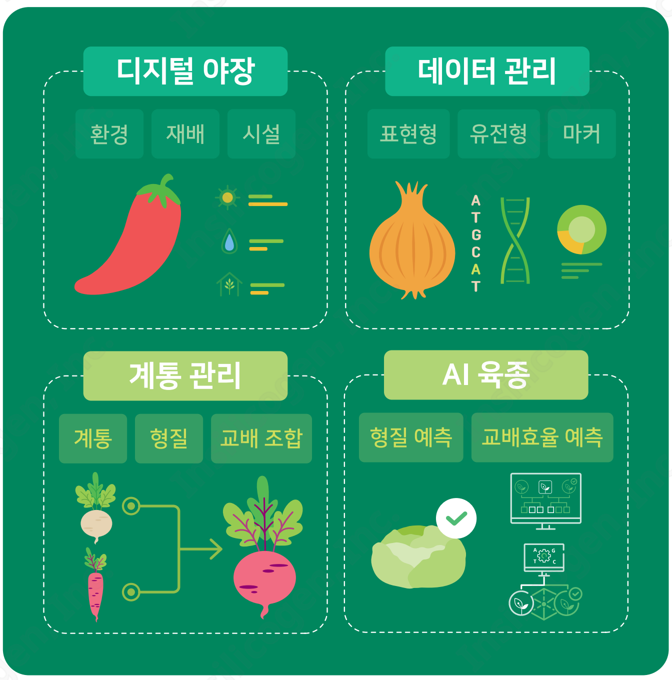
데이터 육종을 위한 All-in-One 플랫폼

아이브리더는 웹 기반 데이터 육종 플랫폼으로, 유전자원, 표현형 및 유전형 정보를 규격화하여 효율적으로 관리합니다.



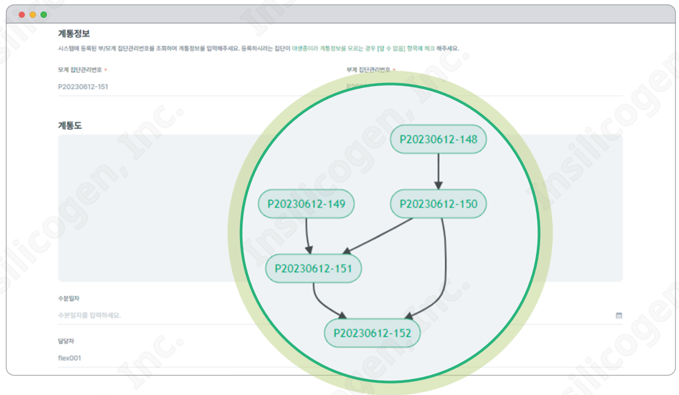
주요 기능

- 작물에 대한 유전자원, 원예 형질 및 유전형 정보 관리
- 작물의 주요 형질에 따른 마커 등록 및 상세 조회
- 통합 검색을 통한 우수 자원 선발
- 인공지능 기반 형질 예측 및 선발
- AI 교배 조합 시뮬레이션으로 신품종 개발 지원
- 사용자 권한 관리를 통한 데이터 보안 유지



집단(계통) 및 교배 대장 관리

- 교배 모부본 관리 및 계통 육성 이력(계통도) 제공
- 육종 목표에 따른 교배 계획 수립
- 교배 조합별 파종, 가식, 정식 이력 관리

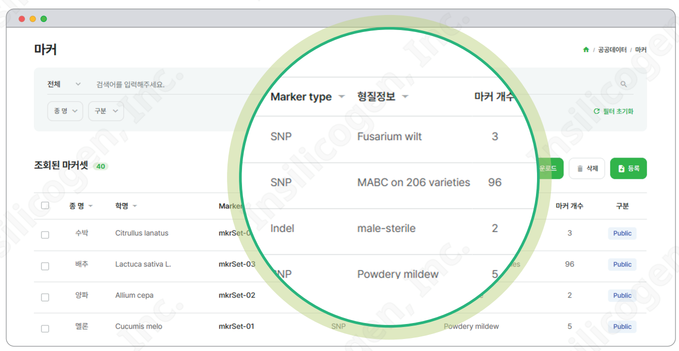


원예 형질 관리(디지털 야장)

- 대용량 야장정보의 손쉬운 등록
- 야장정보 등록 시 자동 오류 검사
- 질적/양적 원예 형질 측정값, 영상, 재배정보 등 다양한 형식 지원
- 관심 원예 형질 검색을 통한 우수 자원 선발

데이터 육종을 위한 데이터 관리

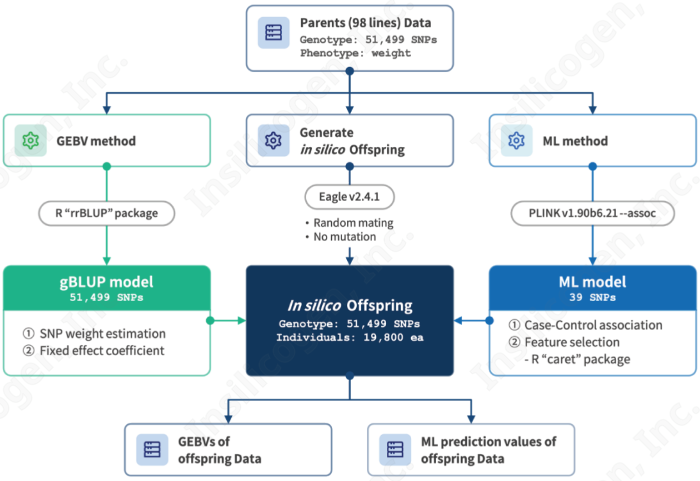
- 유전자원, 유전형, 원예 형질, 마커 데이터 정규 포맷으로 통합 관리
- 전통 육종, 분자 육종, AI 육종에 즉각 활용 가능
- 공공데이터(참조 유전체, 마커) 정보 제공



양파 구중 개량:
시뮬레이션을 통한 교배 조합 작성법

형질	양파 구중
ML model	PLS
모델 정확도	83.2%
마커 수	39 SNPs
학습/검증 데이터 수	60 / 신규 260 개체

- 유전자원: 국내 양파 98계통
- 표현형: 양파 구의 중량
- 유전형: 계통별 1.5 Gb의 GBS 시퀀싱 서열, 51,499 SNPs
- 마커 탐색 방법: GWAS (case-control), feature selection
- 정확도 검증법: 260개 신규 개체의 측정치와 예측치 비교



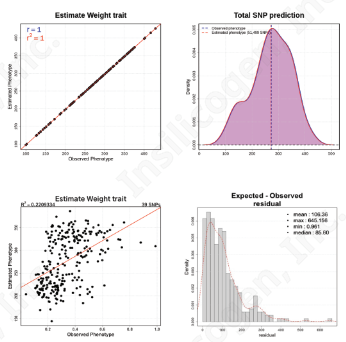
[Fig.1] 양파의 디지털 육종 모식도

1) GEBV method: 모부본의 유전형을 바탕으로 육종가 산출
2) ML method: 기계학습을 이용한 표현형 예측
3) 검증용 신규 데이터를 활용한 method 별 정확도 검증
4) 교배를 위한 자손 세대 인실리코 유전형 생성
5) 인실리코 유전형을 바탕으로 표현형 예측 및 고효율 교배 조합 산출

엘리트 계통의 51,499 SNPs를 이용한 gBLUP 육종가 모델

- 총 98건의 엘리트 계통으로 모델 구축
- gBLUP을 활용한 표현형 예측 모델
- 51,499 SNPs 이용
- r^2=1.00, 평균오차=2.204(g)

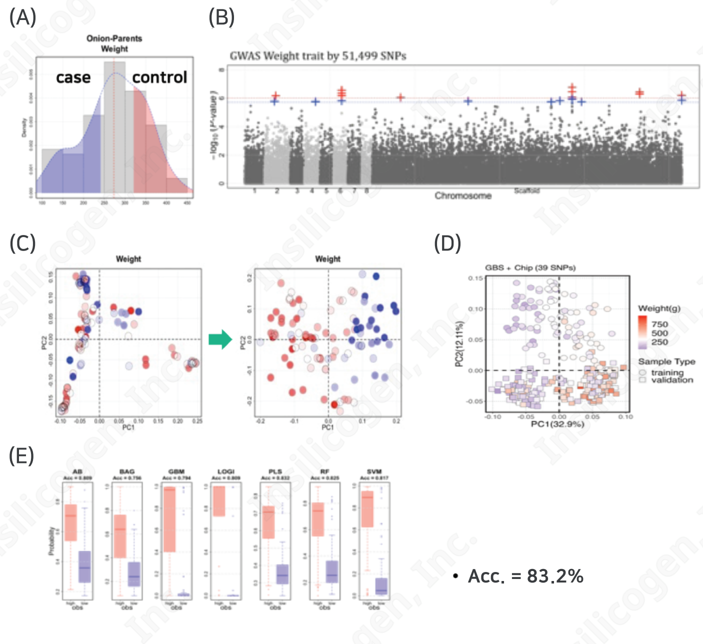
- 총 260건의 신규 개체로 모델 검증
- gBLUP을 활용한 표현형 예측 모델
- 39 SNPs 이용
- r^2=0.22, 평균오차=106.36(g)



[Fig.2] gBLUP을 이용한 육종가 예측 및 표현형 분포도

엘리트 계통의 구중 관련 기계학습 모델 구축

- 1차 GWAS를 통한 주요 1,000 SNPs 선발
- 2차 기계학습의 feature selection을 이용한 최종 39 SNPs 선발
- 39 SNPs을 이용한 양파 구중 예측 모델 구축(정확도: 83.2%)

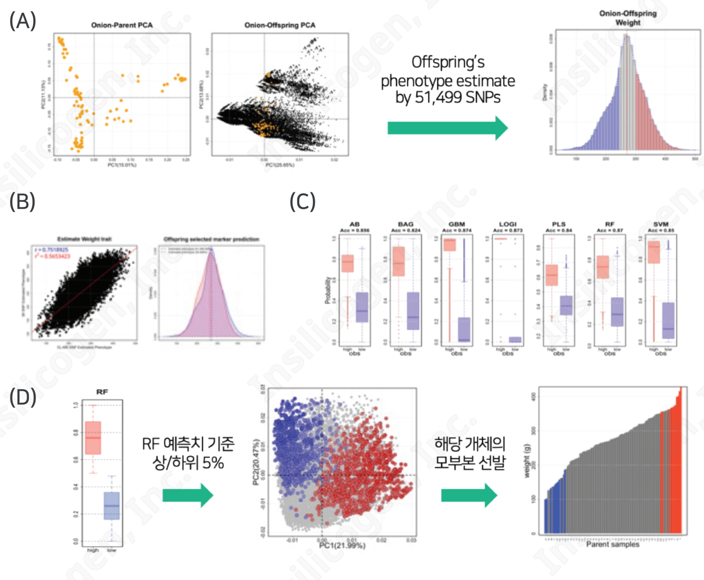


[Fig.3] 구중 관련 마커 탐색 및 예측 모델 구축

(A) 구중 표현형 분포; (B) GWAS 분석 결과 manhattan plot; (C) Feature selection 결과 마커의 표현형 구분력; (D) 엘리트 계통과 검증용 신규 개체의 구중 및 유전적 분포도; (E) ML 알고리즘에 의한 검증 데이터의 구중 예측치

교배 조합 작성: 자손 세대 시뮬레이션(예측 모델 적용)

- 부모 개체의 유전형 정보를 활용한 인실리코 자손 19,800 개체 생산
- 인실리코 자손의 구중 예측치
- 최종 우수 교배 조합 가계 선발



[Fig.4] 교배 조합 선발을 위한 인실리코 자손 시뮬레이션

(A) 인실리코 자손 유전형 생산; (B) gBLUP 모델을 이용한 자손의 구중 예측치; (C) ML 알고리즘에 의한 자손의 구중 예측치; (D) 우수 알고리즘 기준 상/하위 5% 자손의 유전형과 모부본 표현형 분포로 우수 및 미흡 교배 조합 제시

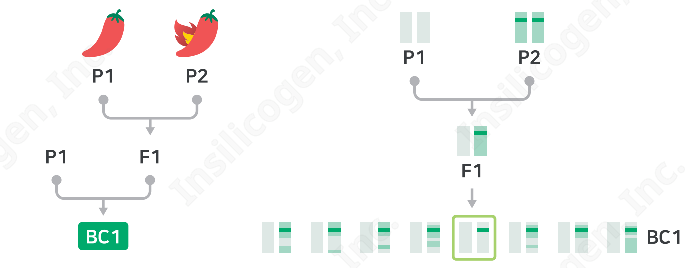
고추 여교잡 육종: MABC 마커 탐색

고추 80계통을 이용한 MABC 마커 탐색

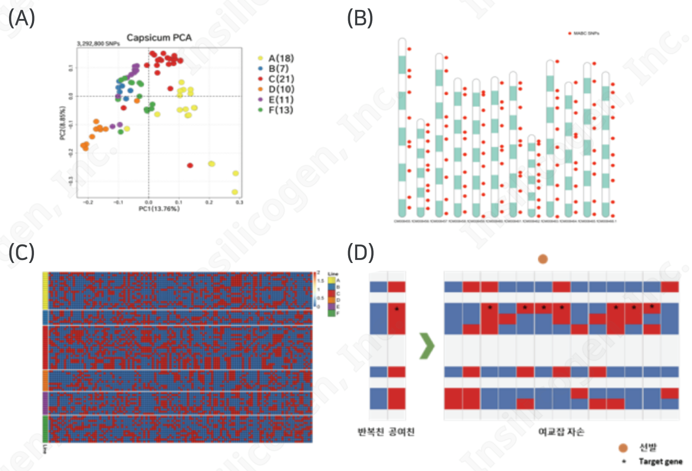
유전자원 고추 80계통

유전형 계통별 34.4 Gb의 WGS 서열, 2,997,866 SNPs

마커 적용 12개 마커: 주요 25계통의 여교잡 자손 background 선발 가능
108개 마커: 전체 80계통의 여교잡 자손 background 선발 가능



[Fig.1] MABC (Marker-assisted backcrossing) 모식도



[Fig.2] 고추 계통 분류 관련 MABC 탐색

(A) 고품질 변이로 확인한 80계통의 유전적 유사도; (B) 108개 MABC 마커의 염색체 분포도; (C) 108개 MABC 마커를 이용한 80계통의 변이 양상; (D) MABC를 이용한 자손 개체 선발

• 제작 가능한 분자 마커 리스트

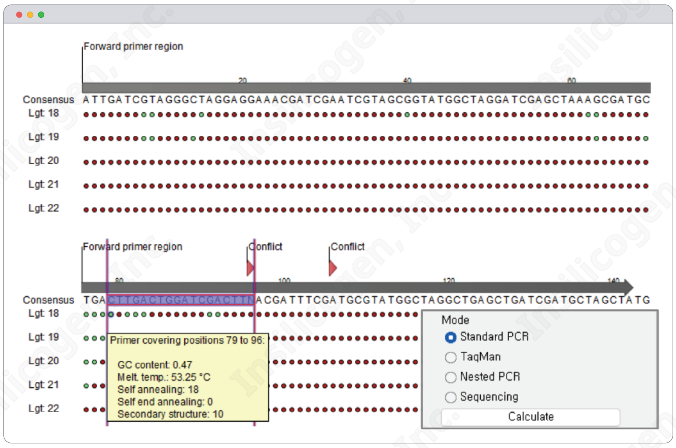
TaqMan	PCR 증폭 시 분해되는 probe의 형광으로 genotyping
KASP	대립 유전형에 맞는 2종 이상 primer의 경쟁적인 증폭으로 genotyping
rhAmp SNP	TaqMan의 형광 기술과 KASP의 경쟁적 증폭을 혼합한 genotyping
CAPS	변이 서열 내 제한효소의 절단 유무로 genotyping
SCAR	PCR을 이용한 특정 변이 영역의 증폭으로 genotyping
HRM	변이 서열에 따른 DNA 이중나선의 용융 온도 차이로 genotyping
Fluidigm chip	Mini SNP chip으로 대량의 샘플을 효율적으로 genotyping
BTseq	NGS를 이용한 타겟 영역의 경제적 genotyping

분자 마커 실험·검증 GUI 솔루션: CLC Main Workbench

분자 마커 프라이머 디자인

• MABC, 웅성불임, 품종판별, 마커 이용선발(MAS) 등 다양한 분자 마커 적용

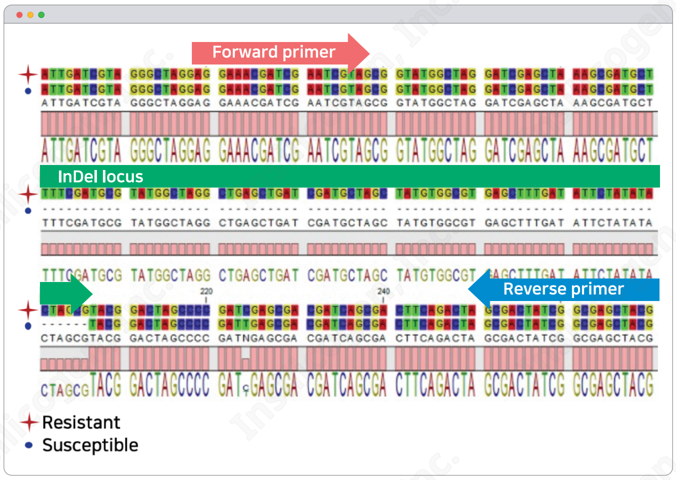
• HRM, KASP, CAPS, SCAR, TaqMan 등 다양한 유형의 프라이머 디자인



계통별 다형성 영역 시각화

• 프라이머 세트에 의해 증폭되는 영역의 유전적 다형성 분석 및 시각화

• 유전적 다형성을 토대로 빠르고 쉽게 표현형 판별



계통도 및 계통 간 친화도 분석

• 계통 간 유전적 유연관계 파악을 위한 계통 분석