

양배추 경제형질 연관 마커 선별과 가상자손 시뮬레이션을 통한 우수 교배 조합 선별

박한용², 임재원¹, 정명희¹, 손효정¹, 임유진¹, 이은채¹, 송예진¹, 장지희¹, 신유희¹, 최수빈², 박병준², 조성현^{1,*}

¹ 경기 용인시 기흥구 흥덕1로 13 흥덕IT밸리 (주)인실리코젠 R&D 센터
² 서울특별시 광진구 능동로 209 세종대학교 바이오산업자원공학과

INSILICOGEN
www.insilicogen.com
|주|인실리코젠



Abstract

양배추(*Brassica oleracea* var. capitata L.)는 십자화과에 속하는 관속식물로, 세계 3대 건강식품으로 선정될 정도로 많은 영양소를 보유하고 있다. 그러나 최근 기후변화로 인한 재배 환경 악화와 병해충 발생 증가로 생산성이 저하되고 있다. 이는 가격 상승과 수입량 증가로 이어지면서 식량 안보의 약화를 유도하고 있다. 이에 대응하여 양배추의 생산량을 증대하기 위해 디지털육종을 활용하여 양배추의 경제형질 연관 마커를 선별하고 *in silico* 예측을 통해 품종 개량 방안을 시뮬레이션하였다. 본 연구에서는 양배추 80개체를 대상으로 WGS (Whole Genome Sequencing)을 수행하였으며, 여러 분석과 QC를 거친 결과 2,910,080개의 고품질 SNPs를 확보할 수 있었다. 양배추의 생산성 연관된 표현형으로는 구형, 밀도, 고갱이 비율, 구중의 4가지를 선정하였으며, 시퀀싱을 수행한 개체 중 56개를 선별하여 측정되었다. GWAS (Genome Wide Association Study), GBLUP (Genomic Best Linear Unbiased Prediction), ML (Machine Learning) 등의 방법을 활용하여 경제형질 연관 마커 96개를 발굴하였다. 발굴한 SNPs 마커를 활용하여 기계학습 기반 예측 모델을 구축하고 예측 정확도를 확인한 결과, 예측 정확도는 평균 89.8%로 산출되었다. 또한, 본 연구를 통해 얻은 마커를 활용하여 가상 자손의 유전형과 표현형을 추정하고 이를 비교함으로써 우수 교배 조합을 선별하였다. 80개의 양배추 개체에서 총 3,240가지의 교배 조합을 산출하였으며, 각 조합에서 총 10,000개의 가상 자손을 생성하였다. 기계학습을 통해 산출된 경제형질 예측치를 비교하여 우수 및 열등 조합을 각각 5개씩 선별하였으며, 실제 해당 조합의 교배 및 재배를 통해 *in silico* 품종 개량 방법의 효용성을 검증하였다. 본 연구의 결과는 우수 개체의 선발 및 교배 조합 작성을 빠르고 정밀하게 할 수 있는 방법을 제시하고, 동시에 이러한 데이터 기반 육종의 효용성을 입증한다. 이는 양배추뿐만 아니라 다양한 작물에 활용할 수 있으며, 생산량 증대 및 식량 안보를 확보하는데, 활용될 수 있을 것으로 기대된다.



Materials & Methods

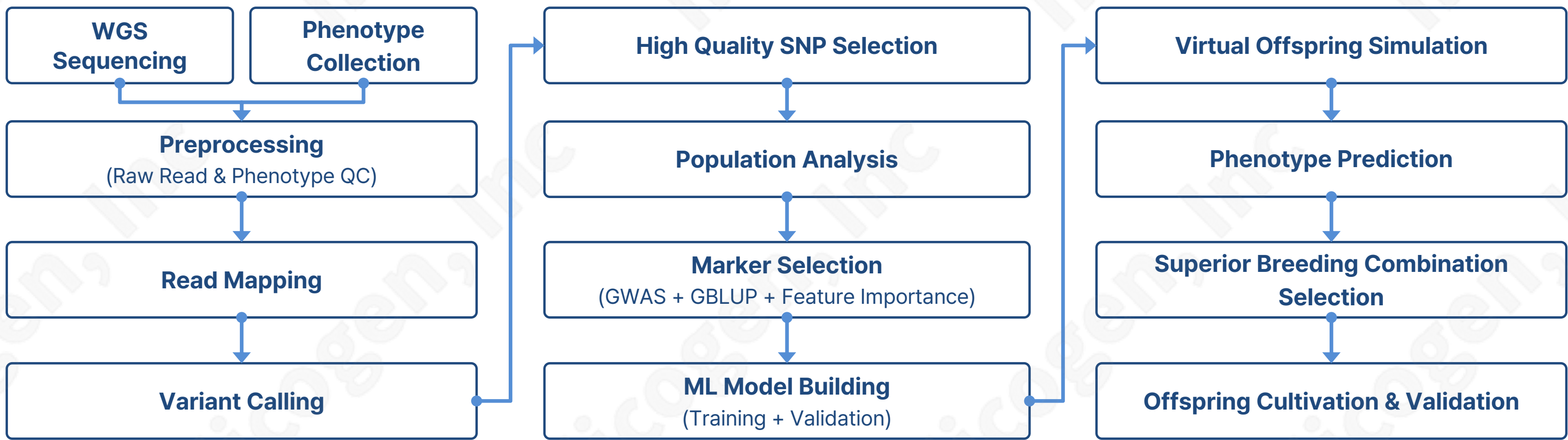


그림 1. 양배추 80 계통의 디지털육종 모식도

Results

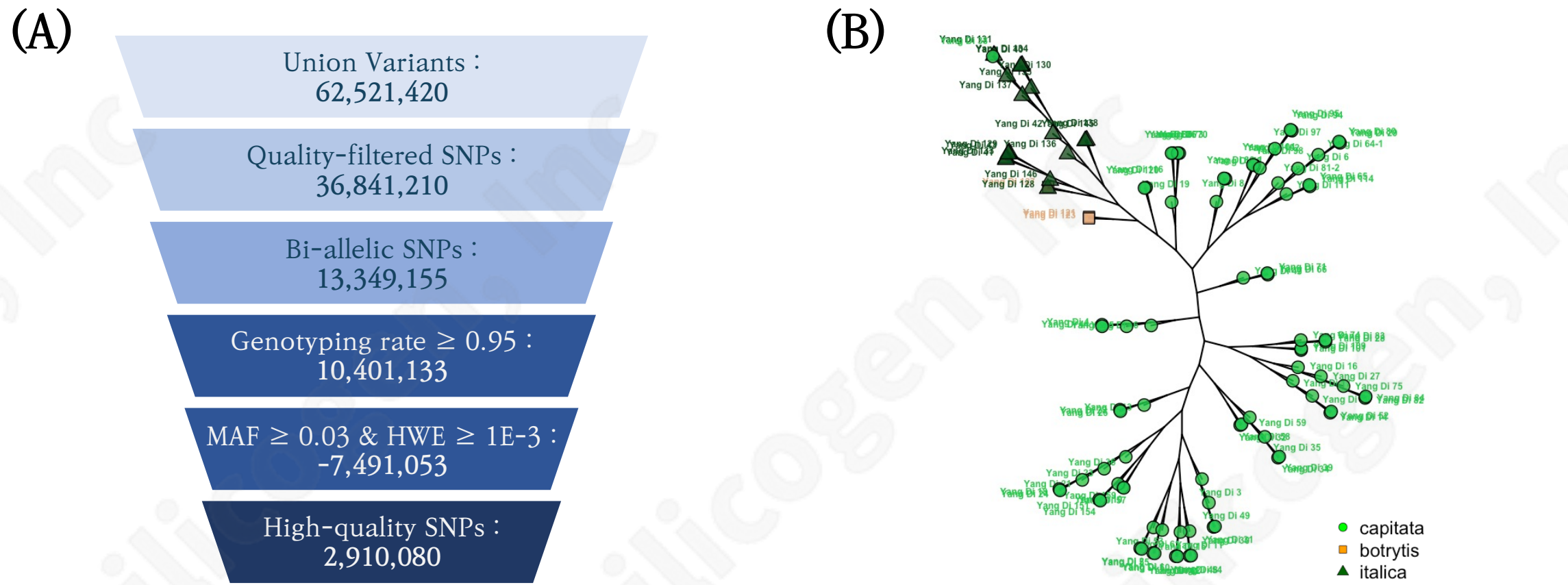


그림 2. 양배추의 유전적 다양성 (A) 선별 HQ SNPs; (B) SNPs를 이용한 양배추 계통 내 유사도 비교

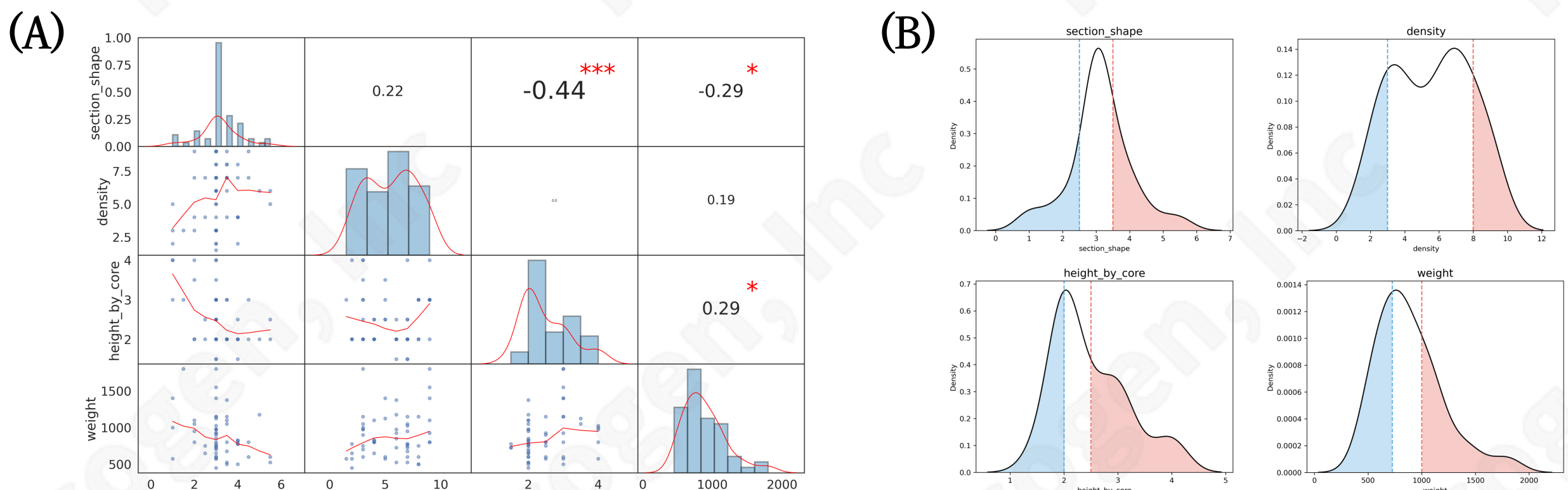


그림 3. 양배추의 표현형 분포 (A) 4개 경제형질 간 상관분석; (B) 우수/열등 개체 선별 기준치 및 표현형 값 분포

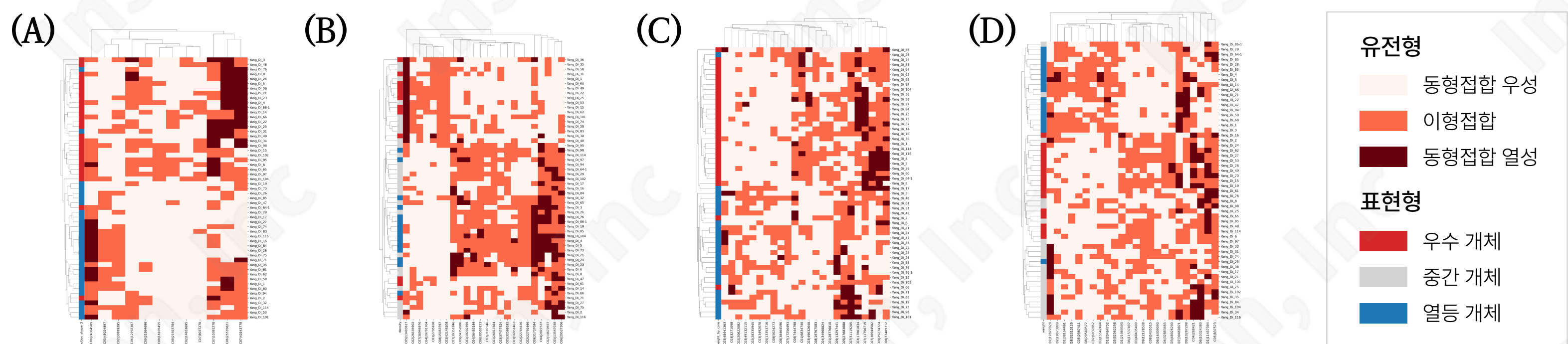


그림 4. 형질별 24개 선별 SNPs 및 Heatmap (A) 구형; (B) 밀도; (C) 고갱이 비율; (D) 구중

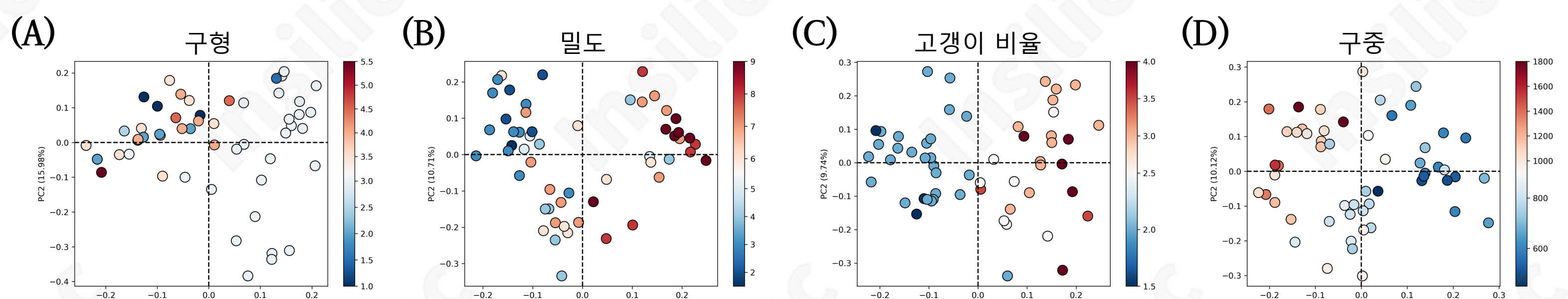


그림 5. PCA (Principal Component Analysis) 및 표현형 분포 (A) 구형; (B) 밀도; (C) 고갱이 비율; (D) 구중

표 1. 경제형질별 예측 알고리즘 및 예측 정확도

RF: Random Forest; PLS: Partial Least Squares; LOGI: Logistic Regression

Trait	Model	Accuracy	Sensitivity	Specificity	Balanced Accuracy
구형	RF	0.978	1	0.800	0.900
밀도	PLS	0.742	0.846	0.667	0.756
고갱이 비율	PLS	1	1	1	1
구중	LOGI	0.871	0.842	0.917	0.879

(A)

Sample	Section shape	Density	Height by core	Weight
Yang_Di_1	3.5	7	2	525
Yang_Di_3	3	1.5	3	575
Yang_Di_19	1	3	4	1,000
Yang_Di_116	2	7	2	950

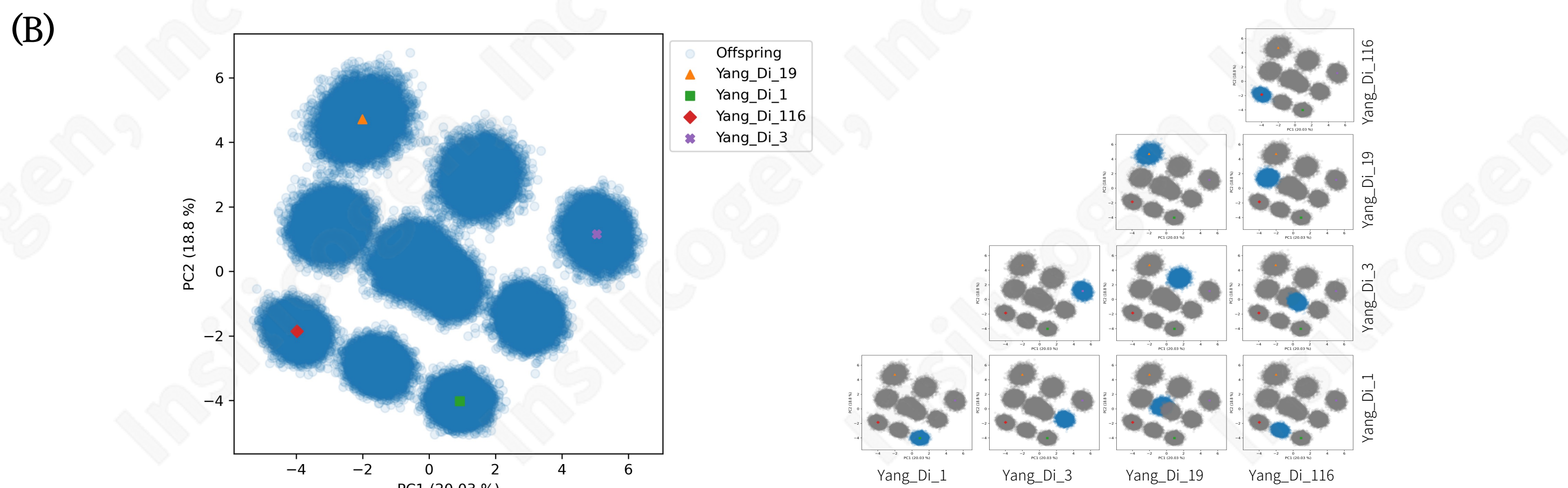


그림 6. *In silico* 방법을 통해 확보한 가상 자손의 유전적 다양성

(A) 시뮬레이션 유용성 확인을 위해 선별된 부모 개체의 표현형 정보; (B) 선별된 4개 부모 개체 유래 가상 자손의 유전적 다양성

표 2. 교배 및 재배를 통한 *in silico* 방법의 효용성 검증

(A) 교배 및 재배를 통한 실제 자손 표현형 측정; (B) 가상 자손의 유전형과 ML 모델을 통한 표현형 예측치; (C) 우수 표현형 발현 가계 확인 및 실측치와 예측치 사이의 비교를 통한 디지털육종의 효용성 확인

(A)

Family	# of samples	Section shape		Density		Height by core		Weight		Rank Sum
		Average	Rank	Average	Rank	Average	Rank	Average	Rank	
19	16	0.908	5	0.0328	5	3.917	16	2.833	1	27
20	38	0.987	12	0.0333	3	3.719	11	2.562	2	28
11	12	0.961	9	0.0334	2	3.778	14	2.350	5	30
6	4	0.975	10	0.0257	16	3.000	5	2.550	3	34
5	1	1.000	13	0.0304	10	3.000	4	2.200	8	35
13	7	1.133	18	0.0310	9	2.000	1	1.975	10	38
15	29	0.890	3	0.0314	8	4.087	18	1.817	14	43
8	6	1.165	19	0.0281	13	2.750	3	2.063	9	44
21	34	0.900	4	0.0299	11	4.400	19	1.856	12	46
10	5	1.096	17	0.0240	17	3.667	12	2.417	4	50
22	32	0.977	11	0.0269	14	3.419	8	1.221	17	50

(B)

Family	Combination	Complex trait	Section shape	Density	Height by core	Weight
19	Yang_Di_53 x Yang_Di_73	2.007	0.449	0.455	0.583	0.976
5	Yang_Di_48 x Yang_Di_53	1.867	0.111	0.689	0.571	0.971
6	Yang_Di_14 x Yang_Di_19	1.617	0.486	0.376	0.499	0.563
13	Yang_Di_14 x Yang_Di_49	1.587	0.182	0.634	0.479	0.644
20	Yang_Di_23 x Yang_Di_27	1.452	0.037	0.361	0.767	0.681
10	Yang_Di_32 x Yang_Di_8	1.283	0.219	0.340	0.732	0.246
22	Yang_Di_23 x Yang_Di_24	1.248	0.173	0.269	0.558	0.551
8	Yang_Di_14 x Yang_Di_32	1.149	0.101	0.384	0.757	0.138
11	Yang_Di_3 x Yang_Di_8	1.013	0.262	0.314	0.531	0.062
15 & 21	Yang_Di_3 x Yang_Di_5	0.861	0.178	0.279	0.540	0.000

(C)

	실제 자손(F1)	가상 자손	집단 간 차이
우수 가계	19, 20, 11, 6, 5,	19, 5, 6, 13, 20	11, 13
열등 가계	13, 15, 8, 21, 10, 22	10, 22, 8, 11, 15, 21	
정확도			81.82% (9 / 11)

Conclusion

본 연구는 양배추 계통에 대한 유전형 정보를 바탕으로 4가지 경제형질(구형, 밀도, 고갱이 비율, 구중)과 연관된 SNP 마커를 제시하며, 우수 개체 및 교배 조합을 정확하게 식별하고 교배 조합 작성 시 정확하고 신속한 의사결정을 가능하게 하는 데이터 기반 육종 실용성을 입증하였다. 즉, 총 96개 SNP 마커를 사용하면 고생산성의 양배추 개체 및 교배 조합을 선별할 수 있다는 것을 의미하며, 향후 양배추뿐만 아니라 다양한 원예작물의 주요 형질 예측 및 정밀 육종에도 폭넓게 활용될 수 있을 것으로 기대한다.

Acknowledgement

본 결과물은 농림축산식품부의 재원으로 농림식품기술기획평가원의 "첨단정밀육종활용고부가산업화 기술개발"사업의 지원을 받아 연구되었음(RS-2025-02216851).