농촌진흥청 | 보도자료 | 국민이 만든 대전환의 길

보도 시점 2026. 6. 11.(목) 11:00 / 6. 12.(금) 조간 배포 2026. 6. 11.(목) 09:00

# ‘누락’ 유전형 정보, ‘양자컴퓨팅 기반 기술’로 채워 넣는다

- 농촌진흥청, 유전형 정보 결측치 복원 기술 개발
- 기존 방식보다 유전형 복원 정확도 높아
- 슈퍼컴퓨팅-양자컴퓨팅 조합으로 농업 난제 풀어갈 것

농촌진흥청(청장 이승돈)은 농생명 핵심 빅데이터인 유전형 정보의 빠진 부분(결측치)을 더욱 정확하게 복원하는 양자컴퓨팅* 기반 기술인 ‘큐임퓨터(QuImputer)’를 개발했다.

* 양자컴퓨팅: 양자역학 원리를 이용해 많은 조합을 가진 최적화 문제를 새로운 방식으로 다루는 차세대 컴퓨팅 기술

생물별 유전체 특징을 담고 있는 유전형 정보는 유용 유전자 탐색, 디지털 육종, 농생명 인공지능 개발 등 다양한 농생명 연구의 핵심 기반 데이터다. 그러나 높은 분석 비용, 시료 확보의 어려움 등 이유로 일부 유전형 정보 누락 문제가 빈번하게 발생한다.

유전형 정보가 빠지면 어떤 개체가 중요한 유전적 특징을 가지고 있었는지 놓칠 우려가 있다. 예를 들어, 누락 유전형 정보에 고온에서도 잘 견디는 특징이 포함돼 있다면 해당 작물의 가치를 분석 과정에서 놓칠 수 있다.

기존에는 빠진 유전형 정보를 복원하기 위해 주변 단서와 유전체 전체의 유형(패턴)을 바탕으로 빈 곳에 들어갈 가능성이 높은 답을 추정하는 통계학 기반의 기술을 이용했다.

하지만, 이 방법도 정보가 빠진 부분이 길게 이어지거나 집단에서 드물게 나타나는 희귀 변이에서는 정확도가 떨어진다. 특히, 빠진 부분이 많을수록 가능한 유전형 조합의 수가 기하급수적으로 늘어나 기존 컴퓨터로는 문제를

해결하기 어려웠다.

이에 연구진은 기존 방식보다 유전형 복원 정확도를 높인 양자컴퓨팅 알고리즘을 세계 최초로 개발했다. 유전변이들이 함께 유전되는 경향, 변이가 집단에서 나타나는 빈도와 변이 사이의 거리를 양자컴퓨터로 계산해 가장 자연스러운 유전형을 찾아 넣는 것이다.

연구진은 먼저 농촌진흥청 초고성능컴퓨터인 나비스(NABIS) 2호기의 양자컴퓨터 시뮬레이터를 이용해 양자컴퓨터 알고리즘을 만들고 시험했다. 이후 미국 아이비엠(IBM)사가 제공하는 실제 양자컴퓨터를 이용해 검증하고, 제대로 작동함을 확인했다. 이 기술을 적용한 결과, 기존 방법보다 희귀 변이 분석에서 복원된 희귀 변이의 정밀도가 29.97%포인트 높게 나타났다.

농촌진흥청은 양자컴퓨팅 기술이 고도화되면 양자컴퓨팅 전용으로 개발해 디지털 육종과 유용 유전자 탐색 등 다양한 농업 난제를 해결할 수 있을 것으로 전망했다. 기존 유전형 복원 방법의 한계를 양자컴퓨터로 해결한 이 기술의 특허출원*을 완료했다.

* 양자컴퓨팅 기반 연관 불균형(LD) 해밀토니안 모델을 활용한 유전형 분석 장치 및 방법(출원번호: 10-2026-0097395)

농촌진흥청 슈퍼컴퓨팅센터 이태호 센터장은 “이 기술은 양자컴퓨팅 기반의 최적화 방식으로 복잡한 생명 빅데이터 문제를 풀어내는 새로운 접근 가능성을 찾은 사례”라며, “앞으로 유전형 정보를 다루는 모든 연구 현장에서 활용할 수 있을 것으로 기대된다.”라고 말했다.

붙임 1. 큐임퓨터(QuImputer) 개요
2. 주요 용어 설명


| 담당 부서 | 국립농업과학원<br>슈퍼컴퓨팅센터 | 책임자 | 센터장 | 이태호 (063-238-4558) |
|---|---|---|---|---|
| | | 담당자 | 연구사 | 김민우 (063-238-4600) |

## 붙임 1 큐임퓨터(QuImputer) 개요

### □ 연구 필요성 및 목적

○ **(필요성)** 유전형 정보는 유용 유전자 탐색, 디지털 육종 등 다양한 농생명 연구의 핵심 데이터이지만 높은 분석 비용, 시료 확보 어려움 등 다양한 이유로 **일부 유전형 정보가 누락되는 문제가 발생**

○ 통계적 추론을 수행하는 기존 보정 방법 등은 전반적으로 성능이 우수하지만 연속 결측에서는 관측 정보가 약해지고, **희귀 변이에서 major allele[1] 쪽으로 예측이 기울 수 있음**


그림 1. 기존 HMM 기반 통계적 추론 방법 모식도

☞ 통계 기반 희귀 변이 지역의 복원은 해당 샘플의 변이를 놓칠 가능성이 높음

○ **(목적)** 양자컴퓨터를 활용하여 결측된 유전형 데이터에서 희귀·저빈도 변이를 생물학적 정보를 기반으로 **복원하는 알고리즘 개발**

### □ 큐임퓨터(QuImputer)의 차별점

○ 큐임퓨터(QuImputer)는 하나의 틀(window) 안에 있는 결측 SNP 전체를 서로 연결된 LD 네트워크로 보고 동시에 최적화해 연관성을 고려함

○ 각 결측 SNP를 양자 큐비트 대응 변수로 바꾸고, 생물학적 정보(Allele Frequency·LD coupling 등)를 하나의 에너지 함수에 넣어 현재 관측된 데이터에 가장 부합한 유전형 조합을 최적해[2]로 선택

◆ QuImputer는 저커버리지 벼 유전체 데이터에서 희귀·저빈도 변이 복원과 희귀 ALT 누락(Type II 오류) 감소에 강점을 보인 양자컴퓨팅 기반 로컬 최적화 방법

◆ Beagle의 대체가 아닌, Beagle의 전역 결측 추정법과 QuImputer를 상호보완적으로 결합하는 hybrid **전략**이 결측된 유전체 데이터를 복원하는데 의미를 가짐

---

1) Allele: 유전학에서 '같은 유전자 위치에 존재하는 서로 다른 형태(변형)'를 뜻함

2) 가능한 모든 해 중에서 목적함수 값을 가장 좋게(최소 또는 최대) 만드는 해

## 붙임 2 주요 용어 설명

- **유전체 (Genome)**: 생물이 가진 모든 유전정보, 즉 DNA 염기서열 전체를 의미함
- **유전형 (Genotype)**: 특정 유전체 위치에서 한 개체가 가진 유전정보의 상태(예를 들어, 기준형인지 변이형인지가 유전형으로 표현)
- **결측 유전형**: 유전체 위치는 알려져 있지만 해당 샘플의 값이 비어 있는 상태
- **SNP (Single Nucleotide Polymorphism, 단일염기다형성)**: 유전체의 특정 위치에서 A, T, G, C 중 하나의 염기가 개체마다 다르게 나타나는 변이
- **유전변이**: 개체나 집단 사이에서 DNA 염기서열이 서로 다르게 나타나는 차이를 의미
- **희귀·저빈도 변이**: 전체 집단에서 낮은 빈도로 나타나는 유전변이
- **LD (Linkage Disequilibrium, 연관 불균형)**: 가까운 유전변이들이 함께 유전되는 경향. 어떤 변이가 나타날 때 주변 변이도 함께 나타나는 관계를 의미함
- **AF (Allele Frequency, 대립유전자 빈도)**: 특정 변이가 전체 집단에서 얼마나 자주 나타나는지를 나타내는 값
- **해밀토니안 모델 (Hamiltonian model)**: 가능한 여러 조합에 에너지 값을 매기고, 가장 낮은 에너지를 갖는 조합을 찾는 방식
- **양자컴퓨팅**: 양자역학 원리를 이용해 많은 조합을 가진 최적화 문제를 새로운 방식으로 다루는 차세대 컴퓨팅 기술
- **슈퍼컴퓨터**: 매우 많은 계산을 빠르게 수행하기 위해 여러 계산 자원을 병렬로 사용하는 고성능 컴퓨터
- **Beagle**: 유전형 데이터의 결측값을 통계적으로 복원하는 대표적인 기존 프로그램
- **HMM (Hidden Markov Model, 은닉 마르코프 모델)**: 주변 정보와 전체 패턴을 이용해 보이지 않는 값을 확률적으로 추정하는 통계 모델
- **Haplotype (일배체형)**: 한 염색체 위에서 여러 변이가 함께 나타나는 배열 패턴 또는 조합을 의미하며, 어떤 순서와 조합으로 함께 나타나는지 보여주는 정보