

농생명 오믹스데이터 통합 및 표준화

김도완, 이태호, 김창국, 설영주, 이동준, 오재현, 백정호, 이준아, 이흥로
국립농업과학원, 국립군산대학교

Challenges in Construction of Omics data integration, and its standardization

Do-Wan Kim*, Chang-Kug Kim, Young-Joo Seol, Dong-Jun Lee, Jae-Hyeon Oh, Jung-Ho Beak,
Juna Kim, Hong-Ro Lee and Tae-Ho Lee

Genomics Div., Department of Agricultural Bio-resource, National Academy of Agricultural Science,
RDA, Jeonju, Korea

Computer Information Engineering, Kunsan National University, Korea

E-mail : littlewinner@korea.kr

요 약

유전체 염기서열 분석비용이 크게 감소하면서 유전체 정보 생산이 본격화됨에 따라 시스템 생물학 기반의 통합 및 표준화된 오믹스 데이터베이스 구축이 필요하다. 이에 따라 현재 진행중인 연구 수행의 결과로 얻어진 차세대유전체서열(NGS) 및 전사체(transcriptome) 등의 대용량 정보를 수집하였고 이를 표준화 형식에 맞춰 농업생명공학정보센터(NABIC)에 등록하였다. 또한 농업생명자원 생물정보를 품목별, 개체별로 통합 저장소를 구축하였으며 농업생명자원 생물정보를 품목별, 개체별로 통합 저장소를 구축하였다. 농업생명공학정보센터 오믹스 정보등록시스템 서비스와의 연계 및 확충작업을 하기위해 시스템 기능 개선 및 유지보수 작업을 수행하였다.

ABSTRACT

We performed integration and standardization of the omics data related agriculture. To do this, we requires progressed computational methods and bioinformatics infrastructures for integration, standardization, mining, and analysis. It makes easier biological knowledge to find. we potentialize registration a row and processed data in NABIC (National Agricultural Biotechnology Information Center) and its processed analysis results were offered related researchers. And we also provided various analysis pipelines, NGS analysis (Reference assembly, RNA-seq), GWAS, Microbial community analysis. In addition, the our system was carried out based on the design and build the quality assurance in management omics information system and constructed the infrastructure for utilization of omics analyze system. We carried out major improvement quality of omics information system. First is Improvement quality of registration category for omics based information. Second is data processing and development platform for web UI about related omics data . Third is development of proprietary management information for omics registration database. Forth is management and development of the statistics module producers about omics data. Last is Improvement the standard upload/ download module for Large omics Registration information

키워드

omics, integration, database, NGS

1. 서 론

유전체 염기서열 분석비용이 크게 감소하면서 유전체 정보 생산이 본격화됨에 따라 시스템 생물학 기반의 통합 및 표준화된 오믹스 데이터베이스

구축이 필요하다. 해외의 경우 유전체 해독 프로젝트 수행에 따른 유전체 정보가 축적됨에 따라 다양한 유전체정보를 분석하여 미국, 유럽 및 일본 등은 이를 국가적으로 빅데이터 수준의 통합데이터베이스를 구축중에 있다. 우리나라의 경우 생명공학의 전반적인 연구는 선진국에 비해 다소 늦은 경향이 있으나 유전체, 전사체, 단백질,

대사체, 표현체, 흐름체등 일부 오믹스 데이터 확보 분야는 상당한 경쟁력을 갖고 있으며 통합 데이터 베이스 구축 및 활용단계는 진행단계에 있다. 현재 구축되어있는 정보센터의 확충 및 연계 시스템 구축으로 오믹스 정보 표준화를 통한 생물정보 통합 DB 구축 작업이 필요하다. 우리는 이 연구에서 표준화된 오믹스 정보 통합데이터베이스를 구축하였고, 데이터베이스 구축과 운영에 필요한 정보시스템의 기능개선 및 유지보수에 관련한 사항을 수행하였다.

II. 농업생명자원 오믹스 정보 통합데이터 베이스 구축

국가 유전체 사업이 확대됨에 따라 국가 농생명 정보에 대한 인프라 구축이 필요하다. 현재 농촌진흥청 농업생명공학정보센터(NABIC), 국가생명연구자원정보센터(KOBIC) 등에 농생물정보 데이터 베이스를 제공하는 정보센터가 구축되어 있으나 통합적인 정보 활용 및 사용자 맞춤형 정보 DB구축에는 미흡한 실정이다. 이에 따라 바이오정보센터 확충 및 연계시스템을 구축하여 각종 바이오 정보의 통합 및 활용의 일원화가 필요하고 이를 바탕으로 각종 유전자원의 구조 및 기능유전체 기반 활용사업을 확대해야한다.

현재 진행중인 국가 유전체 연구 수행의 결과

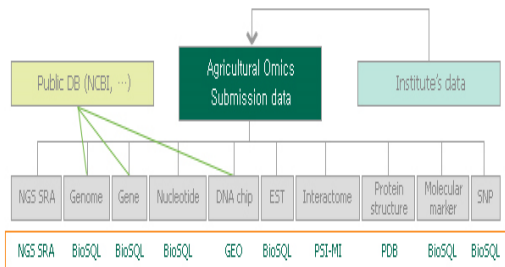


그림 1. 오믹스 표준 데이터 베이스

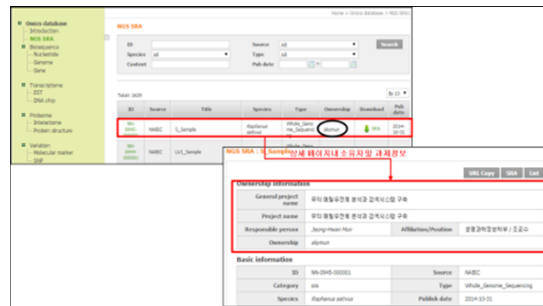
로 얻어진 차세대유전체서열(NGS) 및 전사체(transcriptome) 등의 대용량 정보를 수집하고 이를 표준화 형식에 맞춰 농업생명공학정보센터(NABIC)에 등록하였다. 다부처연구과제에 따라 발생하는 NGS, RNA-seq Transcriptome, Microarray 정보 데이터등을 국제표준화 형식에 맞춰 NGS SRA, Genome, Gene, DNA chip Genome 등의 분야별 오믹스 데이터로 변환 및 가공하였고 이에 의한 정규화 분석 및 유용 정보 해석 모델을 구축하였고 이 데이터들은 EBI SRA, Genbank 등의 포맷으로 활용할 수 있게 하였다.

또한 농업생명자원 생물정보를 품목별, 개체별로 통합 저장소를 구축하였다. 생물정보는 NGS SRA, protein structure, SNP 등 10개의 카테고리로 구

성되어 있고, 각종 바이오정보를 종별로 저장할 수 있게 구축 하였다.

III. 정보 시스템 기능 개선 및 유지보수

현재 구축되어진 농업생명공학정보센터 오믹스 정보등록시스템 서비스와의 연계 및 확충작업을 하기위해 시스템 기능 개선 및 유지보수 작업을 수행하였다. 데이터베이스 정보 등록절차 간소화 및 표준화를 위한 기능 개선과 DB 통계 기능 향상시키는 작업을 하였다. 농업생명공학정보센터의



등록절차의 과정 단계를 줄이고 사용자 편의에 중점을 두고 기능을 개선하였으며, 등록시 오믹스 정보의 표준화작업을 통해 정보의 정규화 분석 및 유용 정보 해석 모델을 구축하였다. 정보센터 내 유용 분자 마커 발굴 시스템, 유전자 구조 정보 유전체 브라우저를 이용한 가시화 기능 개선 및 비교 유전체 분석 시스템 기능 개선작업을 하였다. 농립축산식품 오믹스 정보를 기관 경상과제 및 차세대바이오그린21 사업 수행 결과 발생된 유전체 정보와의 정보 공유 및 통합 관리 체계 수립 하였으며 농생명 오믹스정보 등록 및 품질 관리 기능 갱신체계 구축하였다.

IV. 결 론

『포스트게놈 신사업육성을 위한 다부처 유전체 사업』 등 국가 유전체 사업이 확대됨에 따라 국가 농생명 정보에 대한 인프라 확충이 필요하다. 이에 따라 대량의 유전체 오믹스 관련정보를 체계적으로 관리하고 이를 분석 가공하는 활용체계를 구축하였다. 또한 다양한 유전체 정보의 통합 데이터베이스 구축을 위하여 표준 데이터 모델을 설계하였고 이를 기반으로 표준화된 입출력 인터페이스를 구축하여 일관된 표준 모델로 유전체 정보를 관리 할 수 있는 기반을 구축하였다. 또한 통합데이터베이스 구축과 운영을 위해 필요한 정보시스템의 기능개선 및 유지보수에 관련한 일련의 작업을 수행하였다.

본 연구는 향후의 농업생명산업의 한 단계 도약시키는 주춧돌이 될 것이다. 고부가가치화 기반기술 제공 하고 바이오 유전체 정보를 이용하여 동

식물 분야뿐만 아니라 동식물 품종개량 및 미생물 유래 고부가 신소재 발굴 등 바이오 산업 활성화를 촉진시킬 것이다. 또한 대규모 차세대 유전체 자료에 대한 과학적 품질관리 기반으로 선행적인 시스템 구축을 통해 범정부적 유전체 사업의 선도해 나갈 수 있을 것으로 기대한다.

참고문헌

- [1] <http://nabic.rda.go.kr/>
- [2] <http://nabic.rda.go.kr/nagp/>
- [3] <http://biosharing.org>
- [4] <http://www.obofoundry.org>