SFO Science of Fisheries & Oceanography
ISSN 2734-1011 (Print)
ISSN 2671-4310 (Online)
Vol. 31, No. 2, pp. 97-107 (2022)
https://doi.org/10.22714/SFO.2022.31.2.3

Original Research Article

대한민국의 환경유전자 데이터베이스 플랫폼 시스템 구조 제안

Suggestion of Environmental DNA Database Platform System Structure in Korea

박정환[1,*], 김건희[2]
Junghwan Park[1,*], Keonhee Kim[2]

[1](주)피제이팩토리
[2]건국대학교 휴먼앤에코케어센터
[1]PJ Factory Ltd., Seoul 05115, Republic of Korea
[2]Human and Ecocare Center, Konkuk University, Seoul 05029, Republic of Korea

* Correspondence to Junghwan Park
E-mail: park@pjfactory.com

Received March 20, 2023
Revised March 21, 2023
Accepted March 21, 2023

Abstract : Recent efforts in developing independent databases for Environmental DNA (eDNA) are derived from the need for localized sequence data sets to identify indigenous species and spatial data visualization of species composition over time. However, the uncertainty of the eDNA result (i.e., feces of birds carrying DNA from other environments) requires species composition data through time and resource-consuming alpha classification in the field for comparative analysis. In this paper, we suggest an eDNA-Citizen science database platform with which citizen science projects can be organized and prepared for investigating the species composition in the area near them with the species list from the eDNA data in the platform. The platform design aims to induce the complementary interaction between the eDNA researchers and citizen scientists to create species composition data in a set of eDNA analyses and alpha classification by scientists and citizens, respectively.

Keywords : eDNA, Database, Platform structure, Citizen science, Ecological literature

서 론

환경유전자 분석기법은 특정지역의 생물상을 적은 인원으로 빠르게 파악할 수 있다는 장점이 있다. 이로 인해 환경유전자를 기반으로 하는 데이터베이스들이 기존의 염기서열 데이터베이스가 존재함에도 독립적으로 여러 지역 및 국가에서 산발적으로 구축되고 있다. 이와 같이 환경유전자 데이터베이스가 지엽적이고 산발적으로 구축되는 이유는 1) 해당 데이터베이스의 연구가 특정 지역 또는 특정 생태계의 생물상 또는 특정 종과 같이 지리적인 범위를 가지며, 2) 특정 지역에서만 서식하는 고유종 탐침을 위해서는 대상 지역에서 발견된 생물들의 유전자 염기서열 데이터베이스가 필요하기 때문이다.

그러나 환경유전자 분석기법은 철새의 배설물 등을 통한 외부 유전자의 유입 또는 탐침(Probe)의 잘못된 설계로 인해 존재하는 유전자를 감지하지 못하는 오류를 내포한다. 이를 개선하기 위해 환경유전자 데이터베이스와 더불어 그 결과를 교차검증할 수 있는 현장 관측데이터가 필요하다. 즉, 현장에서 알파 분류보다 경제적이고 효율적인 환경유전자 분석기법을 보다 정교하게 하기 위해서는 알파 분류를 통한 현장조사가 동시에 이루어져야 한다는 딜레마가 발생한다.

시민과학은 생태 분야에서 많은 인력과 시간이 소요되는 생물상 조사를 시민집단과의 협업을 통해 성공적으로 수행한 선례들을 가지고 있다. 연구자들만으로 조사가 불가능한 광범위한 지역으로부터 방대한 양의 데이터를 짧은 시간 내에 수집할 수 있다는 점은 시민과학의 가장 큰 강점이다. 하지만, 정규 훈련을 거치지 않은 시민들에 의해 생성된 데이터는 신뢰성 문제를 항상 내포한다. 따라서, 시민과학은 조류(鳥類)와 같이 역사가 오래되고 나름의 지식을 습득한 애호가 집단을 중심으로 특정 생물군에 제한적으로 이루어지거나, 샘플채집과 같이 별다른 지식이 필요 없는 작업으로 참여가 국한되어왔다. 그로 인해 연구 집단은 시민과학을 통한 유연한 연구를 설계하기 어려웠으며, 시민들은 시민과학 활동을 통해 새로운 지식을 얻거나 연구의 주요한 역할을 수행할 수 없다는 문제를 지니고 있다.

환경유전자 분석은 시민과학자들에게 짧은 시간 내 학습 가

Table 1. Sharing of gene database between INSDC members (https://www.insdc.org)

Data type	DDBJ Center	EMBL-EBI	NCBI
Next generation read	Sequence Read Archive	European Nucelotide Archive (ENA)	Sequence Read Archive
Sequence Read Archive	Trace Archive		Trace Archive
Annotated sequence	DDBJ		GenBank
Samples	Biosample		Biosample
Studies	Bioproject		Bioproject

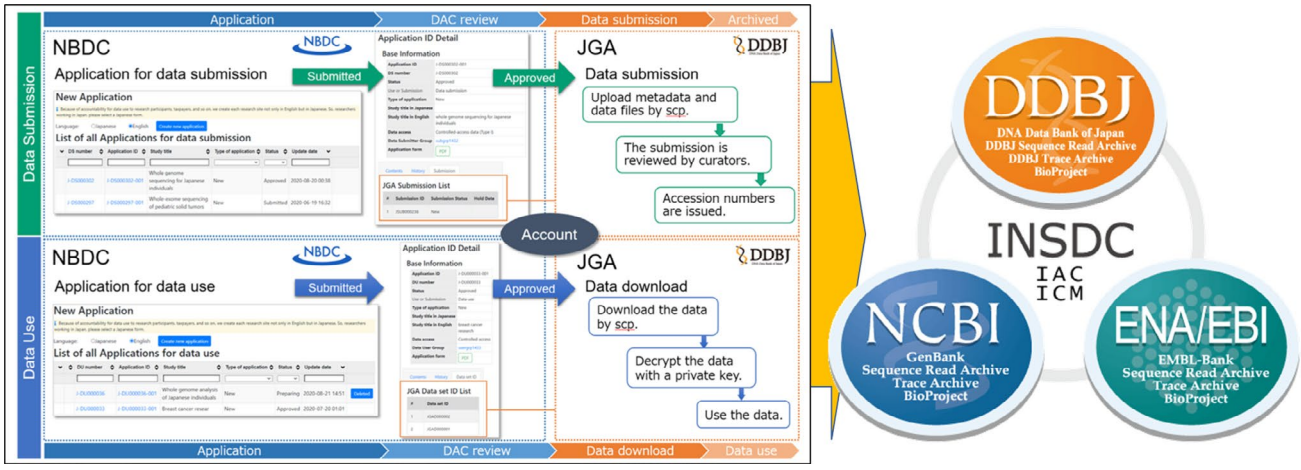

Fig. 1. Submission and application process of genetic information in DDBJ (Fukuda *et al*., 2021).

능한 생물 종 리스트를 제시할 수 있다. 이는 시민과학자들이 조사 대상 종과 조사 지역을 명확하게 하여 현장조사에 필요한 지식 습득과 조사를 계획, 실행할 수 있게 한다. 시민과학자들의 현장조사를 통해 얻어진 확인 종 리스트는 환경유전자 기반 종 리스트와 교차 분석이 가능하며, 이를 통해 환경유전자 분석의 정확도 및 개선사항을 파악할 수 있게 된다. 이와 같이 연구자는 환경유전자 분석을 통해 명확한 조사 대상과 지역을 제시하고, 시민과학자는 이를 기반으로 현장조사를 실시해 서식 종 리스트를 작성하고 공유함으로서 두 집단 간의 상보작용이 성립되게 된다.

본 논문은 이와 같은 상보작용을 기반으로 하는 '환경유전자-시민과학 데이터베이스 플랫폼'을 제안한다. 이를 통해 기존의 환경유전자 데이터베이스 모델을 개선하는 한편, 시민이 연구에 주요한 역할을 수행하는 바람직한 시민과학 활동의 기반을 마련하고자 한다. 이를 통해, 두 집단의 활동으로부터 도출된 연구결과물들이 데이터베이스에 축적될 뿐 아니라 시민들의 생태소양(Ecological Literacy)을 함양하는 다양한 시민과학 프로젝트가 자발적으로 조직되고 실행될 수 있는 플랫폼 구축을 목표로 한다.

국제 환경유전자 데이터베이스 구축 현황

현재 유전자 염기서열 데이터베이스는 크게 3개로 구분할 수 있다. 미국의 NCBI (National Center for Biotechnology Information), 유럽의 EMBL-EBI (European Molecular Biology Laboratory-European Bioinformatics Institute) 및 ENA (European Nucleotide Archive)의 염기서열 데이터베이스, 일본의 DDBJ (DNA Data Bank of Japan)는 전 세계 염기서열 정보가 모이는 주요 염기서열 데이터베이스로써 매우 다양한 유전자 염기서열 정보를 저장하고 있다. 미국, 유럽, 일본의 각 염기서열 데이터베이스는 INSDC (International Nucleotide Sequence Database Collaboration)의 회원으로써 유전자 염기서열 데이터를 수집하고 제공하기 위해 서로 자료를 공유하고 있으며 (Table 1, Fig. 1), 이를 통해 전 세계의 통합된 염기서열 데이터베이스(INSD: International Nucleotide Sequence Database)를 구축하고 있다(Cochrane *et al*., 2016).

NCBI, EMBL-EBI, DDBJ의 염기서열 데이터베이스는 별도의 환경유전자만을 분리하지 않고 전체 데이터베이스에 포함하여 보유하고 있다. 다만 NCBI와 DDBJ는 각 데이터의 성격에 따라서 'Sequence Read Archive' 또는 'Biosample'로 구

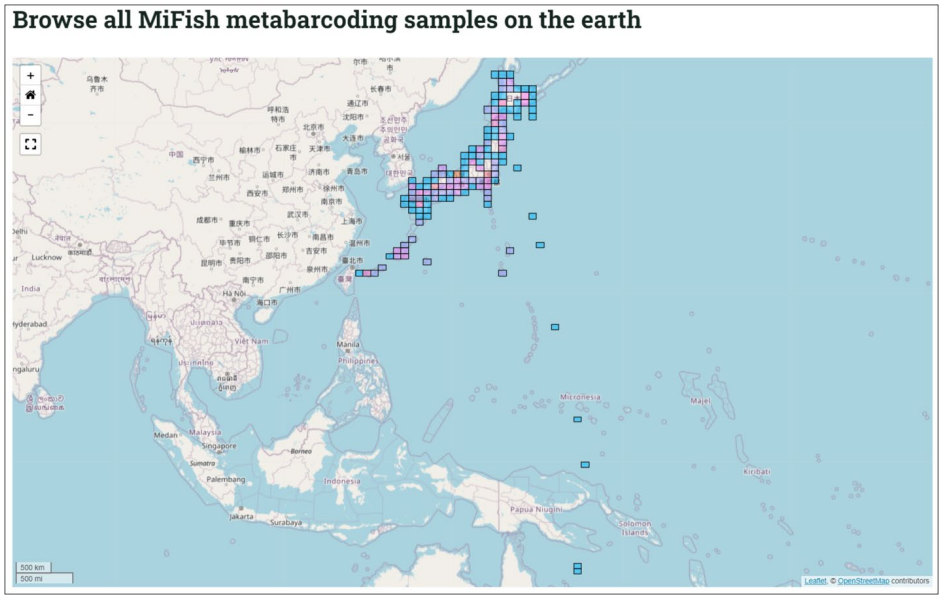

Fig. 2. Fishes eDNA metabarcoding database (https://db.anemone.bio).

분하고 있다. 각 유전자 데이터베이스에서 환경유전자는 'environmental sample sequence' 또는 다양한 환경에서 분석된 'Metagenome'으로 표시되어 있으며 NCBI와 EMBL-EBI는 BLAST 분석을 위해 각각의 염기서열 수준까지 정보를 제공하는 반면에 DDBJ는 메타데이터 라이브러리 수준의 정보를 저장하고 이를 NCBI와 EMBL-EBI에 제공하고 있다. NCBI와 EMBL-EBI는 BLAST 분석 수준에서 환경유전자 염기서열 데이터베이스를 활용하고 있으나 DDBJ는 BLAST 분석을 제공하지 않는다.

현재 세계적으로 환경유전자만을 활용하여 데이터베이스를 구축한 사례는 매우 드물지만 최근 매우 활발하게 데이터베이스 구축이 이루어지고 있다(Büdel *et al*., 2014; Davron *et al*., 2022; Petrosyan *et al*., 2023). 세계 생물다양성정보기구(GBIF: Global Biodiversity Information Facility)에서는 덴마크 지역의 곰팡이(Fungi) eDNA 유전자 데이터베이스, Savu Sea-Indonesia의 해양생물 eDNA 유전자, 호주 인근의 어류 eDNA 유전자 데이터베이스를 구축하였으며(Frøslev and Ejrnæs, 2018; Atlas of Living Australia, 2021; Anggoro, 2022), 이를 통해 eDNA 기반으로 분석된 생물다양성 유전자 데이터베이스와 플랫폼 구축체계를 정립하였다(Andersson, 2021). 가장 최근에 발표된 환경유전자 데이터베이스는 NYK Line (Nippon Yusen Kaisha)과 토호쿠 대학, 미나미산리쿠(Minamisanriku) 마을, 어스워치 재팬(Earthwatch Japan)에서 공동으로 진행한 담수와 해양의 어류 메타데이터베이스(Meta database)로써 플랫폼 회원 가입자들을 대상으로 어류 메타바코딩(Meta barcoding) 시료 분석 결과를 열람할 수 있도록 되어 있다(Fig. 2).

그외에도 GBWG (Genomic Biodiversity Interest / Working Group)에서 environmental sample과 eDNA 데이터베이스에 대하여 데이터를 공유하기 위한 데이터 표준 및 사용 가이드를 개발하고 있으며(https://www.tdwg.org/community/gbwg/enviro/), South California의 Coastal Water Research Project (https://www.sccwrp.org/)에서는 미국 전역에 분포하는 10개 하구지역의 eDNA를 분석하여 메타바코딩 분석 결과를 지속적으로 축적하고 있다.

국내 환경유전자 데이터베이스 현황

국내에서도 한반도의 생물다양성을 파악하기 위해 다양한 생물들의 데이터베이스를 보유하고 있다(Table 2). 하지만 대부분 생물의 형태적 정보를 기반으로 데이터베이스가 구축되어 있으며 일부 정부출연기관 및 사업단에서 형태적 정보가 존재하는 생물들을 대상으로 유전자 데이터베이스 구축을 시도하였다.

과학기술부에서는 생물자원 확보 지원 관리를 통한 국가생물자원센터 구축의 일환으로써 식물유전자원 정보은행구축 연구가 수행되었으며 해양수산부에서는 국가해양생물유전체사업단(NMGP: National Marine Organism Genome Project)을 통해 해양 수산생물 유전체정보 활용(메타게놈)에 대한 연구를 수행하였다. 또한 국립생물자원관에서는 다양한 생물들의 형태적 정보뿐만 아니라 생태적 정보와 유전자 염기서열 정보까지 제공하는 '한반도의 생물다양성' 홈페이지를 운영하고 있다. 그

Table 2. Biodiversity information and DNA database in Korea (Park *et al*., 2005)

Institute	URL	Contents
NIBR (National Institute of Biological Resource)	https://species.nibr.go.kr/index.do	Morphological identification key and ecological information of all taxa
NIFS (National Institute of Fisheries Science)	https://www.nifs.go.kr/frcenter/	Fisheries specimen and genetic information
National Science Museum	http://www.nsm.go.kr	Morphological identification key of all taxa
KCTC (Korean Collection for Type Culture)	https://kctc.kribb.re.kr	Type cultrue of microorganism
BRIC (Biological Research Information Center)	http://bric.postech.ac.kr	Statistical information of all taxa

외에도 수산생명자원정보센터의 경우에는 현재 약 45,000개의 수산동물(어류 및 갑각류 생물 등)의 유전정보를 보유하고 있다.

하지만 현재까지 대한민국을 대표하는 유전자 데이터베이스는 존재하지 않으며, 각 정부출연기관에서 자체적으로 보유하고 있는 데이터베이스 또한 하나로 통합되지 못하는 실정이다. 개인회사로는 유일하게 천랩(현재 CJ 바이오사이언스)에서 유전자 데이터베이스를 구축하였으나 그 범위는 장내 미생물(Gut microbiota)을 중심으로 하는 미생물에 한정되어 다른 생물분류군까지 확장하지 못하는 한계를 지니고 있다.

더욱이 환경유전자 데이터베이스는 국내에 전무한 실정이며, 과학단체 및 정부출연기관에서도 데이터베이스를 구축하지 못한 실정이다. 그럼에도 국내에서 환경유전자를 이용한 생물다양성 분석은 활발하게 수행되고 있고, 이러한 분석 결과는 매우 다양한 생물 분야에서 빠르게 증가하고 있다. 하지만 이러한 환경유전자 연구의 결과물들은 현재 연구자 개인이 보유하고 있는 수준이기 때문에 이를 통합하기 위한 환경유전자 데이터베이스 플랫폼의 필요성이 매우 높다.

시민과학 운영을 고려한 대한민국형 환경유전자 데이터베이스 플랫폼 설계

환경유전자만을 위한 독립된 데이터베이스 시스템 구축은 개인 연구자들이 가지고 있는 유전자 데이터를 한 곳에 모으고 서로 공유 한다는 점에서 기존의 유전자 데이터베이스 시스템과 동일한 기능을 수행한다. 따라서, 이를 위한 데이터 입력, 저장, 검색, 다운로드 등의 기능은 두 시스템 모두 매우 유사한 구조와 사용자 환경을 가지게 될 것이다. 하지만 환경유전자를 통한 생물군집 분석은 포식자의 배설물, 상류로부터의 세포조직의 유입 등으로 인한 오염으로 조사 지역 외 생물 종 정보를 포함할 수 있다. 또한 환경에 존재하는 유전자의 농도가 너무 낮아 탐색이 되지 않거나 탐침(primer) 설계가 잘못되어 해당 종을 탐색할 수 없을 수도 있다. 이러한 환경유전자 연구가 가지는 고유의 문제점으로 인해 그 결과의 검증을 위해서는 조사지의 현장조사가 병행 되어야 하며, 그 결과를 환경유전자 분석 결과와 비교 가능하여야 한다(Fig. 3).

환경유전자 데이터베이스를 기반으로 하는 현장조사자는 연구자 그룹 외에도 시민과학자, 정규, 비정규, 부정규 환경관련 교육자, 정책결정자, 어업종사자와 같은 다양한 시민그룹으로 확장될 수 있다. 그리고, 이를 가능하게 하는 것은 환경유전자 분석데이터이다.

다양한 분석 장비에 접근이 어려운 시민의 연구 참여는 관찰을 통한 조사에 한정될 수 밖에 없다. 그러나 관찰을 통한 연구 또한 오랜 기간의 전문적인 훈련을 필요로 한다. 따라서, 조류관찰(Bird Watching)과 같이 오랜 역사를 통해 준 전문가 수준의 애호가집단이 형성된 극히 일부 분야에서만 일반시민의 관찰을 통한 연구 참여가 이루어 졌다. 따라서, 대부분의 시민과학은 전문지식을 필요로 하지 않는 샘플의 채집, 또는 사진촬영 등의 형태로 연구에 참여하고 있으며, 그 과정 중 시민이 얻을 수 있는 과학적 지식이 미미할 뿐 아니라 연구의 주요한 역할로 부터도 배제되어 왔다.

하지만 환경유전자 분석은 단기간의 훈련으로 조사가 가능하도록 조사 대상 종을 좁혀주는 한편 조사지를 명확하게 제시한다. 따라서, 환경유전자 분석 결과와 조사지의 데이터베이스를 구축하여 시민과학자의 접근을 가능하게 한다면, 다양한 집단이 스스로 조사지를 선정하고, 사전학습을 통해 필요한 지식을 습득하는 한편, 주체적으로 현장조사를 계획하고 실시할 수 있을 것이다. 이를 통해 시민과학자들로부터 도출된 현장 관찰종 리스트는 다시 환경유전자 데이터베이스에 입력되어 동일지역의 환경유전자를 분석한 연구자에게 교차분석이 가능한 중요한 자료를 제공하게 될 것이다.

이와 같이 연구자와 시민과학자가 함께 활용하는 환경유전자 데이터베이스는 단일 사용집단(연구자)을 대응하는, '서비스' 성향이 강한 기존의 유전자 염기서열 데이터베이스 시스템과 달리 복수의 사용자집단을 대응하는 '플랫폼'의 양상을 띠

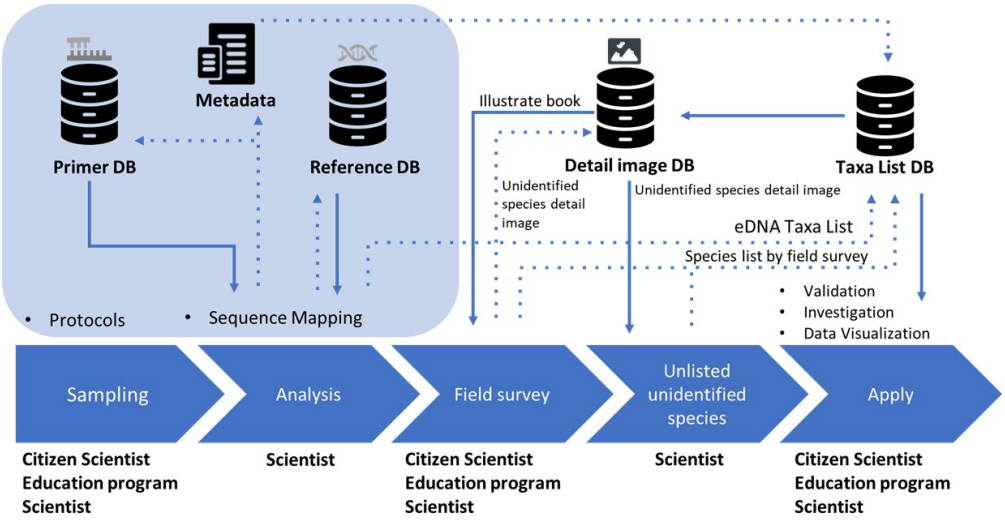

Fig. 3. Diagram of Korea's environmental genetic database platform enabling collaboration between scientists and citizens.

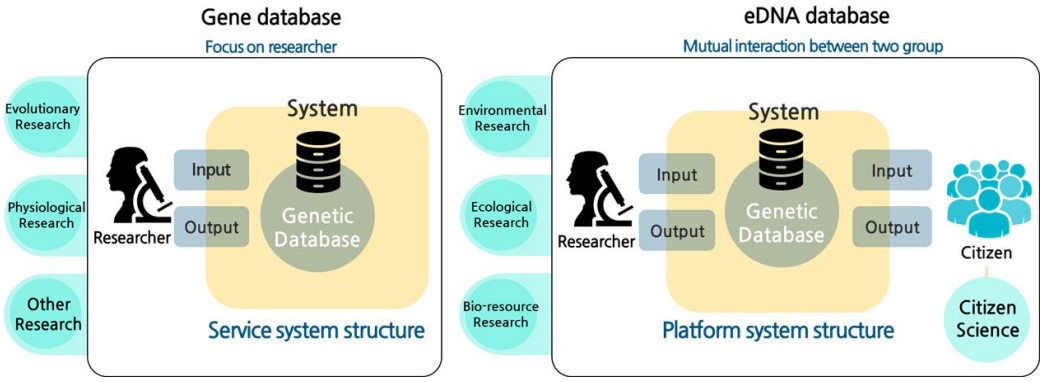

Fig. 4. Difference of database system between DNA barcoding and environmental DNA meta-barcoding.

고 있다. '서비스'와 달리 '플랫폼'은 서로 다른 두 집단이 상이한 목적을 위해 동일한 시스템 상에서 활동하며, 각각의 결과물이 상대집단의 요구사항을 충족시키는 선순환 구조를 형성한다는 것이다. 따라서 '플랫폼'는 그 요구사항 및 구조가 '서비스'와 비교 시 매우 복잡하다.

대한민국형 환경유전자 데이터베이스 플랫폼 구축 시 고려사항

환경유전자 데이터베이스 시스템을 구축함에 있어서 가장 중요한 고려사항은 '데이터베이스의 사용자를 어떻게 규정하고 분류할 것인가?'이다. 이는 해당 데이터베이스가 '서비스'인지, 또는 '플랫폼'인지를 구분 하는 주요 인자이기 때문이다. '서비스'와 '플랫폼'은 해당 시스템이 대응하는 사용자가 단일 집단인지, 서로 상응하는 복수 집단인지로 결정된다. 즉, '클라우드 저장소'의 경우, 저장하는 데이터를 생성하는 사용자와 저장된 데이터를 열람하는 사용자가 동일한 개인 또는 집단이므로 '서비스'로 분류된다. 반면 '소셜 미디어'의 경우, '콘텐츠 생산자'와 '시청자'라는 서로 다른 두 집단이 동일한 서비스를 활용하며 서로 대응하게 되므로 '플랫폼'을 성립한다.

'서비스'의 구축은 사용자의 요구사항을 파악하여 구축하는 상대적으로 단순한 형태이나, 플랫폼의 경우 상응하는 두 집단의 관계를 정의하고, 서로 다른 요구사항을 파악하는 한편, 기능을 통해 두 집단의 소통이 이루어져야 하므로 보다 복잡한 양상을 띠게 된다. 플랫폼은 복수의 집단, 즉 '집단 A'와 '집단 B'의 요구사항을 모두 충족시켜야 하며, 보다 근본적으로는 두 집단이 해당 플랫폼을 사용해야 하는 명확한 이유가 존재하여야 한다.

본 연구에서 제시하는 대한민국형 환경유전자 데이터베이스는 플랫폼의 형태로 연구자 이외에도 시민과학자, 교육자, 환경 NPO들이 함께 활용하는 것을 전제로 한다. 연구자집단과 시민

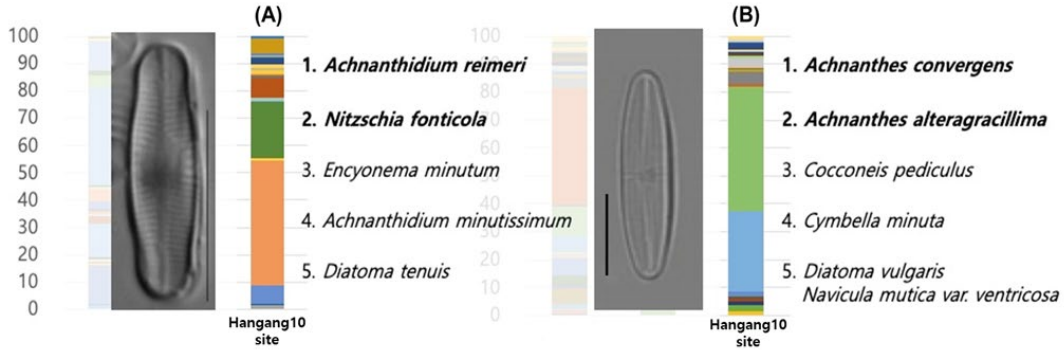

Fig. 5. Identification error caused by deficiency of DNA sequence data. (A) Analysis by eDNA, (B) Analysis by microscope.

과학자로 대표되는 그 외 집단은 각각의 결과물을 플랫폼에 제공하여 상대집단의 연구 활동에 필요한 데이터를 제공하는 것을 목표로 한다(Fig. 4). 즉, 연구자가 입력한 환경유전자 데이터로부터 시민과학자는 해당 지역에서 관찰하여야 할 종의 리스트를 얻게 되며, 현장 관찰을 통해 작성된 관찰 종 리스트는 연구자에게 환경유전자의 primer 설계 및 민감도의 검증 자료로서 제공되게 된다(Fig. 7).

1. 고유종과 외래종을 포함하는 국내 서식 종의 환경유전자 데이터베이스

현재 NCBI, EMBL-EMI, DDBJ는 기본적으로 각 국가에 서식하는 생물들의 유전자 정보를 기반으로 데이터베이스를 구성하였으며, 그 외 다른 국가에서 발견된 생물들의 유전자 정보는 해당 국가의 연구자들이 직접 유전자 정보를 등록하여 데이터베이스를 구축한다. 특히, 논문 투고 시 유전자 데이터베이스의 등록번호가 반드시 필요하기 때문에 해외 유전자 데이터베이스의 유전자 정보 등록은 강제성이 있다. 이러한 이유로 논문으로 발표되지 않은 국내 고유종 및 멸종위기종과 같이 개체수가 매우 적은 생물들의 유전자 정보는 해당 데이터베이스에 존재하지 않으며, 해외 유전자 데이터베이스를 기반으로 국내 생물의 종 다양성을 분석하는 경우 국내 서식 종이 유전자 염기서열과 유사한 다른 종으로 분류될 위험이 존재한다.

식물플랑크톤 돌말류 중에서 국내 하천 수생태계에서 자주 발견되는 *Achnanthes convergens*는 국내에서 주로 발견되지만 일본과 해외에서는 거의 발견되지 않기 때문에 국제 유전자 데이터베이스에 유전자 정보가 존재하지 않는다. 이로 인해 국내 하천에서 채집된 생물막 시료를 메타바코딩 분석하였을 때, *A. convergens*가 우점하는 시료에서 *Achnanthes reimeri*가 우점하는 것으로 나타나게 된다(Fig. 5). *A. remeri*는 *A. convergens*와 형태적으로 다르고 국내에서 한 번도 발견되지 않은 해외 서식 종이지만 *A. convergens*와 유전자 염기서열이 비슷하기 때문에 이러한 동정 오류가 발생하게 된다. *A. convergens* 이외에도 국내 하천 수생태계에서 주로 발견되는 *Achnanthes*속, *Fragilaria*속 분류군의 일부 종들의 유전자 정보가 해외 데이터베이스에 존재하지 않기 때문에 eDNA 기반의 군집분석에서 해당 생물은 군집에서 누락된다(Fig. 6).

국내에서 주로 발견되는 많은 생물들의 유전자 정보를 포함하지 않는 해외 데이터베이스를 기반으로 수행되는 환경유전자 분석은 전 세계 보편종에 대해서는 어느 정도의 신뢰도와 정확도를 보장할 수 있으나 국내 고유종 및 멸종위기종을 그 대상으로 하는 경우 신뢰도와 정확도에 명확한 한계가 존재한다.

최근에는 NCBI-Genbank 데이터베이스에 국내 연구자들이 유전자 정보를 등록하며 국내 고유종 및 멸종위기종에 대한 유전자 정보가 포함되었으나 주로 세균 등과 같은 미생물 분류군이 주를 이루었으며 곤충류의 유전자 정보와 식물체의 유전자 정보가 뒤를 이었다. 국내에서도 고유종 및 멸종위기종에 대한 유전자 데이터베이스를 구축하고자 활발하게 연구를 수행하였다. 국립생물자원관에서는 한반도에 서식하는 자생 생물종 조사연구를 통해 약 18,000종의 신종 및 국내 미기록종을 발견하고 120만 점의 생물자원을 확보하였으며 야생생물자원 DNA 바코드 표준 데이터베이스와 '야생생물 통합 유전정보 시스템(WIGIS)'을 구축하였다(Choi *et al*., 2011). 이렇게 구축된 국내 고유종의 유전자 데이터베이스는 현재 국립생물자원관의 "한반도의 생물다양성" 홈페이지 플랫폼에서 BLAST 분석 서비스 및 유전자 바코드 분석에 활용되는 Primer와 유전자 염기서열 라이브러리 정보를 제공하는 수준으로 활용되고 있다(https://species.nibr.go.kr). 이러한 국내 생물의 유전자 바코드 데이터베이스는 환경유전자 데이터를 분석하기 위한 훌륭한 reference로써 환경유전자 데이터베이스 시스템(Fig. 9)의 reference database로 활용할 수 있다.

2. 환경유전자 검침종의 현장 확인 시스템

환경유전자는 다양한 환경에 존재하는 유전물질을 의미하는 것으로 그 유래가 매우 다양하다(Kim *et al*., 2021) (Fig. 7). 가

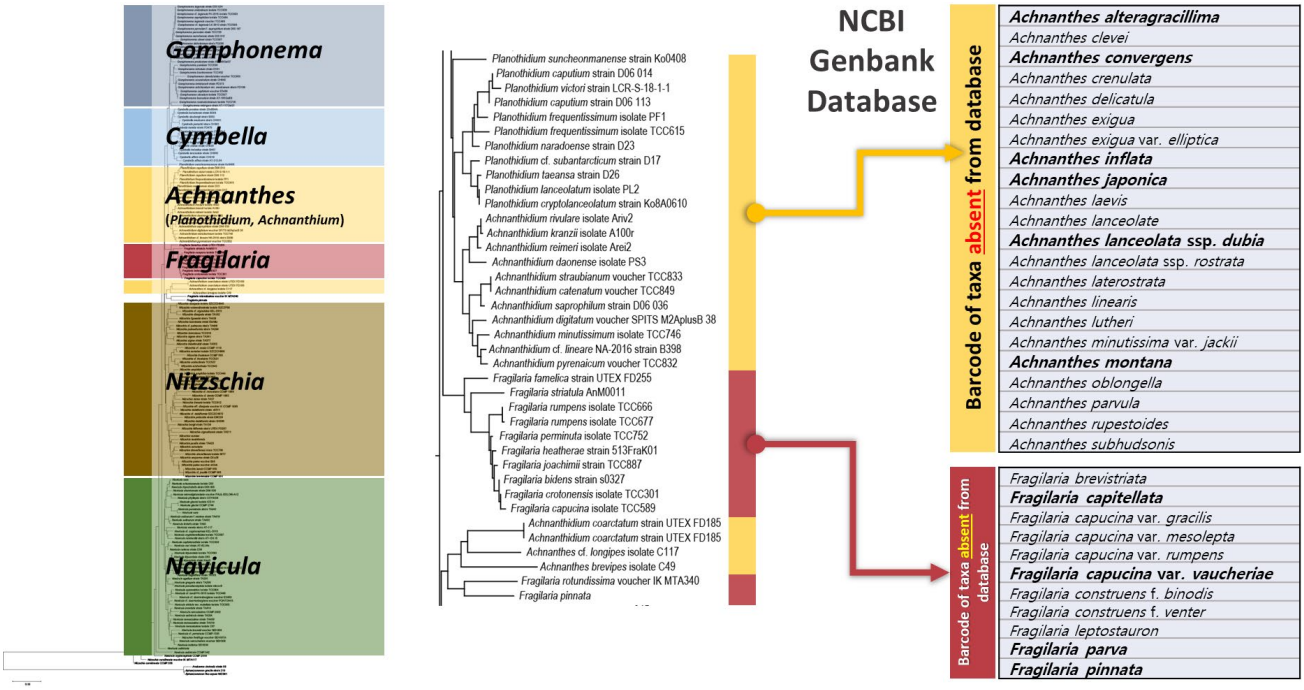

Fig. 6. Deficiency of diatom species DNA sequence data in Genbank. Species name with colored red means absent of DNA barcode data (Kim *et al*., 2019).

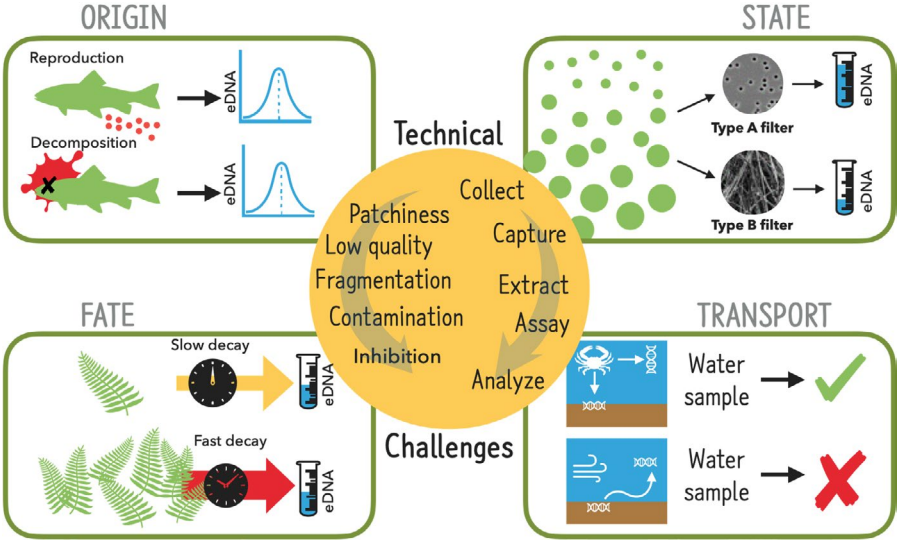

Fig. 7. Ecology of eDNA (Herder *et al*., 2014).

장 기본적으로 생물의 성장 과정에서 개체가 탈피과정을 거치며 생물의 표피에서 분리됨에 따라 환경으로 유입될 수 있으며, 산란과정을 통해 대상 생물의 체세포가 대량으로 환경에 유입될 수 있다. 이외에도 포식성 생물의 배설물을 통해 먹이생물의 유전물질이 환경에 유입될 수 있다.

생물로부터 유래된 환경유전자는 해당 생물의 존재를 파악할 수 있는 단서로써 활용할 수 있기 때문에 대상 생물을 직접 채집하지 않고 존재 여부를 파악할 수 있는 매우 큰 장점으로 작용한다. 하지만 반대로 이는 매우 큰 단점으로 작용할 수 있다. 대상 생물의 유전물질이 유래된 정확한 경로를 알 수 없기 때문에 해당 지역에 서식하는 생물로부터 유래된 유전물질이 아닌, 외부로부터 유전자가 유입되었다면, 이를 유전자 분석 결

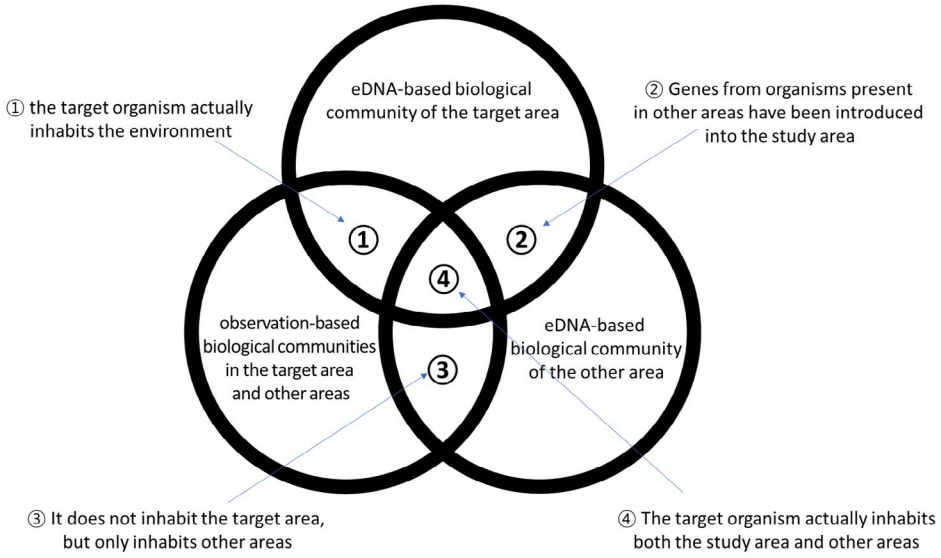

Fig. 8. Biodiversity community information relationship between eDNA and field survey.

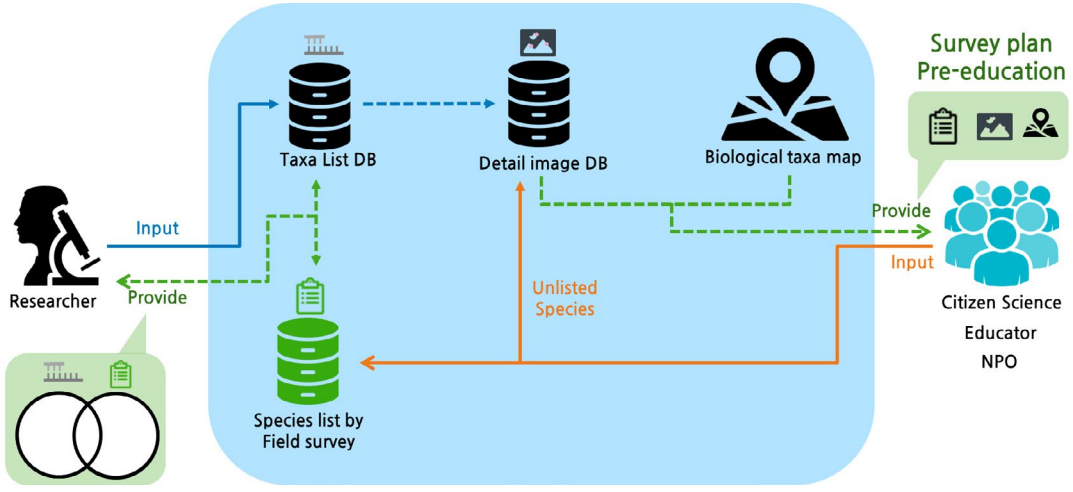

Fig. 9. Collaboration between researchers and citizen scientists using the eDNA database platform.

과만으로 정확히 파악하기 어렵다. 이러한 문제점은 대형생물에서 주로 발견된다.

철새와 같이 장거리를 이동하는 동물이 이동하는 과정에서 배출되는 배설물의 유전자가 유입될 수 있으며, 해양과 같은 수생태계에서는 수 km 밖에도 물의 흐름에 따라 이동할 수 있기 때문에 환경유전자 기반으로 발견된 생물이 해당 환경에 서식한다는 것을 확실하게 보장하기 어렵다(Shogren *et al*., 2017). 또한 수중의 수서무척추 곤충류는 수중에 많은 환경유전자를 배출하지 않으며 이로 인해 생물이 존재하지만 환경유전자 만으로 탐색이 어려울 수 있다(Blackman *et al*., 2019). 결과적으로 환경유전자만을 이용한 생물군집 정보는 생물의 서식 가능성을 의미할 뿐이며, 실제 해당 생물이 환경에 서식하는지 여부는 알 수 없다.

따라서 환경유전자를 이용한 생물군집을 분석함과 동시에 동일한 지역에서 해당 생물의 실체를 관찰함으로써 환경유전자 분석 결과의 신뢰도를 높이는 과정이 수반되어야 한다(Fig. 8). 이러한 생물의 실체 관찰은 두 가지 목적을 가지고 수행한다. ① 환경유전자 기반의 생물군집 분석 결과를 기반으로 군집목록에 존재하는 생물이 실제 환경에 서식하고 있는지, ② 군집목록에 존재하지 않는 생물이 해당 환경에 서식하고 있는지를 판단함으로써 환경유전자를 이용한 생물군집 분석의 한계점을 보완하는 것이다. 멸종위기종 또는 생활 특성이 매우 단편적인 생물의 경우에는 해당 환경에 유전물질의 농도가 매우 적을 수 있으며, 이로 인해 환경유전자로 인한 탐색에서 해당 생물이

누락될 수 있다. 따라서 ②번과 같은 경우에는 환경유전자 데이터베이스의 생물군집 목록에 실제 발견된 생물을 포함하여 보다 명확한 생물군집 데이터베이스를 구축할 수 있다. 이러한 환경유전자 기반의 생물군집구조의 데이터베이스와 실체 관찰을 통한 군집구조의 데이터베이스가 지속적으로 축적되어 하나의 데이터베이스를 구축한다면 보다 명확하고, 다양한 환경유전자 데이터베이스가 구축될 것으로 판단된다.

3. 시민과학을 통한 환경유전자 결과 검증

환경유전자에 의한 생물군집의 조사는 유전자 분석을 통한 생물의 서식을 추정하는 것이다. 즉 추정치에 유전자라는 근거를 통해 힘을 실어주는 방식이다. 그러나 해당 방법론은 샘플링에서부터 유전자 분석에 이르기까지 확률적인 기법에 의존할 수밖에 없다. 따라서, 환경유전자의 방법론적 고도화를 이루기 위해서는 확률에 의한 결과의 정확도를 검증할 수 있는 시스템적인 장치가 필요하다.

조사지에 서식하는 생물 종을 파악하는 가장 확실한 방법은 현장관찰을 통한 고전적 알파 분류방식이다. 그러나 이와 같은 종 동정은 오랜 훈련을 거친 전문 인력을 필요로 하며 조사 자체도 많은 시간과 인력이 소모된다. 하지만 역설적이게도 이와 같은 전문 인력의 부족과 그로 인한 조사 범위의 시, 공간적 한계를 뛰어넘을 수 있도록 하는 것이 환경유전자 분석기법이다. 그러나 관찰에 의한 검증 없이 환경유전자 분석만을 통해 산출된 생물군집 리스트는 '슈뢰딩거의 고양이 (Schrödinger's cat)'를 연상시킨다. 즉, 환경유전자 기법으로 도출된 리스트에 포함된 종은 관찰되기 전까지 그곳에 존재하며, 동시에 존재하지 않는 상태와도 같다. 다시 말해 환경유전자는 생물의 흔적일 뿐 해당 생물은 조사 지역에서 사멸하였거나 다른 지역으로 이동해 더 이상 존재하지 않을 수도 있으며, 강우, 포식자의 배설물 등을 통해 외부로부터 유전자만 유입되는 경우, 실제로는 서식하지 않는 생물의 유전자가 검출될 수도 있다. 이밖에도 환경유전자 기법의 샘플링과 primer 디자인에 의해 누락되는 종들을 고려하면 환경유전자 분석 결과의 불완전성은 더더욱 커지게 된다.

그럼에도 환경유전자 기법에 의한 군집 분석은 일정 수준의 신뢰도를 가진다. 따라서, 이를 활용하면 현장에서 관찰하여야 하는 생물 종의 폭을 크게 줄일 수 있어 비전문가도 단기간의 학습을 통해 해당 종들의 서식 유무 조사를 수행 할 수 있을 것이다.

시민과학은 다양한 분야에서 두각을 나타내고 있으나 특히 광범위한 시, 공간의 생태학 조사에 큰 힘을 발휘한다. 최초의 시민과학 활동인 Christmas Bird Count가 조류의 정성, 정량 조사였음은 시민과학과 태생과 그 성향이 생태학 분야 연구와 매우 밀접함을 보여준다. 그러나 조사를 수행하는 집단이 전문가가 아니라는 점에서 시민과학의 결과물은 언제나 데이터 신뢰성의 문제를 내포한다. 대부분의 시민과학 프로젝트는 이와 같은 한계로 인해 매우 제한적인 종을 대상으로 진행되거나(예: Texas Invasives: https://www.texasinvasives.org/) 전문지식이 필요하지 않은 측정(예: The Sechhi Disk Study: http://www.secchidisk.org/) 또는 시료채취 활동으로 그 범위를 국한시켜 왔다. 이는 시민과학을 단순한 작업이 반복되는 단조로운 활동으로 만드는 가장 큰 원인이다.

환경유전자 데이터베이스는 이와 같은 시민과학 활동을 보다 다채롭고 교육적인 활동으로 개선하는 데 기여할 수 있으며 일방적이던 연구자와 시민의 관계를 상호적으로 변화시킬 수 있다. 이를 위해서는 환경유전자 데이터베이스가 연구자 집단과 시민과학자 집단 사이의 상호 보완적인 연구 활동과 그 결과물의 공유를 가능하게 하는 플랫폼으로서 기능하여야 한다.

환경유전자 데이터베이스 플랫폼에 등록된 생물군집 리스트는 시민과학자들에게 어디서 어떤 생물들의 조사가 이루어져야 하는지에 대한 가이드를 제공하게 된다(Fig. 9). 시민들은 리스트에 존재하는 생물 종만을 대상으로 학습을 수행하므로 비교적 짧은 시간 내에 알파 분류 조사에 필요한 지식을 습득할 수 있으며, 다른 조사지 또는 군집 리스트를 선택하여 학습하는 종의 범위를 점진적으로 넓혀갈 수 있다. 이때 환경유전자 데이터베이스 플랫폼은 학습과 조사지 선정에 필요한 종들의 연결이미지 및 종 분포지도를 제공하는 한편, 시민과학 활동의 결과물인 조사지의 서식 생물 종 리스트를 등록할 수 있도록 구축되어야 한다. 환경유전자 데이터베이스 플랫폼에 축적되는 연구자와 시민과학자의 데이터는 연구자에게 환경유전자 결과의 정확도를 가늠할 수 있는 지표를 제공하는 한편 시민과학자에게는 명확한 활동 목표, 이를 달성하기 위한 정보 및 도구들을 제공하게 된다.

결 론

본 연구에서 제안하는 것과 같이 연구자와 시민과학자를 연결하고 서로의 결과물을 공유하는 한편 이를 통해 상호 집단이 필요로 하는 정보를 제공하는 선순환 구조는 전형적인 '플랫폼'의 형태이다. 이는 연구자가 환경유전자 분석을 통해 도출된 리스트상의 종이 실제 서식하는지 종과 일치하는지를 확인하고자 하는 필요성과, 시민과학 활동을 보다 다채롭고 교육적인 형태로 발전시키고자 하는 두 집단의 요구사항이 상호 보완적이기 때문에 성립할 수 있다. 그러나 이와 같은 형태는 두 집단의 활동에 필요한 요구사항을 모두 충족시킬 수 있는 기능이 한 시스템 안에 구축되어야 하기 때문에 일반적인 유전자 데이터베이스 시스템과 비교 시 매우 복잡한 구조적 특성을 가지게 될 것이다. 따라서, 이와 같은 요구사항을 모두 충족시키는 '대한

민국 환경유전자 데이터베이스 시스템'은 환경유전자 연구의 취약점을 시민과학 활동을 통해 보완 가능한 플랫폼 구조를 전제로 하여 구축되어야 한다.

적 요

최근 두드러지는 환경유전자만을 수집, 관리하는 새로운 데이터베이스 시스템을 구축하기 위한 노력의 배경에는 고유종의 식별이 가능한 '국지적인 유전자 데이터베이스'의 필요성, 그리고 '분석 결과의 시, 공간적 시각화' 같은 환경유전자 분야 연구의 고유한 특성이 존재한다. 하지만, 환경유전자 데이터의 불확실성으로 인해 현장에서의 알파 분류를 통한 종 조성 데이터와 비교분석이 가능한 형태의 환경유전자 데이터베이스 시스템의 구축이 필요하다. '시민과학'은 알파 분류에 소요되는 많은 인적자원을 시민의 연구 참여를 통해 해결할 수 있는 실마리를 제공한다. 또한, 환경유전자 분석을 통해 도출된 종 리스트는 시민과학자들에게 명확한 조사 지역과 대상을 특정하여 스스로 대상 생물을 학습하고 조사 가능한 범위로 좁혀 준다. 본 논문은 이와 같은 환경유전자 연구와 시민과학 활동 간의 상보성을 기반으로 한 '환경유전자-시민과학 데이터베이스 플랫폼'을 구상하였다. 이를 통해 기존의 환경유전자 데이터베이스의 개선을 도모하고, 시민이 연구에 주요한 역할로 참여하는 한편, 궁극적으로 양질의 연구데이터 축적뿐 아니라 시민들의 생태소양(Ecological Literacy) 함양이 가능한 '환경유전자-시민과학 데이터베이스 플랫폼'을 제시하고자 한다.

사 사

이 논문은 2022년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. 2020R1I1A1A01070367).

참고 문헌

Andersson, A.F., A. Bissett, A.G. Finstad, F. Fossøy, M. Grosjean, M. Hope, T.S. Jeppesen, U. Kõljalg, D. Lundin, R.N. Nilsson, M. Prager, C. Svenningsen and D. Schigel, 2021. Publishing DNA-derived data through biodiversity data platforms. v1.0 Copenhagen: GBIF Secretariat.

Anggoro, A., 2022. Savu Sea-Indonesia eDNA Dataset. Version 1.7. BIONESIA. Occurrence dataset https://doi.org/10.15468/hyxfc6 accessed via GBIF.org on 2022-09-25.

Atlas of Living Australia, 2021. Maximising fish detection with eDNA metabarcoding. Occurrence dataset https://doi.org/10.15468/7wphc2 accessed via GBIF.org on 2022-09-25.

Blackman, R., E. Mächler, F. Altermatt, A. Arnold, P. Beja, P. Boets, B. Egeter, V. Elbrecht, A.F. Filipe and J. Jones, 2019. Advancing the use of molecular methods for routine freshwater macroinvertebrate biomonitoring-The need for calibration experiments. *Metabarcoding Metagenome*, 3, 49-57.

Büdel, B., C. Colesie, T.A. Green, M. Grube, S.R. Lázaro, K. Loewen-Schneider, S. Maier, T. Peer, A. Pintado and J. Raggio, 2014. Improved appreciation of the functioning and importance of biological soil crusts in Europe: the Soil Crust International Project (SCIN). *Biodivers. Conserv.*, 23, 1639-1658.

Charles, E.C., R. Lopez, O. Stroe, G. Cochrane, C. Brooksbank, E. Birney and R. Apweiler, 2018. The European Bioinformatics Institute in 2018: tools, infrastructure and training. *Nucleic Acids Res.*, 2019, 47, 15-22.

Choi, N.W., K.Y. Lee, B.C. Kang, J.H. Park, S.C. Lee and S.S. Hwang, 2011. Instruction of wildlife integrated genetic information system (WIGIS). Nakdonggang National Institute of Biological Resources. Final Report.

Cochrane, G., I. Karsch-Mizrachi and T. Takagi, 2016. International Nucleotide Sequence Database Collaboration. *Nucleic Acids Res.*, 44, 48.

Convention on Biological Diversity (https://www.kbr.go.kr/).

Davron, D., A. Temur, T. Umida, I. Sari and T. Komiljon, 2022. Suitable habitat prediction with a huge set of variables on some central asian tulips. *J. Asia Pac. Biodivers.*, 16(1): 75-82.

Eric, W.S., M. Cavanaugh, K. Clark, J. Ostell, D.P. Kim and K.-M. Ilene, 2019. Genbank. *Nucleic Acids Res.*, 47, 37-42.

Frøslev, T. and R. Ejrnæs, 2018. BIOWIDE eDNA Fungi dataset. Danish Biodiversity Information Facility. Occurrence dataset https://doi.org/10.15468/nesbvx accessed via GBIF.org on 2022-09-25.

Fukuda, A., Y. Kodama, J. Mashima, T. Fujisawa and O. Ogasawara, 2021. DDBJ update: streamlining submission and access of human data. *Nucleic. Acids Res.*, 49, 71-75.

GBIF: Environmental DNA as a source for DNA-derived occurrence data (https://docs.gbif.org/publishing-dna-derived-data/1.0/en/#environmental-dna-as-a-source-for-dna-derived-occurrence-data).

Herder, J.E., A. Valentini, E. Bellemain, T. Dejean, J.J.C.W. van Delft, P.F. Thomsenand and P. Taberlet, 2014. Environmental DNA - a review of the possible applications for thedetection of (invasive) species. Stichting RAVON, Nijmegen. Report 2013-104.

INSDC: International Nucleotide Sequence Database Collaboration (https://www.insdc.org).

Kim, K.H., J.H. Ryu and S.J. Hwang, 2021. Sampling and Extraction Method for Environmental DNA (eDNA) in Freshwater Ecosystems. *Korean J. Ecol. Enviro.*, 54(3), 170-189.

Kim, K.H., N.Y. Kim, S.Y. Noh, J.H. Park and S.-J. Hwang, 2019. Assessment of trophic diatom index (TDI) based on eDNA in stream biofilm by Next Generation Sequence (NGS) case study in the streams of Han River watershed. Abstract book of Spring

conference, Korean Society of Limnology.

Max, M.-C., R.-C. Emilio, J.E. David, W. Bettina, B. Büdel, H. Hohne and W.K. Cornwell, 2022. Towards an understanding of future range shifts in lichens and mosses under climate change. *J. Biogeogr.*, 50(2), 406-417.

National Institute of Fisheries Science, Fisheries resource information center (https://www.nifs.go.kr/frcenter/mgrdb/).

National Marine Organism Genome Project (NMGP) (http://www.magic.re.kr/portal/intro/nmgp).

NYK line Fish eDNA database (https://www.nyk.com/english/news/2022/ 20220602_01.html).

Park, H.S., S.S. Ahn, B.Y. Ahn, H.H. Cho, J.H. Kwon, J.Y. Lim, S.J. Lee, G.J. Lee and J.H. Yang, 2005. Establishment of national systematics center of excellency for biological resource banks (Establishment and Research for Integrated Biodiversity Information Network). Korea Institute of Science and Technology Information. Final Report.

Shogren, A.J., J.L. Tank, E. Andruszkiewicz, B. Olds, A.R. Mahon, C.L. Jerde and D. Bolster, 2017. Controls on eDNA movement in streams: Transport, retention, and resuspension. *Sci. Rep.*, 7, 5065.

Varos, P., O. Fedor, F. Irina, D. Natalia, W. Andrey, K. Lyudmila and D. Andrew, 2023. The TOP-100 most dangerous invasive alien species in Northern Eurasia: invasion trends and species distribution modelling. *NeoBiota.*, 82, 23-56.