

간행물 등록번호

NNIBR-2021-0000248-10

환경유전자 정보 기반 담수 관심생물 탐지 기법 개발 (2단계 1차년도)

Validation of eDNA metabarcoding analysis for biodiversity
assessment of freshwater invertebrate community

다양성연구팀

국립낙동강생물자원관

발간등록번호: NNIBR-2021-0000248-10

환경유전자 정보 기반 담수 관심생물 탐지 기법 개발 (2단계 1차년도)

담수생물연구본부 동식물연구실 다양성연구팀

Validation of eDNA metabarcoding analysis for biodiversity
assessment of freshwater invertebrate community

Biodiversity Research Team
Animal & Plant Research Department
Nakdonggang National Institute of Biological Resources

2021

참여 연구진 및 담당 업무

구분	소 속	직급	성명(영문)	담 당 업 무
연구 책임자	다양성연구팀	원급	김지연 (Jiyeon Kim)	- 사업 총괄 - 환경유전자 정보 기반 생물다양성 분석(11지점) - 관심생물 대상 집단별 유전정보 추출(2종) - 환경유전자 연구 매뉴얼 작성(1식) - 논문 2건, 보도자료 1건
	다양성연구팀	선임급 (팀장)	유정남 (Jeong-Nam Yu)	- 담수 관심생물 유전다양성 분석(2종) - 보도자료 1건
내부 참여 연구원	다양성연구팀	원급	이승현 (Seung Hyun Lee)	- 환경유전자 정보 기반 생물다양성 분석(11지점) - 담수생물 표본 확보(2,500점)
	다양성연구팀	공무직	사공진 (Jin Sagong)	- DNA 바코드 정보 확보 지원 - 담수생물 표본 확보 지원
	동물연구팀	책임급 (팀장)	서경인 (Kyong-In Suh)	- 생물 표본 중 동정 및 유전정보 제공 (수서곤충 하루살이류)
	동물연구팀	원급	송지훈 (Ji-Hun Song)	- 생물 표본 중 동정 및 유전정보 제공 (수서곤충 날도래류)
외부 참여 연구원	이화여자대학교	교수	박중기 (Joong-Ki Park)	- 생물 표본 중 동정 및 생태 정보 제공(연체동물) - 환경유전자 분석 및 유전다양성 분석
	한양대학교	교수	노미나 (Mina Rho)	- 환경유전자 정보 분석 및 파이프라인 구축
	이화여자대학교	연구원	곽해나 (Haena Kwak)	- DNA 바코드 정보 수집 및 DB 구축
	고려대학교	연구원	강효정 (Hyo Jeong Kang)	- 생물 표본 중 동정 및 생태 정보 제공 (수서곤충 갈따구류)

목 차

목차	iii
초록	vi
I. 서론	1
II. 연구방법	2
1. 환경유전자 정보 확보 및 생물다양성 분석	2
2. 주요 관심생물 탐지 및 유전다양성 분석	5
3. 환경유전자 데이터베이스 및 파이프라인 구축	6
4. 환경유전자 정보 기반 생물다양성 분석 매뉴얼 작성	7
III. 결과 및 고찰	8
1. 환경유전자 정보 확보 및 생물다양성 분석	8
2. 주요 관심생물 탐지 및 유전다양성 분석	15
3. 환경유전자 데이터베이스 및 파이프라인 구축	18
4. 환경유전자 정보 기반 생물다양성 분석 매뉴얼 작성	22
IV. 종합 결론	23
V. 연구결과 활용 계획	27
VI. 참고문헌	29
VII. 부록(별책)	

환경유전자 기반의 생물 종 다양성 분석 매뉴얼(수서곤충편)

표 목 차

표 1. 환경유전자 조사지점 정보	2
표 2. 환경유전자 조사 및 시료 전처리 방법	3
표 3. 환경유전자 연구 특화 유전자 마커 정보	4
표 4. 유전다양성 분석 연구대상종 선정 사유	5
표 5. DNA 바코드 연구 특화 유전자 마커 정보	6
표 6. PCR 혼합물 조성 및 반응 조건 정보	7
표 7. 환경유전자 대용량 염기서열 종합 정보	8
표 8. 환경유전자 정보 지점별·시기별 생물다양성 비교·분석	9
표 9. 주름다슬기 지점별 유전자 다양성 지수	16
표 10. 주름다슬기 지점 간 유전적 차이	17
표 11. 왕우렁이 지점별 유전자 다양성 지수	17
표 12. 왕우렁이 지점 간 유전적 차이	18
표 13. 환경유전자 바코드 데이터 수집 현황	19
표 14. 환경유전자 정보 기반 생물다양성 분석 흐름도	19

그 립 목 차

그림 1. 환경유전자 정보 기반의 생물다양성 분석 연구 방향.....	1
그림 2. 환경유전자 조사지점 위치 및 현황.....	2
그림 3. 환경유전자 시료 DNA 및 Library QC.....	4
그림 4. 환경유전자 정보 생물상 분석.....	10
그림 5. 환경유전자 정보 생물다양성 지수 분석.....	10
그림 6. 환경유전자 정보 종수 탐지 기대치 분석.....	12
그림 7. 환경유전자 정보 계절별 군집 다양성 분석.....	13
그림 8. 환경유전자 정보 수계별·지점별 군집 다양성 분석.....	13
그림 9. 환경유전자 정보 수계별(세분화) 군집 다양성 분석.....	14
그림 10. 환경유전자 정보 분류군 이중 유사도 분석.....	15
그림 11. 주름다슬기 유전자형 network 및 분포.....	16
그림 12. 왕우렁이 유전자형 network 및 분포.....	18
그림 13. 환경유전자 정보 기반 생물다양성 분석 파이프라인.....	20
그림 14. 환경유전자 정보 대표 서열 선별 분석 방법 비교.....	21
그림 15. 환경유전자 기반의 생물 종 다양성 분석 매뉴얼(수서곤충편).....	22

초 록

본 연구는 환경유전자 정보를 활용한 차세대 생물탐지 및 생물다양성 분석 기법을 개발하고자 다음 4개의 과제로 구성하여 수행되었다.

첫 번째, 환경유전자 정보 확보 및 생물다양성 분석에서는 낙동강 본류 및 지류 11지점을 대상으로 계절별(봄, 여름, 가을, 겨울), 수심별(중층, 저층) 조사를 실시하여 총 176개의 환경유전자 시료를 확보하였다. 환경유전자 분석 연구에 특화된 유전자 마커 3건(mtDNA-CO1 1건, nrDNA-18S 2건)을 적용하여 총 264건의 대용량 환경유전자 정보(정보량 10.7G, Read count 53,543,890건, ASV 133,466건)를 확보하였다. 생물조사방법(환경유전자 연구 vs 생물측정망 연구)에 따른 생물다양성 비교 결과, 환경유전자 연구에서 더 높은 생물다양성이 측정되었으며, 두 생물조사방법 사이에서 분류군 과(Family) 수준에서 60%의 높은 탐지 일치율(공통 탐지 39과/생물측정망 64과)이 확인되었다. 계절별·수계별·지점별 생물다양성 분석(Rarefaction curve, PCoA, Two-way clustering 분석 등)을 통해 군집의 구조적 특성을 확인하였다. 분류군 종수 탐지 기대치 분석 결과, 계절 중 봄(253.0 ± 2.6 종), 수계 중 영강 지류(281.5 ± 1.8 종), 지점 중 영강1(151.4 ± 1.6 종)에서 가장 높은 종수 탐지 기대치가 측정되었다. 분류군 군집 다양성 분석 결과, 계절별로 군집 구조가 독립적으로 형성되었으며, 수계별·지점별로는 종 다양성이 높아 군집이 특정 구조를 형성하지 못하고 비구조화 경향을 보였다. 분류군 이중 유사도 분석 결과, 계절별로 일부 절지·환형·선형·편형동물에서 높은 이중 유사도로 탐지되었으며, 수계별·지점별로는 이중 유사도가 높은 분류군의 경향성이 뚜렷하게 나타나지 않고 다양한 분류군이 탐지되었다. 향후 환경유전자 정보 기반의 생물 탐지 효율과 분석의 신뢰도 향상을 위해서는 국내 서식종을 대상으로 지속적인 환경유전자 데이터베이스 구축(분류체계 정리, 다양한 유전자 부위의 유전 정보 확보 등), 세부적 연구 방법론 보완(환경유전자가 대표할 수 있는 지점의 범위를 고려한 지점 선정, 하천 규모에 따른 채수량 등)의 연구가 필요할 것으로 판단된다.

두 번째, 관심생물 탐지 및 유전다양성 분석에서는 낙동강 본류 및 지류 11지점에서 확보한 대용량 환경유전자 정보(10.7G, Read count 53,543,890건, ASV 133,466건)로부터 관심생물에 해당하는 데이터를 추출하여 집단별 유전다양성 분석에 사용하는 연구 방법을 적용하였다. 연구대상종으로는 연체동물 내에서 생태계 보전·관리를 위해 개체군 증감의 추적이 필요한 한국고유종 주름다슬기, 생태계교란종 왕우렁이를 선정하였다. 주름다슬기와 왕우렁이의 유전자형 분포, 유전자 다양성 지수 산출을 통해 집단별 유전적 특성과 유전적 차이를 확인하였다. 주름다슬기는 11지점 중 5지점에서 서식이 탐지되었으며, 총 5개 집단에서 확보한 mtDNA-CO1 유전자(313bp) 정보 83개에서 총 4개의 유전자형이 도출되었다. 왕우렁이는 11지점 중 4지점에서 서식이 탐지되었으며, 총 4개 집단에서 확보한 mtDNA-CO1 유전자(313bp) 정보 405개에서 총 4개의 유전자형이 도출되었다. 주름다슬기와 왕우렁이 모두 하류 지점에서 높은 유전다양성이 확인되었으며, 이는 상류 지점의 유전자 유입, 농수 확보를

위한 하천수 사용으로 인한 영향으로 해석된다. 향후 환경유전자 정보 기반의 관심생물 탐지 및 유전다양성 분석의 신뢰도 향상을 위해서는 개체군 분석에 필요한 유전자 정보 수 확보를 위한 방법론(공통 유전자 마커, 특정 종 맞춤형 유전자 마커 혼합 적용 등)의 개선이 필요 할 것으로 판단된다.

세 번째, 환경유전자 데이터베이스 구축을 위해서 환경유전자 바코드 데이터 929,649건을 수집 및 정제하였다. 공개 DB(NCBI, BOLD, SILVA)에서 주요 4개 분류군(연체·절지·환형·선형동물)을 대상으로 mtDNA-CO1 및 nrDNA-18S 바코드 데이터 929,495건을 수집하였다. 자체적으로는 국내에 자생하는 연체동물 17종, 절지동물 13종에 대한 mtDNA-CO1 바코드 데이터 154건을 확보하였다. 환경유전자 정보 기반의 생물다양성 분석 연구에 특화된 파이프라인 구축을 위해서 분석 프로그램 선정 및 안정화 검증을 수행한 후, 환경유전자 분석 맞춤형 파이프라인 기반을 구축하였다. 4단계 분석 프로세스는 Quality filtering & assembly, Clustering, Taxonomy assignment, Biodiversity statistics로 구성하였다.

네 번째, 당해연도 연구 방법과 결과를 기반으로 환경유전자 연구에 적용할 수 있는 표준화된 생물다양성 분석 매뉴얼(환경유전자 정보 기반의 생물다양성 분석 매뉴얼 - 수서곤충편)을 작성하였다. 환경유전자 시료 채집 및 전처리 방법, 환경유전자 추출 및 유전자 마커 적용 방법, 생물다양성 분석 방법 등 환경유전자 연구 전 과정에 대한 연구 방법이 포함되었다.

본 연구를 통해 환경유전자 정보를 이용하여 특정 담수생물 탐지, 계절별·수계별·지점별 생물다양성(종, 유전자) 분석, 군집 구조 분석(Rarefaction curve, PCoA, Two-way clustering 분석 등)의 가능성을 확인하였다. 연구 결과는 향후 유전자 연구 분야의 학술적 발전, 정부 정책 수립 및 추진, 환경부 및 지자체의 현안문제 대응 전략 수립, 담수생물다양성 평가 체계 마련을 위한 과학적 근거 자료로 활용할 계획이다.

I. 서 론

1. 연구 배경 및 필요성

가. 최근 기후·환경 변화에 따른 담수 관심생물의 대발생, 사멸 등으로 환경적·경제적 피해가 증가되고 있음에도 불구하고,

- (1) 생물 기초정보(종 정보, 분포 현황 등) 미흡으로 원인 파악과 대응이 어려움
- (2) 담수 생물 종의 변화 추이를 실시간으로 모니터링 할 수 있는 새로운 개념의 생물다양성 평가 및 관리에 관한 표준화된 기술 개발이 필요한 시점

나. 전통적 방식의 생물다양성 연구 방법에 최첨단 차세대 유전자 분석 기술*을 도입하여, 생물다양성 평가 연구를 강화하고 담수 관심생물 관련 환경부 및 지자체의 현안문제 대응을 위한 연구 결과의 활용성 확대 필요

* 차세대 유전자 분석 기술: 담수 환경시료에서 수집 가능한 환경유전자(Environmental DNA, eDNA)를 이용하여, 다양한 생물의 대응량 유전자 정보를 분석(Metabarcoding)하고 관심종의 유무, 종 판별, 종 다양성 등을 확인할 수 있는 첨단 복합 기술(Thomsen et al., 2012; Bohmann et al., 2014; Wangensteen and Turon, 2017; Andújar et al., 2018)

다. 최근 정부의 ‘디지털 뉴딜’ 추진에 따른 D.N.A.(Digital·Network·AI) 생태계 강화를 위해 빅데이터 기반의 최신 생물다양성 분석 기술 도입 필요

2. 연구 목적 및 내용

가. 환경유전자 정보 기반의 담수 관심생물 탐지 기법 개발을 통해,

- (1) 지표종, 생태계 교란종, 외래종, 멸종위기종 등 국내 서식에 대해 관심이 필요한 담수생물의 다양성(종, 분포, 확산 등) 분석
- (2) 환경유전자 정보 기반의 생물다양성 분석 매뉴얼 정립 후 관계기관으로의 연구 기술 보급

나. 연구 내용(그림 1)

- (1) 환경유전자 정보 확보 및 생물다양성 분석
- (2) 주요 관심생물 탐지 및 유전다양성 분석
- (3) 환경유전자 데이터베이스 및 파이프라인 구축
- (4) 환경유전자 정보 기반 생물 다양성 분석 매뉴얼 작성



그림 1. 환경유전자 정보 기반의 생물다양성 분석 연구 방향

II. 연구방법

1. 환경유전자 정보 확보 및 생물다양성 분석

※ 환경유전자 기반의 생물다양성 분석 기법(담수무척추동물편) 적용(국립낙동강생물자원관, 2020)

가. 조사지점 : 낙동강 본류 및 지류 11지점, 총 4회(봄, 여름, 가을, 겨울) 조사(그림2, 표 1)

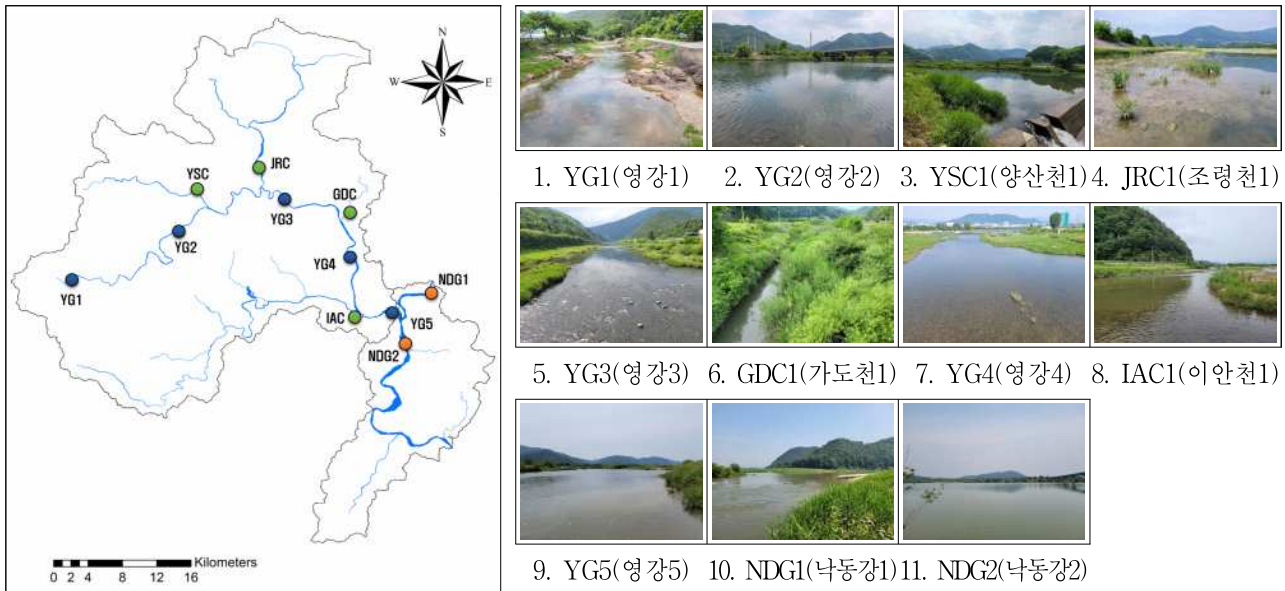


그림 2. 환경유전자 조사지점 위치 및 현황

* ● : 영강 본류, ● : 영강 지류, ● : 낙동강 본류

표 1. 환경유전자 조사지점 정보

번호	지점(영문)	중권역	주소
1	영강1(21YG1)	영강	경상북도 상주시 화북면 청화로 148
2	영강2(21YG2)	영강	경상북도 문경시 가은읍 작약로 870
3	양산천1(21YSC1)	영강	경상북도 문경시 가은읍 도리실길 54
4	조령천1(21JRC1)	영강	경상북도 문경시 마성면 문경대로 1768
5	영강3(21YG3)	영강	경상북도 문경시 호계면 문경대로 1030
6	가도천1(21GDC1)	영강	경북 문경시 호계면 구산리 343-1
7	영강4(21YG4)	영강	경북 문경시 영신동 463-1
8	이안천1(21IAC1)	영강	경북 상주시 함창읍 금곡리 264
9	영강5(21YG5)	영강	경북 문경시 영순면 말응리 산 90-1
10	낙동강1(21NDG1)	낙동강	경북 예천군 풍양면 삼강리길 27
11	낙동강2(21NDG2)	낙동강	경북 상주시 사벌국면 매호리 산 31-3

나. 조사 및 시료 전처리 방법(표 2)

(1) 조사방법

- 조사지점 내 다양한 환경(여울·소 등)을 대상으로 수행
- 중층시료 : VanDorn water sampler, eDNA sampler(Smith-root Inc.; Smith et al., 2003, Thomals et al., 2018)를 이용하여 용량별(500ml, 1L, 3L) 채수 후 냉장보관(4℃)
- 저층시료 : 표준채(45 μ m, 75 μ m)를 이용하여 현장수(45L)와 하상재질(1kg)을 혼합하여 채질 후 채수병에 냉장보관(4℃)
- 생물시료 : 정성채집법(Hand net, Kick net)을 이용하여 연체동물·수서곤충 등을 채집 후 알코올(70%) 고정

(2) 시료 전처리 방법 : Filter/Gast pump(5 in. Hg)를 이용한 필터링(pore size : 0.2, 1.2, 3, 5 μ m) 후 냉동보관(-20℃)

※ 저층시료의 경우 45 μ m 이하의 하상 재질 포함

표 2. 환경유전자 조사 및 시료 전처리 방법

구분	중층시료	저층시료	생물시료
현장 조사			
시료 전처리			

다. 환경유전자 정보 확보

(1) DNA 추출 및 QC

- DNA 추출 : 혼합물을 한 번에 파쇄하고 추출할 수 있는 Qiagen DNeasy power soil kit 적용(Qiagen kit protocol 이용)
- DNA QC : 추출한 DNA의 Quantity(fluorescence-based method), Quality(gel electrophoresis method), Size Check(그림 3)

(2) 유전자 마커 적용을 통한 Library 제작 및 QC

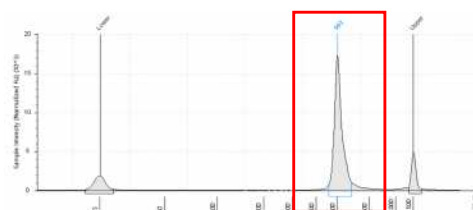
- 유전자 마커 mtDNA-CO1 1건, nrDNA-18S 2건 적용(표 3)

- Library Preparation kit(Herculase II Fusion DNA Polymerase Nextera XT Index Kit V2)를 이용한 Library 제작 및 QC(그림 3)

※ 예상 유전자 사이즈에서 단일 피크 확인 시, Library 제작 QC 통과

- (3) 대용량 환경유전자 정보 확보(264건) : Illumina Miseq platform을 이용하여 약 300bp 사이즈의 Illumina short read 서열 생산

Sample Name	Conc. (ng/μl)	Final Volume(μl)	Total Amount(μg)	Result
21YG1F1_CO1	9.043	35	0.317	Pass



<DNA QC>

<Library QC>

그림 3. 환경유전자 시료 DNA 및 Library QC(예시 시료 21YG1F1_CO1)

표 3. 환경유전자 연구 특화 유전자 마커 정보

No.	Primer name	Strand	sequence	Target size(bp)	Citation
1	CO1(Leray mCO1intF)	F	GGWACWGGWTGAACWGTWTAYCCYCC	313	Leray et al. (2013)
2	CO1(HCO2198)	R	TAAACTTCAGGGTGACCAAAAAATCA	313	Folmer et al. (1994)
3	18S(V1/2)	F	GCTTGTCTCAAAGATTAAGCC	400	Blaxter et al. (1998)
4	18S(V1/2)	R	GCCTGCTGCCTTCCTTGGA	400	Blaxter et al. (1998)
5	18S(V4)	F	AGGGCAAKYCTGGTGCCAGC	530	Zhan et al. (2013)
6	18S(V4)	R	GRCGGTATCTRATCGYCTT	530	Zhan et al. (2013)

라. 환경유전자 정보 기반 생물다양성 분석 : 자체적으로 구축하는 환경유전자 데이터베이스 및 분석 파이프라인을 활용하여, 대용량 환경유전자 raw data로부터 Assembly, Preprocessing, Clustering, Taxonomic assignment, Diversity statistics 등의 순차적인 생물다양성 분석 수행

- (1) Preprocessing & Clustering : DADA2 프로그램을 이용해 raw data로부터 분석에 사용할 대표 서열 선별(Callahan et al., 2016)
- (2) Taxonomic assignment & Diversity statistics
 - 자체 구축된 환경유전자 데이터베이스를 reference로 대표 서열에 대한 taxonomic assignment 수행
 - Alpha-diversity, Taxonomy-diversity, Alpha-rarefaction, PCoA, UPGMA Tree,

Heatmap 등의 다양한 다양성 분석(diversity analysis) 수행

※ 환경유전자 데이터베이스를 구축 중인 주요 5개 분류군(연체·절지·환형·선형·편형동물)을 대상으로 생물다양성 분석 수행

※ Alpha-diversity : 우점도 지수(Dominance index; DI), 다양도 지수(Diversity index; H'), 균등도 지수(Evenness index; EI, J'), 풍부도 지수(Richness index; RI)

2. 주요 관심생물 탐지 및 유전다양성 분석

가. 연구대상종 선정(연체동물 2종 : 주름다슬기, 왕우렁이)

- 연체동물은 전국적으로 서식 범위가 넓으나 하천 규모 등 서식환경 요인에 따라 채집이 어려워 분포범위 및 규모가 명확하게 파악되지 못하고 있는 분류군으로 환경유전자 탐지 기법을 적용하여 신속·정확한 분포 범위 및 서식 유무 파악이 가능함
- 연체동물 내에서 생태계 보전·관리를 위해 개체군 증감의 추적이 필요한 한국고유종 주름다슬기, 생태계 교란종 왕우렁이 선정(표 4)

표 4. 유전다양성 분석 연구대상종 선정 사유

과명-학명	학명	선정 사유 / 관련 근거
다슬기과 (Pleuroceridae) - 주름다슬기 (<i>Semisulcospira forticosta</i>)		- 국내 3속 10종 서식 보고됨 - 환경부 지정 멸종위기종(1종, 염주알다슬기)과 한국고유종(6종) 포함 - 서식 범위가 넓으나 수환경 규모에 따라 채집이 어려운 분류군 - 수환경 평가에 널리 이용되고 있음 - 국외반출 승인대상 생물자원, 한국고유종
사과우렁이과 (Ampullariidae) - 왕우렁이 (<i>Pomacea canaliculata</i>)		- 국내 1속 2종 서식 보고됨 - 세계자연보전연맹(IUCN) 지정 생태계교란종 - 서식 범위가 넓으나 수환경 규모에 따라 채집이 어려운 분류군 - 생태계교란종

* 출처 : 국립생물자원관 한반도의 생물다양성(<https://species.nibr.go.kr>)

나. 유전적 다양성 분석 : Haplotype network, Haplotype diversity, Nucleotide diversity 등 다양한 유전적 다양성 분석 수행

3. 환경유전자 데이터베이스 및 파이프라인 구축

가. 환경유전자 바코드 데이터 수집

- (1) 공개 데이터 : 공개 DB(NCBI, BOLD, SILVA)에서 4개 분류군(연체·절지·환형·선형동물)을 중심으로 mtDNA-CO1 및 nrDNA-18S 유전자 바코드 데이터 수집
- (2) 자체 데이터 : 정성채집법으로 확보한 연체동물·수서곤충을 대상(30종)으로 mtDNA-CO1 유전자 바코드 데이터 확보
- (가) Genomic DNA 추출 : DNeasy Blood & Tissue kit 사용(Qiagen kit protocol 이용)
 - 1) 다리, 날개 등의 조직 일부를 적출하여 1.5ml tube에 넣음
 - 2) ATL buffer 180 μ l, proteinase K 20 μ l를 첨가 및 혼합한 뒤, 56℃에서 조직을 완전히 용해
 - 3) 15초간 vortexing한 뒤, AL buffer 200 μ l를 첨가하고 다시 vortexing하여 혼합
 - 4) 순수 에탄올 200 μ l를 첨가하고 vortexing하여 혼합
 - 5) 혼합액을 DNeasy Mini spin column으로 옮긴 뒤, 8,000rpm으로 1분간 원심분리
 - 6) 새로운 2ml collection tube로 교체한 뒤, AW1 buffer 500 μ l를 첨가하고 8,000rpm에서 1분간 원심분리
 - 7) 새로운 2ml collection tube로 교체한 뒤, AW2 buffer 500 μ l를 첨가하고 14,000rpm에서 3분간 원심분리
 - 8) 2ml collection tube를 버리고 새로운 1.5ml tube로 교체
 - 9) 멸균한 3차 증류수 또는 AE buffer 100 μ l를 첨가하고 5분간 실온에 둔 뒤, 8,000rpm에서 1분간 원심분리하여 최종산물 DNA 추출
 - 10) 추출된 Genomic DNA는 0.7% agarose 겔 또는 분광 광도기(Nanodrop, 미국)를 이용하여 Quality 검증(순도, 농도 등)
- (나) 유전자 마커 선정 : 가장 대표적인 바코드 DNA 유전자 마커(mtDNA-CO1 유전자 영역, LCO1490/HCO2198) 적용(Folmer et al., 1994, 표 5)

표 5. DNA 바코드 연구 특화 유전자 마커 정보

No.	Primer name	Strand	Sequence	Target size(bp)	Citation
1	LCO1490(CO1)	F	GGTCAACAAATCATAAAGATATTGG	650	Folmer et al. (1994)
2	HCO2198(CO1)	R	TAAACTTCAGGGTGACCAAAAAATCA		

- (다) PCR 및 Purification : 선정된 유전자 마커를 이용한 유전자 증폭(TaKaRa Ex Taq, TaKaRa Bio.) 및 증폭 산물 정제(QIAquick PCR Purification Kit, QIAGEN Co.)(표 6)

표 6. PCR 혼합물 조성 및 반응 조건 정보

PCR mixture		PCR condition		
TaKaRa Ex Taq	0.5 μ l	Initial	95℃	1min
10x Ex Taq Buffer(20mM Mg2+ plus)	5 μ l	Denaturation	95℃	30S (40C)
dNTP Mixture	4 μ l	Denaturation	45~50℃	30S (40C)
Primer F	2 μ l	Annealing	72℃	1min (40C)
Primer R	2 μ l	Extension	72℃	10min
D.W.	32.5 μ l	Final extension	4℃	10min
gDNA	4 μ l			
Total volume	50 μ l			

(라) Sequencing 및 데이터 정리

- 1) ABI PRISM 3730XL Analyzer(Applied Biosystems Co.)를 이용한 Sequencing 수행
- 2) 확보한 raw data의 Assembly 및 Blast를 통한 최종 종 동정(Geneious v.11.0.5)

나. 환경유전자 분석 파이프라인 구축

- 환경유전자-메타바코딩 분석 방법과 담수생물에 특화된 생물다양성 분석 파이프라인 구축을 위해 분석 프로그램 선정 및 안정화 검증을 수행한 후, 환경유전자 분석 맞춤형 파이프라인 기반 구축

4. 환경유전자 정보 기반 생물다양성 분석 매뉴얼 작성

- 환경유전자의 조사방법에서 분석방법에 이르기까지 환경유전자를 이용한 생물 다양성 연구의 전반적인 업무 매뉴얼 작성
- ※ 환경유전자 시료 채집 및 전처리 방법, 환경유전자 추출 및 유전자 마커 적용 방법, 생물다양성 분석 방법 등에 대한 내용 포함

III. 연구결과 및 고찰

1. 환경유전자 정보 확보 및 생물다양성 분석

가. 환경유전자 대용량 염기서열(NGS) 확보

- Illumina Miseq platform을 이용해 환경유전자 88개 시료에 대한 유전자 마커별(3종) 대용량 염기서열 확보(264건, 표 7)

※ 정보량 10.7G, Read count 53,543,890건, ASV 133,466건 확보

- Raw data 정제 및 ASV 대표 서열 선별을 통해 확보한 총 ASV 133,466건의 데이터를 기반으로 생물다양성 분석 수행

표 7. 환경유전자 대용량 염기서열 종합 정보

조사시기	Data size	Target region	Total read bases(bp)	Read count(건)	ASV(건)
겨울 (n=22)	2.2G	COI	1,175,293,630	3,904,630	12,794
		18S(V1/2)	1,049,088,544	3,485,344	5,272
		18S(V4)	985,392,730	3,273,730	5,769
봄 (n=22)	2.7G	COI	1,515,308,046	5,034,246	27,569
		18S(V1/2)	1,406,632,598	4,673,198	9,646
		18S(V4)	1,451,458,722	4,822,122	11,058
여름 (n=22)	2.6G	COI	1,392,787,200	4,627,200	16,375
		18S(V1/2)	1,382,452,064	4,592,864	6,911
		18S(V4)	1,246,828,085	4,142,286	7,921
가을 (n=22)	3.2G	COI	1,550,211,404	5,150,204	13,589
		18S(V1/2)	1,554,123,980	5,129,980	8,038
		18S(V4)	1,417,133,886	4,708,086	8,524
합계 (n=88)	10.7G	Merging	16,126,710,889	53,543,890	133,466

* n : 환경유전자 시료 수. 총 88개=11지점×2개 유형(저층·중층)×4계절

나. 환경유전자 기반 생물다양성 분석

(1) 환경유전자 기반 생물다양성 분석 결과

- 환경유전자 정보를 기반으로 조사지점(11지점)의 4계절(봄, 여름, 가을, 겨울) 생물다양성을 분석한 결과, 주요 5개 분류군에서 총 20강 65목 229과 409속 604종의 담수생물이 확인됨(표 8)
- 분류군별 생물다양성 비교 결과, 절지동물이 10강 32목 132과 247속 358종으로 가장 많은 종이 탐지 되었으며, 다음으로 선형동물 2강 12목 49과 73속 97종, 환형동물 2강 8목 11과 38속 61종, 편형동물 4강 8목 21과 32속 57종, 연체동물 2강 5목 16과 19속

31종 순으로 많은 종이 탐지됨(그림 4)

- 계절별 생물다양성 비교 결과, 봄에 264종으로 가장 많은 생물이 탐지 되었으며, 다음으로 가을 228종, 여름 205종, 겨울 150종 순으로 많은 종이 탐지됨(그림 4)
- 지점별 생물다양성 비교 결과, 영강1 지점에서 155종으로 가장 많은 종이 탐지되었으며, 다음으로 양산천 148종, 영강3, 조령천 141종, 영강2 137종, 가도천 132종, 낙동강2 122종, 영강5 106종, 영강4·낙동강1 105종, 이안천 102종 순으로 많은 종이 탐지됨(그림 4)
- 계절별, 수계별* 생물다양성 지수 분석 결과, 낙동강 본류-겨울에 가장 높은 우점도, 낙동강 본류-가을에 가장 높은 다양도 및 균등도, 영강 지류-가을에 가장 높은 풍부도 지수가 측정됨(그림 5)

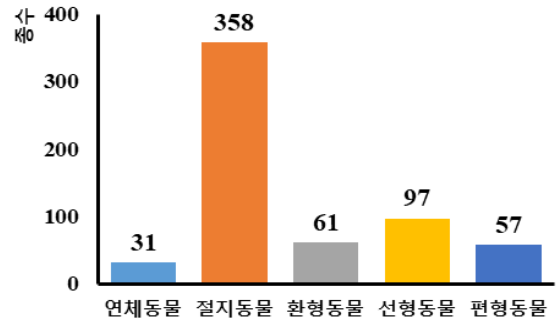
* 11개 조사지점을 3개 수계로 구분 : 영강 본류(영강1-5), 영강 지류(양산천, 조령천, 가도천, 이안천), 낙동강 본류(낙동강1-2)

표 8. 환경유전자 정보 지점별·시기별 생물다양성 비교·분석

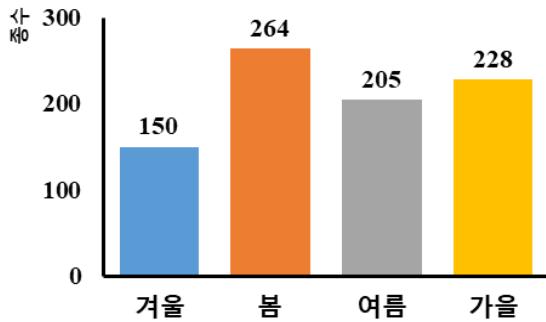
구분	겨울					봄					여름					가을					합계				
	강	목	과	속	종	강	목	과	속	종	강	목	과	속	종	강	목	과	속	종	강	목	과	속	종
영강1 (21YG1)	10	14	21	34	50	14	28	64	100	130	13	24	46	76	111	11	21	40	68	104	16	36	96	106	↑155
영강2 (21YG2)	14	24	31	43	62	13	29	49	80	110	12	22	34	45	73	15	27	49	80	119	16	41	81	117	137
양산천1 (21YSC1)	9	11	17	24	31	14	32	53	78	110	14	29	51	77	122	12	26	52	81	123	17	42	99	118	148
조령천1 (21JRC1)	12	20	29	39	55	13	26	44	64	85	12	19	29	49	70	13	29	51	77	122	15	43	88	111	141
영강3 (21YG3)	13	20	34	44	58	14	30	63	77	111	13	27	55	81	119	13	27	46	70	104	16	40	107	119	147
가도천1 (21GDC1)	13	22	35	49	69	14	24	43	64	92	12	22	35	49	67	16	33	59	92	129	17	38	80	102	132
영강4 (21YG4)	8	15	17	22	31	13	27	41	54	80	8	9	11	13	15	12	22	37	68	103	14	35	62	85	105
이안천1 (21IAC1)	9	15	19	26	41	14	24	31	39	66	13	22	28	47	67	13	25	38	60	100	15	32	57	82	102
영강5 (21YG5)	11	16	23	38	60	13	24	37	59	83	11	19	26	47	68	12	22	36	55	85	14	32	62	86	106
낙동강1 (21NDG1)	10	15	22	38	54	10	20	30	45	62	12	25	31	40	60	14	27	39	58	84	15	36	66	85	105
낙동강2 (21NDG2)	11	17	22	26	36	12	22	36	55	83	14	28	46	65	104	12	28	45	65	101	15	35	78	92	122
합계	17	37	89	100	150	16	49	163	214	↑264	16	44	118	156	205	18	50	129	180	228	20	65	229	409	604

* 11지점의 생물다양성(종)을 비교하여 가장 높은 수치 앞에 ↑(빨간색 위 화살표) 표시

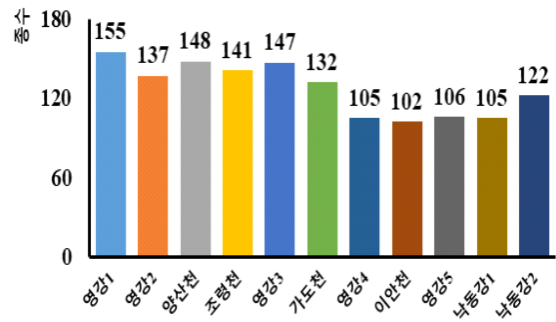
* 4계절의 생물다양성(종)을 비교하여 가장 높은 수치 앞에 ↑(파란색 위 화살표) 표시



<분류군별 생물다양성 비교>

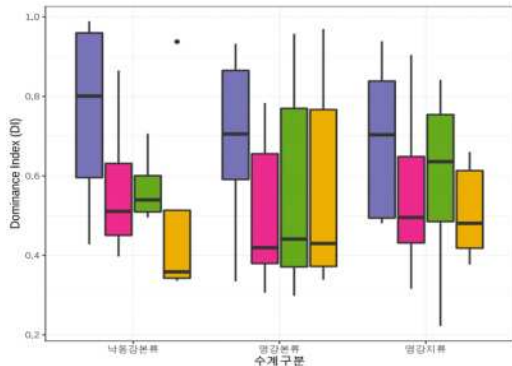


<계절별 생물다양성 비교>

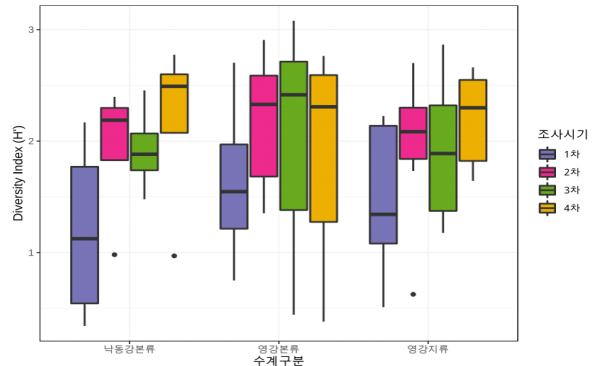


<지점별 생물다양성 비교>

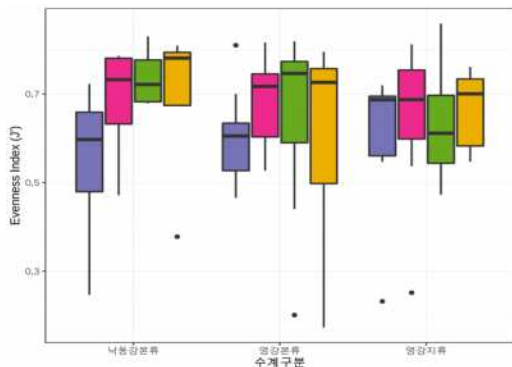
그림 4. 환경유전자 정보 생물상 분석



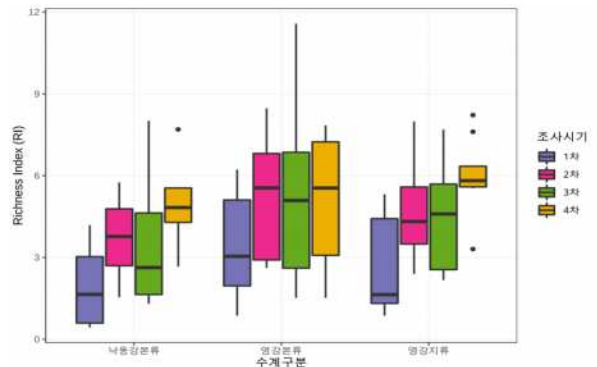
<우점도 지수>



<다양도 지수>



<균등도 지수>



<풍부도 지수>

그림 5. 환경유전자 정보 생물다양성 지수 분석

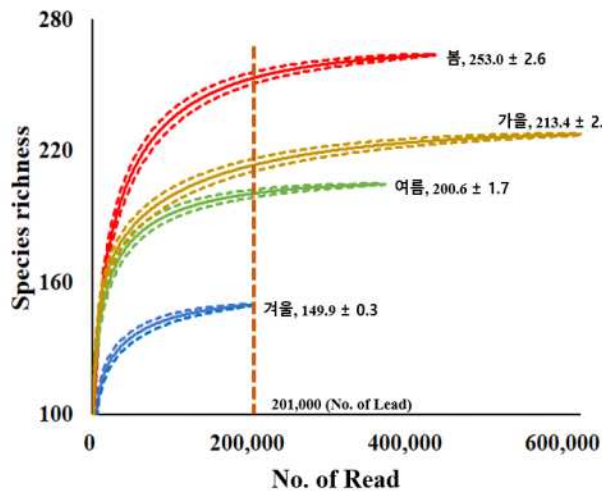
* 조사시기 index 정보 : 1차(겨울), 2차(봄), 3차(여름), 4차(가을)

(2) 생물조사방법에 따른 생물다양성 비교·분석

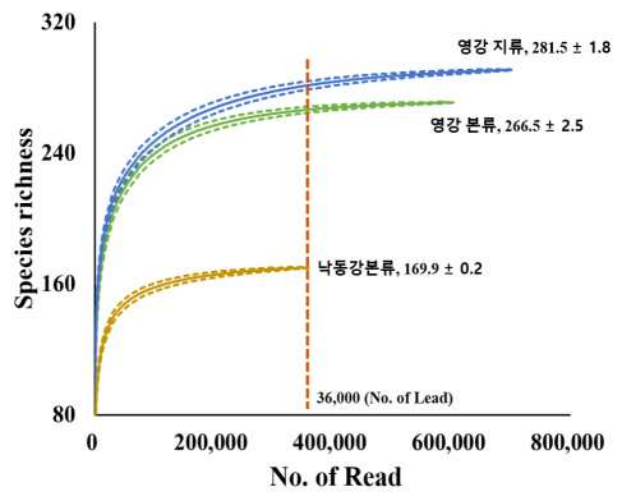
- 본 연구의 환경유전자 정보 기반 탐지 기법과 국립환경과학원에서 운영하는 생물측정망(저서성대형무척추동물)의 생물다양성을 비교·분석함
- ※ 생물측정망 자료의 범위는 2011~2018년까지 매년 2회씩 총 16회 수행된 자료임
- 두 가지 생물조사방법에서 공통으로 확인된 담수생물은 총 3문 3강 6목 17과 27속 36종으로 꼬마하루살이과(Baetidae) 5속, 납작하루살이과(Heptageniidae) 2속, 줄날도래과(Hydropsychidae) 2속, 다슬기과(Pleuroceridae) 1속 등이 대표적임
- 환경유전자 정보 기반 탐지 기법에서만 확인된 담수생물은 총 3문 20강 64목 232과 395속 568종으로 물지렁이과(Naididae) 18속, 깔따구과(Chironomidae) 39속, 선형동물(Nematoda) 69속, 편형동물(Platyhelminthes) 32속* 등이 대표적임
- * 현장 또는 실험실에서 종 동정이 어려운 미소생물과 분류체계 정립이 필요한 담수생물로 본 연구의 환경유전자 정보 기반 탐지 기법으로 신속·정확한 종 동정이 가능
- 생물측정망 조사방법에서만 확인된 담수생물은 총 3목 25과 82속 114종으로 민강도래과(Nemouridae) 1속, 각다귀과(Tipulidae) 2속, 뱀잠자리과(Corydalidae) 2속* 등이 대표적임
- * 환경유전자 데이터베이스에 유전 정보가 없는 담수생물로 국내 서식종 대상의 유전 정보 확보의 필요성 확인
- 두 가지 생물조사방법을 비교한 결과, 분류군별로 과(Family) 수준에서는 60%의 높은 탐지 일치율(공통 탐지 39과/생물측정망 64과)을 보였으나, 하위 분류군 수준으로 내려갈수록 탐지 일치율이 낮아지는 경향을 보임. 이는 국내 서식종에 대한 유전 정보 부재로 인한 결과로 환경유전자 정보로 분류군의 탐지 효율을 높이기 위해서는 국내 서식종을 대상으로 지속적인 DNA 바코드 정보 및 환경유전자 데이터베이스 구축이 필요함

(3) 환경유전자 기반 주요 분류군 종수 탐지 기대치 분석(그림 6)

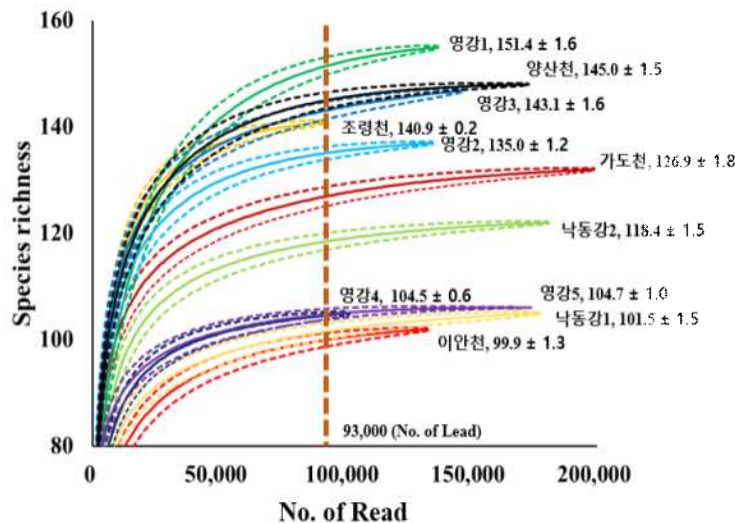
- Rarefaction curve 분석 방법으로 계절별, 수계별, 지점별로 생물 종수 탐지를 위해 필요한 최적의 유전 정보량과 종수 탐지 기대치를 분석함
- 계절별 탐지 종수 기대치 분석 결과, 유전 정보량 201,000건 기준에서 종수 탐지 기대치는 봄이 253.0 ± 2.6 종으로 가장 높은 수치로 분석되었으며, 겨울이 149.9 ± 0.3 종으로 가장 낮은 수치로 분석됨. 이는 봄에 생물의 활동성이 높아 담수 내 생물 조직의 부유량이 증가한 영향으로 판단됨
- 수계별·지점별 종수 기대치 분석 결과, 수계별로 영강 지류에서 281.5 ± 1.8 종(유전 정보량 36,000건 기준), 지점별로 영강1(21YG1)에서 151.4 ± 1.6 종(유전 정보량 93,000건 기준)으로 가장 높은 종수 탐지 기대치가 분석됨. 하천 상류에서 높은 종수 기대치가 확인 되었으나, 향후 지점 특성(하폭, 유량, 유속 등)에 따른 채수 방법 보완 등의 추가 연구가 필요할 것으로 판단됨



<계절별 종수 기대치 분석>



<수계별 종수 기대치 분석>



<지점별 종수 기대치 분석>

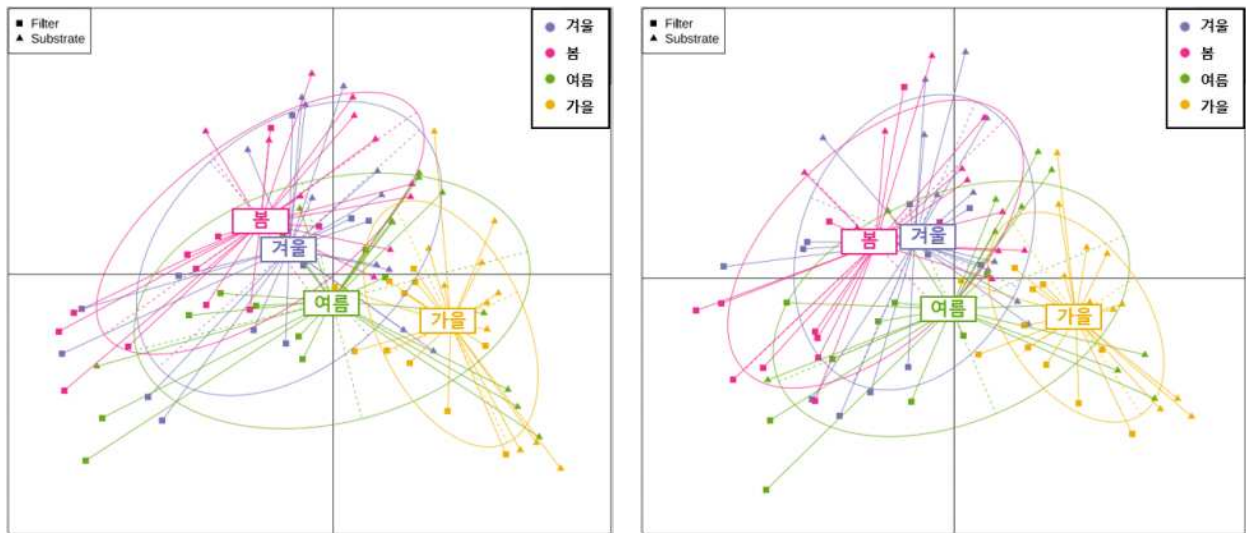
그림 6. 환경유전자 정보 종수 탐지 기대치 분석

(4) 환경유전자 기반 주요 분류군 군집 다양성 분석

- PCoA 분석 방법으로 주요 분류군의 탐지에 따라 계절별, 수계별, 지점별 나타나는 요인으로 군집 다양성 분석하였음
- 계절별 군집 다양성 분석 결과, 각 계절별 탐지 분류군의 양상이 다르게 분석되어 계절별로 군집 구조가 독립적으로 형성됨. 가을의 군집 구조는 겨울, 봄, 여름의 군집 구조와 가장 큰 차이를 보였으며, 이는 생물의 생활사 또는 우기의 영향(유지유량의 증가로 지류 또는 상류로부터 유입된 유전정보를 포함한 부유물 증가)으로 기존의 군집 구조에 변화가 있었던 것으로 판단됨(그림 7)
- 수계별·지점별 군집 다양성 분석 결과, 각 수계 및 지점을 구성하는 종의 다양성이 높아 군집이 특정 구조를 형성하지 못하고 비구조화 경향을 보임(그림 8). 이에 수계별

분석을 4계절로 분리하여 심화 분석을 수행하였으나, 계절에 영향을 받지 않고도 군집이 특정 구조를 형성하지 못하는 비구조화 경향을 동일하게 보임(그림 9)

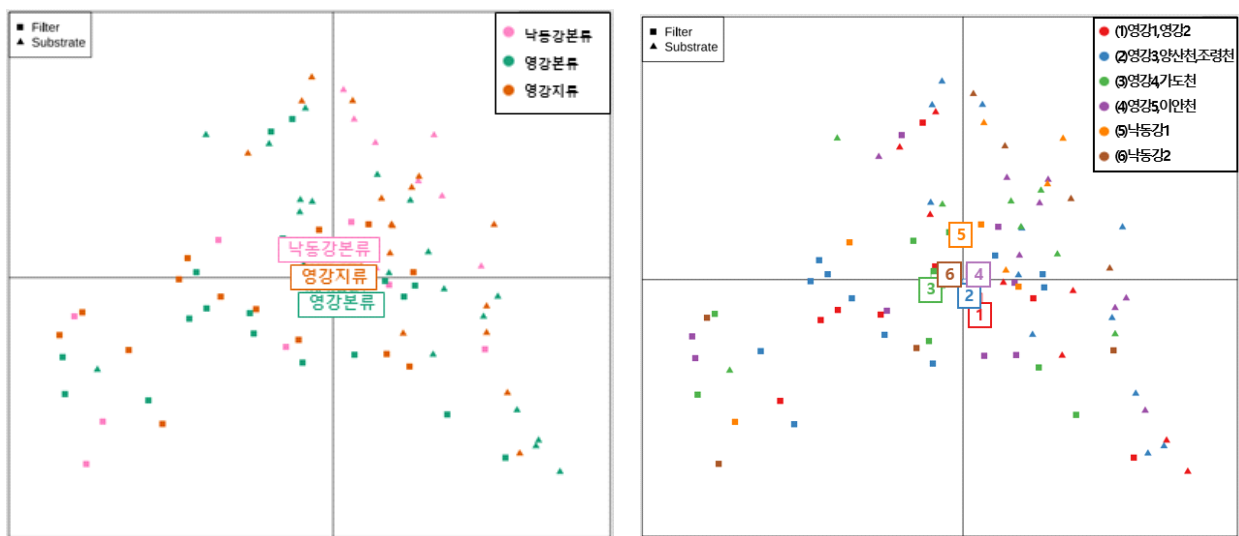
- 향후 하천의 연속성에 따른 군집 다양성의 비교 분석을 위해서는 환경유전자가 대표할 수 있는 지점의 범위를 반영한 지점 선정 등의 방법론적인 보완이 필요할 것으로 판단됨



<속(genus) 수준의 군집 다양성 분석>

<종(species) 수준의 군집 다양성 분석>

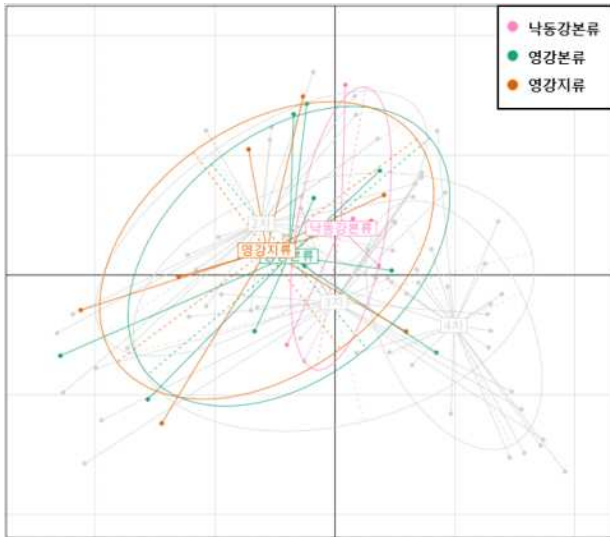
그림 7. 환경유전자 정보 계절별 군집 다양성 분석



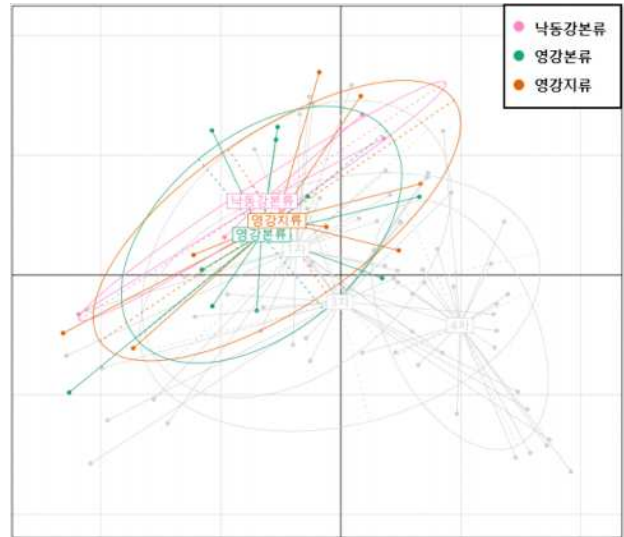
<수계별 군집 다양성 분석>

<지점별 군집 다양성 분석>

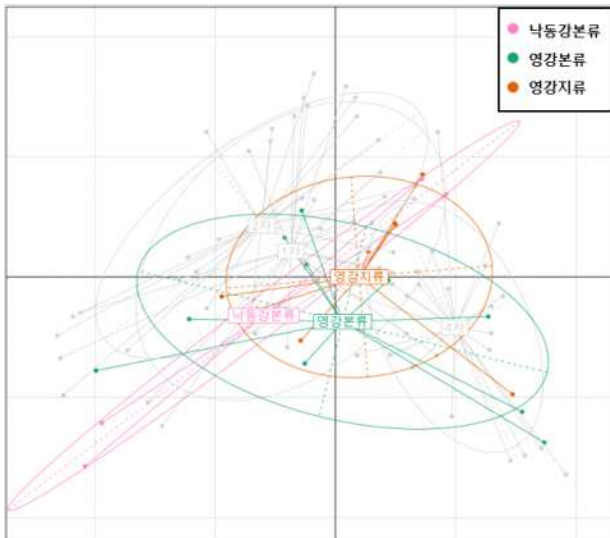
그림 8. 환경유전자 정보 수계별·지점별 군집 다양성 분석



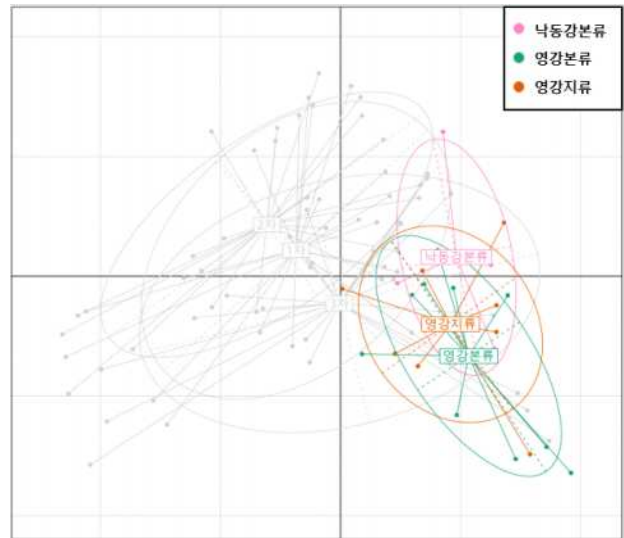
<겨울-수계별 군집 다양성 분석>



<봄-수계별 군집 다양성 분석>



<여름-수계별 군집 다양성 분석>



<가을-수계별 군집 다양성 분석>

그림 9. 환경유전자 정보 수계별(세분화) 군집 다양성 분석

* 조사시기 index 정보 : 1차(겨울), 2차(봄), 3차(여름), 4차(가을)

(5) 환경유전자 기반 주요 분류군 이중 유사도 분석(그림 10)

- Two-way clustering 방법으로 분류군별, 계절별, 조사지점별 유사도를 분석함
- 계절별 이중 유사도 분석 결과, 봄과 가을에 일부 절지·환형·선형·편형동물에서 높은 이중 유사도로 탐지되었으며, 연체동물은 계절에 따른 뚜렷한 경향성이 없이 4계절 고르게 탐지됨
- 수계별·지점별 이중 유사도 분석 결과, 수계별·지점별로 이중 유사도가 높은 분류군의 경향성이 뚜렷하게 나타나지 않고 다양한 분류군이 탐지됨

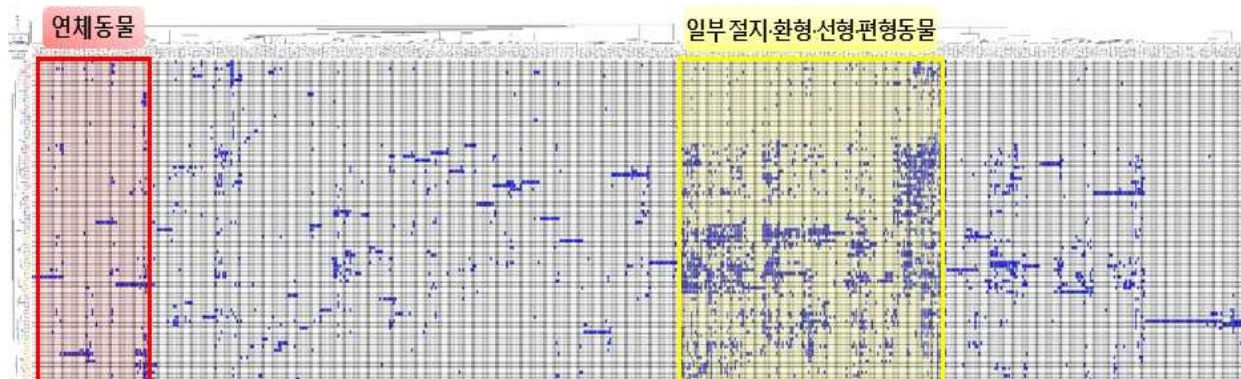


그림 10. 환경유전자 정보 분류군 이중 유사도 분석

* 종 탐지 비율이 높아질수록 해당 종의 위치에 **파란색**의 농도가 진해짐

- (6) 환경유전자 정보를 이용하여 주요 담수생물을 탐지하고 계절별, 수계별, 지점별 생물다양성과 군집 구조 분석의 가능성을 확인하였으나, 탐지 효율과 분석의 신뢰도를 높이기 위해서는 국내 서식종을 대상으로 지속적인 환경유전자 데이터베이스 구축(유전 정보 확보, 분류체계 정리 등), 세부적인 연구 방법론 보완(지점간 거리, 하폭에 따른 채수량 등)이 필요함

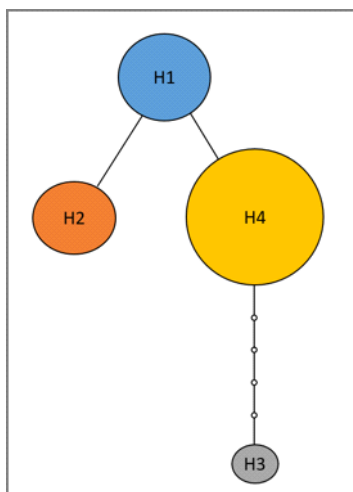
2. 주요 관심생물 탐지 및 유전다양성 분석

가. 낙동강 본류 및 지류 11지점에서 확보한 대용량 환경유전자 정보(10.7G, Read count 53,543,890건, ASV 133,466건)로부터 관심생물에 해당하는 데이터를 추출하여 집단별 유전다양성 분석에 사용하는 연구 방법 적용

나. 주름다슬기(*Semisulcospira forticosta*) 탐지 및 유전다양성 분석(그림 11, 표 9-10)

- (1) 11지점 중 5지점*에서 주름다슬기의 서식이 탐지되었으며, 총 5개 집단에서 주름다슬기 mtDNA-CO1 유전자(313bp) 정보 83개 확보
 - * 영강1(21YG1), 양산천(21YSC1), 조령천(21JRC1), 영강3(21YG3), 가도천(21GDC1)
- (2) TCS 프로그램을 이용하여 유전자형 network 분석 결과, 주름다슬기에서 총 4개의 유전자형(Haplotype)이 도출됨(그림 11)
- (3) 영강1, 양산천, 조령천, 영강3 지점에서는 각각 1개의 유전자형이 관찰되었고, 유전자형 다양도(Hd , Haplotype diversity) 및 염기 다양도(π , Nucleotide diversity)가 모두 낮게 관찰됨
- (4) 가도천 지점에서는 총 3개의 유전자형이 관찰되었고, 양산천, 조령천, 영강3 지점의 유전자형을 모두 공유하고 있었으며, H3 유전자형은 가도천에서만 확인된 고유 유전자형으로 확인됨
- (5) 가도천의 유전자형 다양도(Hd) 및 염기 다양도(π)는 각각 0.668 ± 0.002 와 0.010 으로 다른 4개 지점에 비해 상대적으로 높게 관찰되었음(표 9)

- (6) 각 지점 간 유전적 분화지수(F_{ST}) 측정 결과, 영강1, 조령천과 양산천, 영강3 지점 간에 유전적 차이가 높게 나타났으나, 가도천과 나머지 4개 지점 간에서는 상대적으로 낮은 유전적 차이가 나타났으며(표 10), 이는 가도천이 다른 지점에서 확인된 유전자형을 공유하고 있기 때문으로 사료됨
- (7) 지점별 유전다양성 분석 결과, 가도천은 공간적으로 하류에 위치하여 상류 지점에 서식하는 집단들의 유전자 유입, 농수 확보를 위한 하천수(펌프 등) 사용으로 타 지점의 개체 유입 등의 영향에 의해 다른 지점에 비해 상대적으로 높은 유전적 다양성을 유지하고 있는 것으로 추정되나(그림 11, 표 9-10), 향후 양산천, 조령천, 영강 지점의 추가적인 유전자 정보 확보를 통해 세부적인 연구 수행이 필요함



<유전자형 network>



<집단별 유전자형 분포 양상>

그림 11. 주름다슬기 유전자형 network 및 분포

- * 유전자형 network : 하나의 원은 각기 다른 유전자형을 표현(유전자형별로 다른 색상으로 표현). 숫자는 유전자형 번호. 원의 크기는 유전자형 빈도와 비례
- * 유전자형 분포 : 파이그래프는 지역별 각 유전자형의 빈도 분포를 나타냄

표 9. 주름다슬기 지점별 유전자 다양성 지수

지점	No. ASV	No. Haplotype	Haplotype diversity(Hd)	Nucleotide diversity(π)
영강1	36	1	0.000±0.000*	0.000*
양산천	4	1	0.000±0.000*	0.000*
조령천	6	1	0.000±0.000*	0.000*
영강3	11	1	0.000±0.000*	0.000*
가도천	26	3	0.668±0.002	0.010
Total	83	4	0.696±0.001	0.001

* 집단이 하나의 유전자형으로 구성되어 지수값이 0으로 측정된 경우

표 10. 주름다슬기 지점 간 유전적 차이

F_{ST}	영강1	양산천	조령천	영강3	가도천
영강1		+	+	+	+
양산천	1.00*		+	-	-
조령천	1.00*	1.00*		+	-
영강3	1.00*	0.00*	1.00*		+
가도천	0.46	0.09	0.16	0.19	

* 비교 집단이 각각 하나의 유전자형으로 구성되어 F_{ST} 값이 0 또는 1로 측정된 경우

다. 왕우렁이(*Pomacea canaliculata*) 탐지 및 유전다양성 분석(그림 12, 표 11-12)

- (1) 11지점 중 4지점*에서 왕우렁이의 서식이 탐지되었으며, 총 4개 집단에서 왕우렁이 mtDNA-CO1 유전자(313bp) 정보 405개 확보

* 조령천(21JRC1), 가도천(21GDC1), 이안천(21IAC1), 낙동강1(21NDG1)

- (2) TCS 프로그램을 이용하여 유전자형 network 분석 결과, 왕우렁이에서 총 4개의 유전자형(Haplotype)이 도출됨(그림 12)
- (3) 이안천, 낙동강1의 경우 각각 1개의 유전자형이 관찰되었고, 유전자형 다양도(Hd , Haplotype diversity) 및 염기 다양도(π , nucleotide diversity)가 모두 낮게 관찰됨(표 11)
- (4) 조령천 지점에서는 총 4개의 유전자형이 관찰되었고, 이안천, 낙동강1, 가도천 지점의 유전자형을 모두 공유하고 있었으며, 유전자형 다양도 및 염기 다양도는 각각 0.606 ± 0.0003 과 0.021 로 다른 2개 지점에 비해 가장 높게 관찰되었고, 가도천이 그 다음으로 높게 관찰됨($Hd=0.272 \pm 0.008$, $\pi=0.011$)
- (5) 각 지점 간 유전적 분화지수(F_{ST}) 측정 결과, 이안천, 낙동강1, 가도천, 조령천 지점 간에 유전적 차이가 $0.0-0.2$ ($P\text{-value} < 0.01$)으로 4개 지점 간에서는 낮은 유전적 분화가 나타났음(표 12). 이와 같은 지점별로 확인된 상대적으로 낮은 유전다양성 분석 결과는 외래종이 이입된 이후 창시자 집단의 급격한 개체군 성장에서 통상적으로 발견되는 높은 유전적 다양성의 예측과는 다른 것임. 이는 창시자 집단의 개체군 성장 이후, 주변 서식지 환경의 변화로 인해서 조사된 집단의 개체군이 병목현상(population bottle-necking)을 경험한 결과로서 해석될 수 있음(그림 12, 표 11-12). 따라서 조사지역에서의 서식 환경 변화와 왕우렁이 개체군의 크기 변화의 상관성을 규명하기 위해서는 조사 지역에 서식하고 있는 왕우렁이 개체군의 집단유전학적 연구 수행이 추가적으로 진행되어야 할 것으로 사료됨

표 11. 왕우렁이 지점별 유전자 다양성 지수

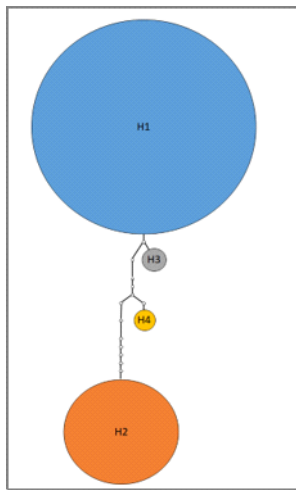
지점	No. ASV	No. Haplotype	Haplotype diversity(Hd)	Nucleotide diversity(π)
조령천	325	4	0.606 ± 0.0003	0.021
가도천	32	2	0.272 ± 0.008	0.011
이안천	41	1	$0.000 \pm 0.000^*$	0.000^*
낙동강1	7	1	$0.000 \pm 0.000^*$	0.000^*
Total	405	4	0.552 ± 0.0004	0.019

* 집단이 하나의 유전자형으로 구성되어 지수값이 0으로 측정된 경우

표 12. 왕우렁이 지점 간 유전적 차이

F_{ST}	조령천	가도천	이안천	낙동강1
조령천		+	+	+
가도천	0.09		+	-
이안천	0.24	0.15		-
낙동강1	0.19	0.02	0.00*	

* 비교 집단이 각각 하나의 유전자형으로 구성되어 F_{ST} 값이 0 또는 1로 측정된 경우



<유전자형 network>



<집단별 유전자형 분포 양상>

그림 12. 왕우렁이 유전자형 network 및 분포

- * 유전자형 network : 하나의 원은 각기 다른 유전자형을 표현(유전자형별로 다른 색상으로 표현). 숫자는 유전자형 번호. 원의 크기는 유전자형 빈도와 비례
- * 유전자형 분포 : 파이그래프는 지역별 각 유전자형의 빈도 분포를 나타냄

라. 환경유전자 정보를 이용하여 특정 종을 탐지하고 종내 집단 수준의 유전다양성 분석의 가능성을 확인하였으나, 개체군 분석에 필요한 유전자 정보 수를 확보하기 위한 방법론(분류군 공통 유전자 마커 적용 후, 특정 종 맞춤형 유전자 마커 병행 연구 수행 등)이 개선될 필요성이 있다고 판단됨

3. 환경유전자 데이터베이스 및 파이프라인 구축

가. 환경유전자 바코드 데이터 수집 및 정제(929,649건, 표 13)

- (1) 공개 데이터 수집 및 정제 : 주요 4개 분류군(연체·절지·환형·선형동물)의 mtDNA-COI 및 nrDNA-18S 바코드 데이터 929,495건 수집

- (가) 공개 DB(NCBI, BOLD, SILVA)에서 다양한 분류 조건(분류체계, 유전자 영역 등)을 이용해 대용량 유전 정보를 수집
- (나) 분류 조건에 부합하는 데이터 및 sequence similarity에 따라 99% 이상 일치하는 sequence는 하나로 clustering하여 정제
- 정제 프로그램 CD-HIT clustering, 조건 sequence similarity 99% 적용
- (다) mtDNA-CO1 바코드 데이터(NCBI, BOLD) 연체동물 38,630건, 절지동물 95,910건, 선형동물 5,441건, 환형동물 14,132건 및 nrDNA-18S 바코드 데이터(BOLD, SILVA) 연체동물 2,000건, 절지동물 21,479건, 선형동물 6,971건, 환형동물 2,140건 확보
- (2) 자체 데이터 : 연체동물 17종, 절지동물 13종, 선형동물 13종에 대한 mtDNA-CO1 바코드 데이터 154건 확보

표 13. 환경유전자 바코드 데이터 수집 현황

(‘21.12. 기준)

구분	mtDNA-CO1(건/종)			nrDNA-18S(건/종)		Total(건)
	NCBI	BOLD	자체	BOLD	SILVA	
연체동물	2,520건/1,668종	36,110건/12,915종	76건/17종	98건/58종	1,902건/1,525종	929,649건
절지동물	13,144건/8,685종	82,766건/13,778종	39건/13종	1,129건/315종	20,350건/14,775종	
선형동물	227건/161종	5,214건/1,909종	39건/13종	2건/1종	6,969건/2,873종	
환형동물	276건/154종	13,856건/3,368종	-	37건/25종	2,103건/1,469종	

나. 환경유전자 분석 파이프라인 구축

- (1) 환경유전자 정보 기반의 생물다양성 분석 연구에 특화된 파이프라인 구축을 위해, 4단계 분석 프로세스(Quality filtering & assembly, Clustering, Taxonomy assignment, Biodiversity statistics) 기반 구축(표 14, 그림 13)

표 14. 환경유전자 정보 기반 생물다양성 분석 흐름도

Process		Program		Result
Assembly		DADA2		Assembled reads(Se,fastq)
Pre-processing		DADA2		Removed noise data
Clustering		DADA2		ASV
Diversity Analysis	Alpha-diversity	In-house program using R package	alpha_diversity.py	ASVs, Chao1, Shannon, Simpson etc
	Taxonomy-diversity		BLAST/UCLSUT	Taxonomy Composition
	Alpha-rarefaction		alpha_rarefaction.py	Rarefaction curve graph
	PCoA		make_2d_plots.py	PCoA graph(2D, 3D)
	UPGMA Tree		upgma_cluster.py	UPGMA Tree graph
Heatmap		make_otu_heatmap_html.py		ASV Heatmap

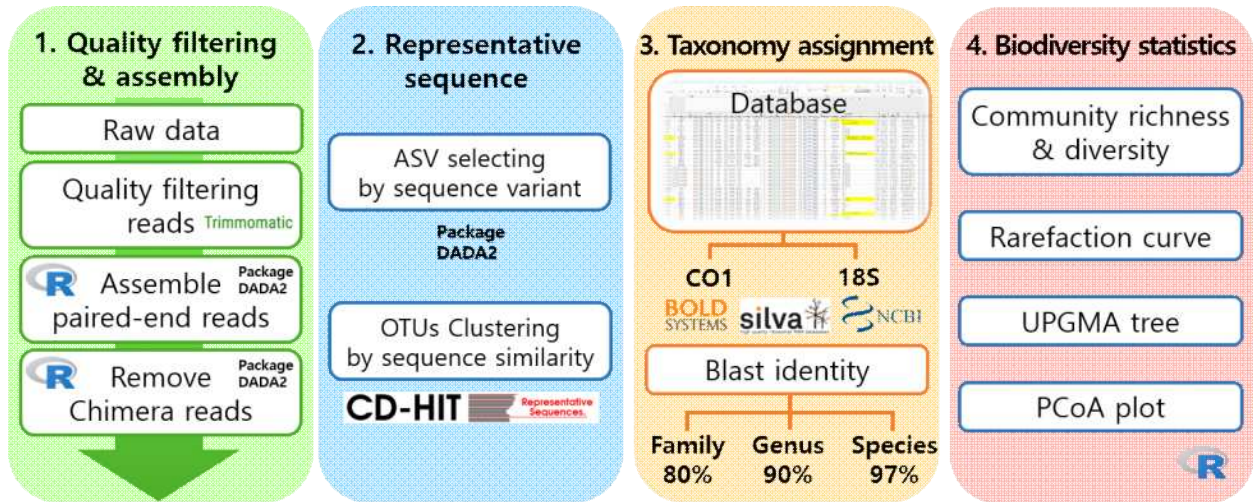


그림 13. 환경유전자 정보 기반 생물다양성 분석 파이프라인

(2) 1단계 Quality filtering & assembly

(가) NGS 분석을 통해 확보한 대용량 Raw data 정제 및 조립

(나) 상세 프로세스(사용 프로그램) : Raw data → Quality filtering reads(Trimmomatic, Bolger et al., 2014) → Assemble paired → end reads(DADA2, Callahan et al., 2016) → Remove Chimera reads(DADA2, Callahan et al., 2016)

(3) 2단계 Clustering

(가) 조립된 데이터로부터 대표 서열 선별(OTU 또는 ASV 분석 방법 적용, 그림 14)

(나) OTU 분석 방법(clustering by sequence similarity)

- 1) OTU(Operational Taxonomic Unit)는 PCR error를 감안한 cluster로 분류 단위를 생성하는 것이 목적
- 2) Raw data를 similarity(97%, 99%)를 기준으로 유사한 서열끼리 clustering하여 대표 서열을 선정하는 방법(CD-HIT, Fu et al., 2012)

(다) ASV 분석 방법(selecting by sequence variant)

- 1) ASV(Amplicon Sequence Variants)는 PCR 및 sequencing 과정에서 생성된 error 서열을 제거한 대표서열로 분류 단위를 생성하는 것이 목적
- 2) Raw data에서 PCR 및 sequencing 과정에서 생성된 error를 제거한 후, 복구된 서열을 기반으로 대표 서열을 선정하는 방법(DADA2, Callahan et al., 2016)

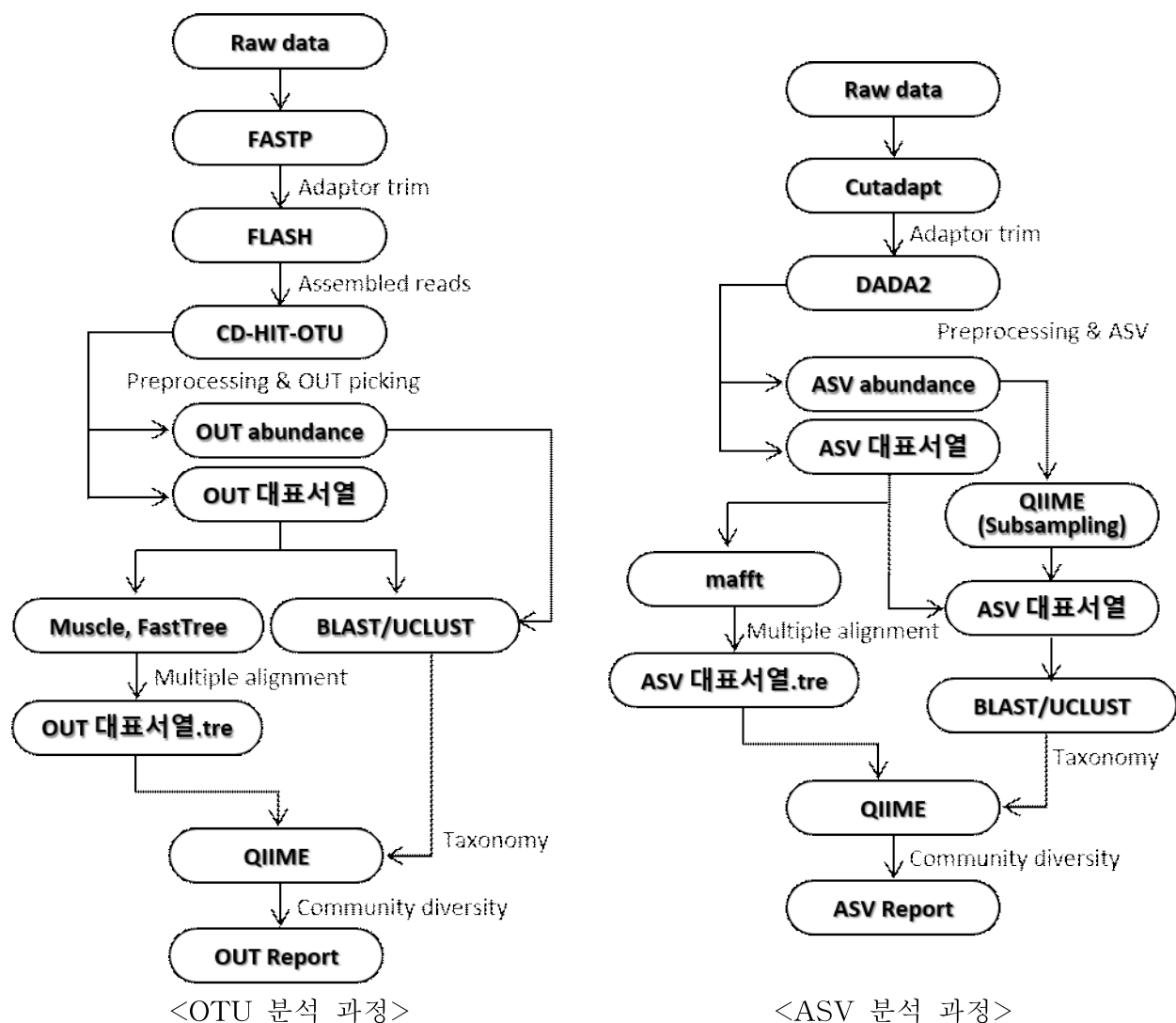


그림 14. 환경유전자 정보 대표 서열 선별 분석 방법 비교

(4) 3단계 Taxonomy assignment

(가) 선별된 대표 서열의 reference 기반 종 판별 및 목록 구축

(나) 분류 수준에 적합한 Blast identity 적용(Family 80%, Genus 90%, Species 97%)

※ 분류군 특성에 따라 Blast identity 수정 필요

(다) Blast 결과에 따른 종 목록 구축

(5) 4단계 Biodiversity statistics

(가) 구축된 종 목록을 기반으로 다양한 생물다양성 분석 수행

(나) 세부 분석 방법(사용 프로그램) : Alpha-diversity(Dominance index, Diversity index, Evenness index, Richness index), Alpha-rarefaction, PCoA, UPGMA Tree, Heatmap 등(R, R Core Team, 2020)

4. 환경유전자 정보 기반 생물다양성 분석 매뉴얼 작성

가. 제목 : 환경유전자 기반의 생물 종 다양성 분석 매뉴얼(수서곤충편)

나. 목적 : 최신 연구 방법(환경유전자, 메타바코딩 등)을 활용하여 환경을 훼손하지 않고 생물 탐지 및 종 다양성 분석을 할 수 있는 연구 기법을 정리한 지침서로 실전에 적용할 수 있는 표준화된 연구 기법 제공

다. 주요내용 : 환경유전자 연구의 전 과정에 대한 수행 절차, 방법(시약·재료, 장비·물품 등), 결과 예시 등 포함(그림 15)

(1) 환경유전자 시료 채집 및 전처리 방법

(2) 환경유전자 추출 및 유전 정보 확보 방법

(3) 생물다양성 분석 방법

- 대용량 환경유전자 정보 정제 방법
- 환경유전자 대표 서열 선정 방법
- 자체 환경유전자 데이터베이스 기반 담수생물 종 목록 작성
- 생물다양성 종합 분석 방법

환경유전자 기반의 생물 종 다양성 분석 매뉴얼 - 수서곤충편 - 2021. 12.  국립낙동강생물자원관	목차 I. 개요 / 2 1. 목적 2 2. 활용 분야 2 II. 환경유전자 시료 채집 및 전처리 방법 / 5 1. 채집 방법 5 2. 전처리 방법 7 III. 환경유전자 추출 및 유전 정보 확보 방법 / 9 1. 환경유전자 추출 방법 9 2. 유전자 마커 적용 방법 10 3. 대용량 환경유전자 정보 확보 방법 12 IV. 생물다양성 분석 방법 / 14 1. 대용량 환경유전자 정보 정제 방법 14 2. 환경유전자 대표 서열 선정 방법 16 3. Custom DB 기반 담수생물 종 목록 작성 18 4. 생물다양성 종합 분석 방법 20 V. 서식 / 22 1. 대용량 환경유전자 정보량표 21 2. 생물다양성 분석 결과표 22	I. 개요 1. 목적 환경유전자 정보를 기반으로 생물다양성을 분석하기 위한 표준화된 채집 및 분석 방법 제공을 목적으로 함  환경유전자 채집 → 환경유전자 추출 → 유전 정보 확보 → 환경유전자 분석 → 생물 종 다양성 분석 환경유전자 기반의 생물 종 다양성 분석 [기본 개념] · 환경유전자란? 환경유전자(Environmental DNA, eDNA)는 생물로부터 직접 유전자를 추출하는 것이 아니라 담수, 해수, 토양, 공기 등 다양한 환경에 부유하고 있는 생물의 조직에서 채취한 유전자를 증명 
--	--	--

그림 15. 환경유전자 기반의 생물 종 다양성 분석 매뉴얼(수서곤충편)

IV. 종합 결론

1. 기존의 생물탐지 및 생물다양성 분석 연구를 보완하기 위해, 본 연구에서는 환경유전자 정보를 활용한 차세대 생물탐지 및 생물다양성 분석 기법을 개발하고자 ① 환경유전자 정보 확보 및 생물다양성 분석, ② 관심생물 탐지 및 유전다양성 분석, ③ 환경유전자 데이터베이스 및 파이프라인 구축을 통해 최종적으로 환경유전자 정보 기반의 생물다양성 분석 매뉴얼(수서곤충편)을 작성함

2. 환경유전자 정보 확보 및 생물다양성 분석

가. 주요 내용

- 낙동강 본류 및 지류 11지점을 대상으로 계절별(봄, 여름, 가을, 겨울), 수심별(중층, 저층) 조사를 실시하여 총 176개 환경유전자 시료를 확보함
- 환경유전자 DNA 추출, 유전자 마커 3건 적용 등을 통해 총 264건의 대용량 환경유전자 정보(정보량 10.7G, Read count 53,543,890건, ASV 133,466건)를 확보함
- ※ 유전자 마커 : mtDNA-CO1 1건, nrDNA-18S 2건(V1-2, V4 부위) 적용
- 확보한 대용량 환경유전자 정보 기반의 생물다양성을 분석한 결과, 주요 5개 분류군(연체·절지·환형·선형·편형동물)에서 총 20강 65목 229과 409속 604종의 담수생물이 확인됨
- ※ 분류군별, 계절별, 지점별 생물다양성을 비교한 결과, 분류군별로 절지동물이 358종, 계절별로 봄이 264종, 지점별로 영강1이 155종으로 가장 높은 종수가 탐지됨
- 생물조사방법에 따른 생물다양성 비교·분석(환경유전자 연구 vs 생물측정망 연구) 결과, 환경유전자 연구에서 더 높은 생물다양성이 측정되었으며, 주로 현장 또는 실험실에서 종 동정이 어려운 미소생물과 분류체계 정립이 필요한 담수생물이 대표적으로 탐지됨. 두 생물조사방법 사이에는 분류군 과(Family) 수준에서 60%의 높은 탐지 일치율(공통 탐지 39과/생물측정망 64과)을 보였으나, 하위 분류군 수준으로 내려갈수록 탐지 일치율이 낮아지는 경향을 보임
- 분류군 종수 탐지 기대치 분석 결과, 계절별로 봄이 253.0 ± 2.6 종(유전 정보량 201,000건 기준), 수계별로 영강 지류에서 281.5 ± 1.8 종(유전 정보량 36,000건 기준), 지점별로 영강1(21YG1)에서 151.4 ± 1.6 종(유전 정보량 93,000건 기준)으로 가장 높은 종수 탐지 기대치가 측정됨
- 분류군 군집 다양성 분석 결과, 계절별로 군집 구조가 독립적으로 형성되었으며, 수계별·지점별로는 종 다양성이 높아 군집이 특정 구조를 형성하지 못하고 비구조화 경향을 보임

- 분류군 이중 유사도 분석 결과, 계절별로 일부 절지·환형·선형·편형동물에서 높은 이중 유사도로 탐지되었으며, 수계별·지점별로는 이중 유사도가 높은 분류군의 경향성이 뚜렷하게 나타나지 않고 다양한 분류군이 탐지됨

나. 고찰

- 환경유전자 정보를 이용하여 주요 담수생물을 탐지하고 계절별, 수계별, 지점별 생물다양성과 군집 구조 분석(Rarefaction curve, PCoA, Two-way clustering 분석 등)의 가능성을 확인함
- 또한 특정 서식 분류군의 낮은 탐지 효율, 분류체계 혼란으로 인한 종 판별 오류, 환경유전자 시료 확보 시 각 지점의 현장 상태를 반영한 조사 방법 적용 등 환경유전자 연구의 보완점이 확인됨
- 이에 환경유전자 정보 기반의 생물 탐지 효율과 분석의 신뢰도 향상을 위해서는 국내 서식종을 대상으로 지속적인 환경유전자 데이터베이스 구축(분류체계 정리, 다양한 유전자 부위의 유전 정보 확보 등), 세부적 연구 방법론 보완(환경유전자가 대표할 수 있는 지점의 범위를 고려한 지점 선정, 하천 규모에 따른 채수량 등) 등 후속 연구가 필요함

3. 주요 관심생물 탐지 및 유전다양성 분석

가. 주요 내용

- 낙동강 본류 및 지류 11지점에서 확보한 대용량 환경유전자 정보(10.7G, Read count 53,543,890건, ASV 133,466건)로부터 관심생물에 해당하는 데이터를 추출하여 집단별 유전다양성 분석에 사용하는 연구 방법을 적용함
- 주름다슬기는 11지점 중 5지점에서 주름다슬기의 서식이 탐지되었으며, 총 5개 집단에서 확보한 주름다슬기 mtDNA-CO1 유전자(313bp) 정보 83개에서 총 4개의 유전자형이 도출됨. 가도천 지점에서는 총 3개의 유전자형이 관찰되었고, 양산천, 조령천, 영강3 지점의 유전자형을 모두 공유함. 가도천의 유전자형 다양도 및 염기 다양도는 각각 0.668 ± 0.002 와 0.010으로 다른 4개 지점에 비해 높게 관찰됨. 영강1, 조령천과 양산천, 영강3 지점 간에 유전적 차이가 높게 나타났으나, 가도천과 나머지 4개 지점 간에서는 낮은 유전적 차이가 나타남. 주름다슬기의 지점별 유전다양성 분석 결과는 공간적으로 하류에 위치하는 가도천으로 상류지점에 서식하는 집단들의 유전자 유입, 농수 확보를 위한 하천수(펌프 등) 사용으로 타 지점의 개체가 유입된 영향으로 판단됨
- 왕우렁이는 11지점 중 4지점에서 왕우렁이의 서식이 탐지되었으며, 총 4개 집단에서 확보한 왕우렁이 mtDNA-CO1 유전자(313bp) 정보 405개에서 총 4개의 유전자형이 도출됨. 조령천 지점에서는 총 4개의 유전자형이 관찰되었고, 이안천, 낙동강1, 가도천 지점의 유전자형을 모두 공유함. 유전자형 다양도 및 염기 다양도는 각각 0.606 ± 0.0003 과 0.021로 다른 2개 지점에 비해 가장 높게 관찰됨. 이안천, 낙동강1, 가도천, 조령천 지점

간에 유전적 차이가 0.0-0.2(P-value<0.01)으로 4개 지점 간에서는 낮은 유전적 차이가 나타남. 왕우렁이의 지점별 유전다양성 분석 결과는 외래종, 이입종 등의 유전다양성 경향성과 유사한 것으로 판단되며, 왕우렁이가 도입될 당시 유전자풀(창시자 집단)이 작은 상태로 도입된 후 우리나라에 정착하면서 급격하게 개체군 성장이 일어난 것으로 판단됨

나. 고찰

- 환경유전자 정보를 이용하여 특정 종 탐지 및 집단 유전다양성 분석의 가능성을 확인함
- 또한 개체군 분석에 필요한 유전 정보 확보, sequencing error 비율 적용 등 환경유전자 연구의 보완점이 확인됨
- 이에 환경유전자 정보 기반의 관심생물 탐지 및 유전다양성 분석의 신뢰도 향상을 위해서는 개체군 분석에 필요한 유전자 정보 수 확보를 위한 방법론(공통 유전자 마커, 특정 종 맞춤형 유전자 마커 혼합 적용 등) 개선, 유전다양성 분석 시 sequencing error의 적용 방법 등 후속 연구가 필요함

4. 환경유전자 데이터베이스 및 파이프라인 구축

가. 주요 내용

- 주요 4개 분류군(연체·절지·환형·선형동물)의 mtDNA-CO1 및 nrDNA-18S 바코드 데이터 929,649건을 수집하고 정제함
 - ※ 공개 DB(NCBI, BOLD, SILVA)에서 주요 4개 분류군(연체·절지·환형·선형동물)의 mtDNA-CO1 및 nrDNA-18S 바코드 데이터 929,495건 수집
 - ※ 자체적으로 연체동물 17종, 절지동물 13종에 대한 mtDNA-CO1 바코드 데이터 154건 확보
- 환경유전자 정보 기반의 생물다양성 분석 연구에 특화된 파이프라인 구축을 위해, 분석 프로그램 선정 및 안정화 검증을 수행한 후, 환경유전자 분석 맞춤형 파이프라인 기반을 구축함. 4단계 분석 프로세스는 Quality filtering & assembly, Clustering, Taxonomy assignment, Biodiversity statistics로 구성됨

나. 고찰

- 환경유전자 연구에 가장 핵심이 되는 환경유전자 데이터베이스 및 분석 파이프라인 기반을 구축함
- 환경유전자 정보 기반의 생물다양성 분석을 수행하기 위해 다양한 형식으로 산재되어 있는 대용량 데이터를 수집하고 분석 reference로 활용하기 위한 정제 과정을 수행함. 지속적인 표준화된 데이터 축적, 생물정보학 비전문가도 쉽게 다룰 수 있는 데이터베이스 구축 및 운영이 필요함
- 또한 환경유전자 연구에 특화된 생물다양성 분석 파이프라인의 기반을 구축하였으며,

세부 분류군의 유전적 특성을 반영한 복합 분석을 수행하기 위한 분석 파이프라인 고도화 과정이 필요함

5. 환경유전자 정보 기반 생물다양성 분석 매뉴얼 작성

가. 주요 내용

- 환경유전자로부터 생물다양성을 확인하는 연구 전 과정에 대한 방법론(1식, 수서곤충편)을 작성함
- 환경유전자 시료 채집 및 전처리 방법, 환경유전자 추출 및 유전자 마커 적용 방법, 생물다양성 분석 방법 등에 대한 내용이 포함됨

나. 고찰

- 당해연도 연구 방법과 결과를 기반으로 환경유전자 연구에 적용할 수 있는 표준화된 생물다양성 분석 매뉴얼(수서곤충편)을 작성함
- 후속 연구를 통해 보완될 환경유전자 정보 확보 및 생물다양성 분석 방법(환경유전자 조사 방법, 분류군별 특화된 연구 방법론 등), 관심생물 탐지 및 유전다양성 분석 방법, 환경유전자 데이터베이스 및 분석 파이프라인 구축 등의 내용을 해당 매뉴얼에 지속적으로 수정·보완하는 매뉴얼 고도화 과정이 필요함

V. 연구결과 활용 계획

1. 환경유전자 연구 분야의 학술적 발전을 위한 핵심 자료 제공

- 가. 환경유전자 연구는 전 세계적으로 2010년 이후 다양한 분류군에서 활용하고 있으나, 국내에서는 일부 국가 연구기관과 대학을 중심으로 초기 단계의 연구를 수행 중
- 나. 본 연구를 통해 확보한 대용량 환경유전자 정보와 생물다양성 분석 결과로 국내 환경유전자 연구 분야의 학술적 발전에 기여
 - 환경유전자 연구 결과를 저명한 저널에 게재하고 학회에 발표하여, 관련 분야에서 우리관의 입지 확보 및 국제적 공인
 - 환경유전자 연구 방법의 표준화를 위한 매뉴얼 작성에 핵심 자료로 적극 활용(국립생물자원관 협업)(국립생물자원관, 2020)

2. 정부 정책 수립 및 추진을 위한 유전자 분야 근거 자료 제공

- 가. 담수생물의 보전·관리를 위한 정책 수립에 활용
 - 실시간 생물 탐지를 통해 보전·관리가 우선적으로 필요한 종 선정 및 관리 지점 선정
 - 단시간 광범위한 지점의 생물다양성 측정을 통해 지점별 유전적 건강성 평가 수행
- 나. 최근 정부의 ‘디지털 뉴딜’ 추진에 따른 D.N.A.(Digital·Network·AI) 생태계 강화를 위한 유전자 빅데이터의 활용성 제고
 - 전 세계적으로 공개된 유전자 데이터 및 국가 사업으로 확보된 비공개 데이터 등 방대한 유전자 데이터가 넘쳐나고 있으나, 빅데이터의 표준화된 관리 부재로 다양한 유전 연구에 즉시 적용하는 데 한계가 있음
 - 본 연구를 통해 산재된 유전자 빅데이터를 수집하고 양식을 표준화한 정제 과정으로 구축한 환경유전자 데이터베이스는 다양한 환경유전자 분석 연구에 활용 가능
 - 국가 사업을 통해 확보된 유전자 빅데이터의 공개를 통한 공공성 확보 예정
 - ※ NCBI, BOLD 등 공개 DB 등록을 통한 우리나라 유전자원 선점
 - ※ 생물자원관 협의체(환경부 국립생물자원관, 국립낙동강생물자원관, 국립호남권생물자원관)를 통한 유전자 정보 표준화 및 기간 관 유전자 정보 공유 추진 중

3. 환경부 및 지자체의 현안문제(환경 피해 또는 인적·경제적 피해) 대응 전략 수립을 위한 과학적 증거 마련

- 가. 담수생물의 대발생, 사멸 등의 현안문제 발생 시, 생물 기초정보(종 정보, 분포 현황 등) 미흡으로 원인 파악과 대응이 어려운 실정임

나. 본 연구를 통해 연체동물 2종(주름다슬기, 왕우렁이)을 대상으로 환경유전자 정보를 활용한 신속한 생물 탐지 방법론의 가능성을 확인함

다. 해당 방법론은 담수 생물 종의 변화 추이를 실시간으로 모니터링 할 수 있는 새로운 개념의 생물탐지 및 관리에 관한 표준화된 기술 개발*에 활용 가능

- * 담수 생태계를 대표하는 지표종, 생태계에 영향을 주는 생태계교란종·위해종·외래종, 보전이 필요한 멸종위기종 등 국내 서식에 대해 관심이 필요한 담수생물의 서식 여부를 신속·정확하게 탐지할 수 있는 자동화 시스템. 형태적 동정을 생략하고 다수의 지역에 대한 탐지를 단시간에 수행할 수 있으며, 이를 위해서는 특정 생물에 대한 유전자 마커 개발 및 자생종 대상의 다양한 유전 정보 확보가 필수임

4. 환경유전자 정보를 이용한 담수생물다양성 평가의 가능성 제시

가. 전통적인 생물다양성 평가 방법은 분류군별 조사 방법, 생물다양성 지수 산정 및 평가 방법이 상이하여 담수생물의 통합 평가에 어려움이 있음

나. 본 연구를 통해 확보한 대용량 환경유전자 정보는 담수생태계에 서식하는 다양한 분류군 대상의 생물다양성 평가 지수 개발에 활용 가능

- 다양성 기대값, 종 탐지 특성, 먹이망 구조 분석 등으로 생물다양성(종·유전·생태) 평가를 위한 다변량 메트릭 구축
- 메트릭별 생물다양성 보전을 위한 의미를 부여하여 기존에 활용 중인 평가 시스템의 보완

다. 국가생물축정망을 이용한 건강성 평가의 보완을 위한 기초 자료 제공 가능

- 본 연구의 주요 분류군을 포함하는 저서성대형무척추동물 평가지수(BMI)는 오타치(s)와 지표가중치(g)의 중요도가 높음
- 연체·유선형·편형동물 등 서식환경에 따른 분포 범위를 환경유전자 정보 기반의 기초자료를 통해 지표가중치의 정확도 향상에 기여

VI. 참고문헌

- 국립낙동강생물자원관. 2020. 담수생물의 유전다양성 보전·관리 연구(Ⅱ).
- 국립생물자원관. 2020. 환경DNA를 이용한 수생태계 종다양성 분석 기반 연구.
- 국립생물자원관 한반도의 생물다양성(<https://species.nibr.go.kr/>).
- Andújar C, Arribas P, Gray C, Bruce C, Woodward G, Yu DW, Vogler AP. 2018. Metabarcoding of freshwater invertebrates to detect the effects of a pesticide spill. *Molecular Ecology*, 27(1): 146–166.
- Blaxter ML, De Ley P, Garey JR, Liu LX, Scheldeman P, Vierstraete A, Vanfleteren JR, Mackey LY, Dorris M, Frisse LM, Vida JT, Thomas WK. 1998. A molecular evolutionary framework for the phylum Nematoda. *Nature*, 392(6671): 71–75.
- Bohmann K, Evans A, Gilbert MTP, Carvalho GR, Creer S, Knapp M, Yu DW, de Bruyn M. 2014. Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology and Evolution*, 29(6): 358–367.
- Bolger AM, Marc L, Bjoern U. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15): 2114–2120.
- Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13(7): 581–583.
- Folmer O, Black M, Hoeh W, Lutz P, Vrijenhoek R. 1994. DNA Primers for Amplification of Mitochondrial Cytochrome C Oxidase Subunit I from Diverse Metazoan Invertebrates, *Molecular Marine Biology and Biotechnology*, 3: 294–299.
- Fu L, Niu B, Zhu Z, Wu S, Li W. 2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28(23): 3150–3152.
- Leray M, Yang JY, Meyer CP, Mills SC, Agudelo N, Ranwez V, Boehm JT, Machida RJ. 2013. A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral

- reef fish gut contents. *Frontiers in zoology*, 10(1): 1–14.
- Smith DR, Villella RF, Lemarié DP. 2003. Application of adaptive cluster sampling to low-density populations of freshwater mussels. *Environmental and Ecological Statistics*, 10(1): 7–15.
- Team R Core. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>
- Thomas AC, Goldberg CS, Howard J, Nguyen PL, Seimon TA. 2018. ANDeTM: A fully integrated environmental DNA sampling system. *Methods in Ecology and Evolution*, 9(6): 1379–1385.
- Thomsen PF, Kielgast J, Iversen LL, Wiuf C, Rasmussen M, Gilbert MTP, Willerslev E. 2012. Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular Ecology*, 21(11): 2565–2573.
- Wangensteen OS, Turon X. 2017. Metabarcoding techniques for assessing biodiversity of marine animal forests. *Marine animal forests. The ecology of benthic biodiversity hotspots*, 1: 445–503.
- Zhan A, Hulák M, Sylvester F. 2013. High sensitivity of 454 pyrosequencing for detection of rare species in aquatic communities. *Methods in Ecology and Evolution*, 4: 558 – 565.

주 의

1. 이 보고서는 국립낙동강생물자원관에서 시행한 연구과제 결과보고서입니다.
2. 이 보고서 내용을 발표할 때에는 반드시 국립낙동강생물자원관에서 시행한 연구과제의 연구결과임을 밝혀야 합니다.
3. 국가과학기술 기밀유지에 필요한 내용은 대외적으로 발표 또는 공개하여서는 아니 됩니다.
4. 이 보고서와 관련된 문의사항은 국립낙동강생물자원관 다양성연구팀(전화 054-530-0833)로 문의하시면 됩니다.