

극지 미세조류 유래 유용물질 탐색 및 대량배양

Investigation and mass production of functional materials
from polar microalgae



한 국 해 양 과 학 기 술 원
부 설 극 지 연 구 소

제 출 문

극지연구소장 귀하

본 보고서를 “극지 미세조류 유래 유용물질 탐색 및 대량배양” 과제의 최종보고서로 제출합니다.



“ : 민슬기

“ : 채현식

“ : 김은재

“ : 서대건

“ : 신승철

“ : 유지현

보고서 초록

과제관리번호	PE19270	해당단계 연구기간	2019. 02. 01~ 2019. 12.31	단계 구분	1/ 1
연구사업명	중 사 업 명	기관목적사업			
	세부사업명	KIST-KOPRI 협력사업			
연구과제명	중 과 제 명				
	세부(단위)과제명	극지 미세조류 유래 유용물질 탐색 및 대량배양			
연구책임자	김상희	해당단계 참여연구원수	총 : 8 명 내부 : 8 명 외부 : 명	해당단계 연구비	정부: 천원 기업: 천원 계: 10,000 천원
연구기관명 및 소속부서명	극지생명과학연구부		참여기업명	-	
국제공동연구	상대국명 :		상대국연구기관명 :		
위 탁 연 구	연구기관명 :		한국과학기술연구원	연구책임자 : 관철호	
요약				보고서 면수	54
<p>○ 전세계가 생물자원의 권리를 확보하기 위한 경쟁이 치열하나 이미 신규 자원 고갈로 지구상 미개척지로 눈을 돌리고 있는 추세임</p> <p>○ 후발주자들은 남북극 같은 극한지에서 새로운 종의 발굴과 기능성 탐색을 통해 유용자원의 권리 확보 및 관련 시장 장악을 위한 집중적 노력이 필요한 시점임</p> <p>○ 극한 환경에서 생물들이 만들어내는 천연 방어물질들은 고효능 신물질 가능성이 높아 학계와 산업계에 폭발적 관심과 대안으로 떠오르고 있음</p> <p>○ 그러나 극지 생물연구는 현재까지 주로 세균을 중심으로 연구되어 왔고 세균 다음으로 다양성이 높은 진핵 미세조류에 대한 연구는 초기 단계임</p> <p>○ 극지미세조류 발굴과 활용은 세계적 추세인 지속가능한 자원(sustainable resource)에 부합하며 신규자원개발과 환경문제에 대한 대응책을 제시할 수 있음</p> <p>○ 극지 미세조류의 낮은 성장속도와 저온성장 등 까다로운 배양 조건을 해결하고 충분한 바이오매스를 확보할 수 있다면 극지생물자원 활용 신산업 분야를 개척할 것으로 확신함</p> <p>○ 한국과학기술연구원과 협력 연구를 통해 신규 물질 발견 가능성을 높임으로써 연구 활성화와 산업적 접근이 가능해짐</p> <p>○ 극지연구소가 확보한 극지 미세조류에 대한 연구 활성화와 산업적 접근이 필요한 시점이며 발굴, 배양부터 활성물질 분리, 활성평가까지 체계적인 다학제 협력이 가능해짐</p> <p>○ 한국과학기술연구원은 극지연구소에서 확보하고 있는 극지 환경의 생물자원을 분양 받아 연구함으로써 신규 물질 발견 가능성을 높일 수 있음</p>					
색 인 어 (각 5개 이상)	한 글	극지 미세조류, 바이오매스, 전장 유전체, 배양조건탐색, 대량배양			
	영 어	Polar microalgae, Biomass, Whole-genome sequence, Screening of culture condition, Large-scale cultivation			

요 약 문

I. 제 목

극지 미세조류 유래 유용물질 탐색 및 대량배양

II. 연구개발의 목적 및 필요성

전세계적으로 생물자원의 권리를 확보하기 위한 경쟁이 치열하나 이미 신규 자원 고갈로 지구상 미개척지로 눈을 돌리고 있는 추세이다. 후발주자들은 남북극 같은 극한지에서 새로운 종의 발굴과 기능성 탐색을 통해 유용 자원의 권리 확보 및 관련 시장 장악을 위한 집중적 노력이 필요한 시점이다. 극한 환경에서 생물들이 만들어내는 천연 방어물질들은 고효능 신물질 가능성이 높아 학계와 산업계에 관심과 대안으로 떠오르고 있다. 그러나 극지 생물연구는 현재까지 주로 세균을 중심으로 연구되어 왔고 세균 다음으로 다양성이 높은 진핵 미세조류에 대한 연구는 초기 단계이다. 극지미세조류 발굴과 활용은 세계적 추세인 지속가능한 자원(sustainable resource)에 부합하며 신규자원개발과 환경문제에 대한 대응책을 제시할 수 있다. 극지 미세조류의 낮은 성장속도와 저온성장 등 까다로운 배양 조건을 해결하고 충분한 바이오매스를 확보할 수 있다면 극지생물자원 활용 신산업 분야를 개척할 것으로 기대한다. 한국과학기술연구원과 협력 연구를 통해 신규 물질 발견 가능성을 높임으로써 연구 활성화와 산업적 활용을 개진하고자 한다.

II. 연구개발 내용 및 범위

○난배양성 극지 균주 대량배양 기술 개발

- 극지균주에 맞는 광량을 고려한 광배양기를 제작하고, 개선된 광배양기를 기반으로 배양배지, 배양온도, 최적 광량 등을 스크리닝 함으로서 단기간 내 남극 미세조류의 목적 농도 달성
- 극지연구소 내에 약 27m³ 규모의 저온배양실 시범 운영 중이며 200L 규모의 배양장치의 자동화 설비 구축

○극지 미세조류의 전장 유전체 염기서열 분석

- WGS기술(Whole Genome Sequencing technology)을 통한 단백질 또는 ORF 서열 정보 획득

○극지 미세조류 이차대사산물의 클러스터분석

- 남극 유래 조류 *Micractinium* sp. KSF0031와 녹조류 근연종인 *Chlorella variabilis*의 이차대사산물 생합성 유전자 클러스터를 antiSMASH v4.2.0

- (antibiotics and secondary metabolite analysis shell)를 이용하여 예측
- 신규 극지 균주의 추가 발굴
 - 현재 극지연구소에는 100여종의 극지 미세조류 생물자원이 확보되어 있으며 각 균주들의 최적 생리조건에서 균주들을 배양하는 동시에 이들의 생태, 생리적 특성, 유전정보 등을 탐색 중
 - 남극, 북극의 미탐사 지역을 방문하여 신규 균주를 확보하기 위한 지속적인 탐사
 - AI 기반 기능 예측: *in silico* 화합물 DB, ORF 정보 활용

IV. 연구개발결과

1. 난배양성 극지 균주 대량배양 기술 개발
 - 가. 남극·북극 해양 및 담수로부터 확보한 극지 미세조류들의 종 분류를 통해 대표성을 지닌 15균주를 선별하였다.
 - 나. 선별된 15균주를 2L 소규모 종배양을 수행한 후 10L로 스케일업하여 2g 이상의 건조체를 확보하였다.
 - 다. 15종의 건조중량 확보와 활성효능이 있는 미세조류를 탐색하였고 5종의 미세조류를 선별하였다. 최적 배양조건을 탐색하기 위해 온도와 배지 테스트를 진행함. 배양 온도는 8, 12, 20, 25도로 각각 설정하였고, 배지는 TAP, BG-11, F/2를 사용하여 담수와 해수배지에서 성장성을 평가하였다.
 - 라. 배지 구성성분 최적화, 온도, 광도 등의 환경조건을 개선하고 저온광생물반응기 (low temperature photobioreactor)를 활용하여 단시간내 최대 바이오매스 확보를 위한 배양 프로토콜을 확립하였다.
2. 극지미세조류의 전장 유전체 염기서열 분석
 - 가. 남극미세조류 KSF0031의 유전체 지도 작성 - *de novo* genome sequencing을 위해서, Nanopore sequencing과 Illumina sequencing을 진행하였다.
 - 나. Algae KSF00031의 genomic DNA를 이용해 총 34,976,686,284 bases의 Illumina sequencing결과 대략 83 Mbase로 추정되어지며, Gene annotation을 위해서 RNA를 Illumina sequencing 한 결과 Contig의 구성을 위한 Nanopore sequencing의 경우 4,617,230,585bases였으며, 평균 길이는 4,957base이었다.
 - 다. 남극 미세조류 KSF0031의 어셈블리 - genomic DNA의 Nanopore Sequencing 결과 103개의 contig에 90Mbases의 유전체로 구성되어 있음을 확인하였다. Contig의 N50 크기는 대략 1.9Mbase정도였다. GC content는 63.37%이며, Plastid는 하나의 circular DNA로 119,977 base로 이루어져 있으며, mitochondrial DNA 역시 circular DNA로 65,047 base로 이루어져 있음을 확인하였다.
 - 라. 남극 미세조류 KSF0031 contig의 gene annotation - Chlorophyta *odf10*의 데이터베이스를 대상으로 대략 90.3%의 완성도를 확인할 수 있었다. 총 유전자 수는 12,734로 예측이 되었으며, 유전자를 이루는 평균 길이는 5,604 bases이며, 유전자당 대략 9개의 coding region으로 구성되며, coding sequence의 평균 길이는 1,588 bases임을 확인하였다. BLAST, InterPROScan, KEGG Automatic Annotation Server를 이용하여 gene annotation을 진행하였으며, 총 12,734 유전

자중 8,284 유전자의 기능을 확인할 수 있었다.

마. KSF0031의 Chloroplast genome와 mitochondrial genome의 유전자 지도를 작성하였다.

바. Orthologous group - KSF0031과 유사한 3종의 algae (*Chlorella variabilis* CHNC64, *Micractinium conductrix* LHPF02, *Chlorella sorokiniana* LHPG02)와 유전자 분석을 진행하였다.

사. KEGG 분석을 통한 지질대사 경로 관련유전자의 비교 분석 - KEGG의 organism에 등록되어있는 11개의 green algae 그리고 이전에 orthologous protein의 비교연구에서 사용된 3종의 green algae를 포함해서 총 14개 green algae와 비교하여 지질대사 경로의 유전자의 유무를 확인하였다.

3. 극지 미세조류 이차대사산물의 클러스터분석

가. 남극 미세조류 *Micractinium* sp. KSF00031와 근연종인 *Chlorella variabilis*의 이차대사산물 생합성 유전자 클러스터를 antiSMASH v4.2.0 (antibiotics and secondary metabolite analysis shell)를 이용하여 예측하였다. 실행조건에서 'bacteria' 모드를 선택했을 때 더 많은 클러스터가 예측되는 것을 알 수 있었으며, 일부 클러스터들은 'fungi' 모드의 결과와 예측된 영역이 겹치기도 하였다.

나. 이차대사산물 생합성 유전자 발굴- *C. variabilis*, *C. sorokiniana*, *M. conductrix* 와 KSF00031의 유전체로부터 예측된 protein-coding 유전자는 각각 9,791개, 10,384개, 10,070개, 12,734개였다. 이차대사산물 생합성에 관여하는 유전자를 추가적으로 탐색하기 위하여 InterPro Scan v5.33-72.0를 활용한 단백질 도메인 예측을 수행하고, 이를 기반으로 KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway 및 gene ontology (GO) 분석을 수행하였다

4. AI 기반 기능 예측

가. 선발한 미세조류 1종의 전장 유전체 염기서열 분석으로 ORF 정보를 획득하며, ORF 정보를 활용하여 AI 기반으로 잠재적 기능을 예측함으로써 활용의 확장성 검토

V. 연구개발결과의 활용계획

1. 미세조류 연구가 다각화 되면서 생리활성물질, 의약품 원료물질 같은 고부가 가치를 창출하는 연구가 본격화되고 있다. 특히 극지 미세조류 연구는 새로운 개척분야로써 항암, 항세균, 항진균, 항바이러스, 신경활성과 같은 신규 생리활성물질들의 생산 가능할 것으로 예측되어 진다. 최근 극지 미세조류 유래 천연물질의 주름개선, 항염, 보습 효과를 보고하였고 정상세포에 독성을 나타내지 않는 특성을 보여 세포보호 물질로써 사업화 타당성을 검증하였다.
2. 현 사업과제를 통해 신규 물질들의 활용성이 입증되고, 분자 기전등이 분석된다면 건강식품, 의약품등의 시장 진입이 용이해질 것이다. 극지의 청정 이미지가 합쳐져 현재 전세계 웰빙 트렌드에 맞는 고부가 상품으로 독보적인 분야를 구축가능할 것으로 예상된다.
3. 극지 생물들이 극한의 환경에 적응하기 위해 생산하는 신규 이차대사산물들의 탐색과 기능개선, 분자생물학적 기초, 응용 연구를 통해 새로운 분자기작과 약리 기전을 규명하고 신기능 유용물질 개발의 방향을 제시할 수 있다.

4. 극지 미세조류 유래 유용물질을 활용한 고부가 가치 제품 개발로 미래원천 기술을 확보함으로써 새로운 먹거리창출과, 일자리 증대로 인한 경제활성화, 연구 개발 투자 증대, 신물질 활용 추가 제품개발 등의 시장 선순환 고리를 형성할수 있다.



S U M M A R Y

(영 문 요 약 문)

I. Title

Investigation and mass production of functional materials from polar microalgae

II. Purpose and Necessity of R&D

Competition for securing the rights of biological resources is fierce all over the world, but the trend is already turning to unexplored regions due to the depletion of new resources. It is time for latecomers to focus their efforts on securing the right of useful resources and taking control of the relevant market by discovering new species and exploring their functionality in extreme regions such as the North and South Pole. Natural defenses produced by living organisms in extreme environments are likely to be high-efficiency new materials, emerging as an interesting and alternative to academia and industry. Polar bioresearch, however, has been mainly focused on bacteria, and research on eukaryotic microalgae, which is the most diverse after bacteria, is at an early stage. The discovery and utilization of polar microalgae is in line with the global trend of sustainable resources and can provide countermeasures for new resource development and environmental issues. If it can solve the difficult culture conditions such as low growth rate in low temperature of the polar microalgae and secure enough biomass, it is expected to open up a new industrial field utilizing polar biological resources. Through cooperation with the Korea Institute of Science and Technology, we will enhance the possibility of finding new materials and promote research and industrial use.

III. Contents and Extent of R&D

- Development of mass culture technology for non-cultivated polar strains
 - Developing technology for stable and large quantity production of polar microalgae
 - Operating a low-temperature culture room of about 27m³ inside the KOPRI and build an automated facility for 200L culture equipment.
- Whole genome sequencing

- Obtain protein or ORF sequence information through whole genome sequencing technology
- Further excavation of new polar strains
 - At present, the KOPRI has more than 100 kinds of polar microalgal bioresources and is cultivating strains under optimal physiological conditions of each strain and searching for their ecological, physiological and genetic information.
 - Ongoing exploration to acquire new strains by visiting unexplored regions of Antarctica and Arctic
- AI-based function prediction: *in silico* compound DB, utilizing ORF information

IV. R&D Results

1. Polar microalgae screening

- 15 representative strains were selected through the classification of polar microalgae obtained from the Antarctic and Arctic oceans and freshwater.
- Selected 15 strains were subjected to 2L small-scale cultivation and scaled up to 10L to obtain a dried biomass of 2g or more.
- We searched 15 kinds of microalgae with dry weight and active effect, and selected 5 kinds of microalgae. Temperature and media tests are conducted to find the optimal culture conditions. The culture temperature was set to 8, 12, 20, 25 degrees, respectively, and the medium was evaluated for growth in fresh water and seawater medium using TAP, BG-11, F/2.
- Secured 30g DCW of strain KSF0031 from 200L for physiological activity verification
- Statistical analysis of Plackett-Burman design and Box-Behnken design to confirm culture optimization of KNM0029C
- Improved environmental conditions such as media composition optimization, temperature and light intensity, and established a culture protocol to secure maximum biomass in a short time by using a low temperature photobioreactor

2. Whole genome sequencing

- Genomic Mapping of Antarctic Microalga KSF0031-Nanopore sequencing and Illumina sequencing were performed for de novo genome sequencing.
- As a result of Illumina sequencing, confirmed 34,976,686,284 bases using genomic DNA of KSF00031. It was estimated to be 83 Mbase. As a result

of Illumina sequencing of RNA for Gene annotation, Nanopore sequencing for contig's composition was 4,617,230,585 bases and average length was 4,957bases.

- The assembly of Antarctic microalga KSF0031–Nanopore sequencing of genomic DNA confirmed that it consists of 90Mbases in 103 contig. Contig's N50 size was approximately 1.9 Mbase. GC content was 63.37%, Plastid was composed of 119,977 bases with one circular DNA and 65,047 bases with mitochondrial DNA.
 - Annotation of the gene annotation of the Antarctic microalga KSF0031 contig–Chlorophyta odb10 was found to be approximately 90.3% complete. The total number of genes was predicted to be 12,734, and the average length of genes was 5,604 bases, consisting of approximately 9 coding regions per gene, and the average length of coding sequences was 1,588 bases. Gene annotation was performed using BLAST, InterPROScan, and KEGG Automatic Annotation Server, and the function of 8,284 genes among 12,734 genes was confirmed.
 - Genetic maps of the Chloroplast genome and mitochondrial genome of KSF0031 were prepared.
 - Orthologous group–Three kinds of microalgae similar to KSF0031 (*Chlorella variabilis* CHNC64, *Micractinium conductrix* LHPF02, *Chlorella sorokiniana* LHPG02) were analyzed.
 - Comparative analysis of genes related to lipid metabolism pathways through KEGG analysis–14 green microalgae registered in KEGG organisms used in comparative studies of orthologous proteins. The presence or absence of genes related to lipid metabolism pathway was confirmed.
3. AI-based feature prediction
- Full-length genome sequencing of one microalgae selected to obtain ORF information, examination of utilization by using ORF information to predict potential functions based on AI

V. Application Plans of R&D Results

1. As microalgae research is diversified, researches that create high value such as bioactive substances and pharmaceutical raw materials are in earnest. In particular, polar microalgae research is expected to produce new bioactive substances such as anticancer, antibacterial, antifungal, antiviral and neurological activity as a new pioneering field. Recently, the anti-wrinkle,

- anti-inflammatory, and moisturizing effects of natural materials derived from polar microalgae have been reported.
2. If the current project demonstrates the availability of new substances and analyzes the molecular mechanisms, it will be easier to enter the market for health foods and pharmaceuticals. The clean image of the polar region is expected to combine to create a unique field with high value products that fit the current well-being trends around the world.
 3. By exploring and improving the new secondary metabolites produced by polar organisms to adapt to the extreme environment, molecular biology, and applied research, new molecular mechanisms and pharmacological mechanisms can be identified and new functional useful materials can be developed.
 4. The development of high valuable products using polar microalgae-based useful materials can create a virtuous cycle of new market creation such as future food technologies, revitalizing the economy by increasing jobs and R&D investment.



C O N T E N T S

Chapter 1 Introduction

I. Objective and scope of research

1. Research objective
2. Research scope
3. Research goals and research contents
4. Need for research

Chapter 2 Current R&D Status in Korea and Other Nations

I. Trends in technical development

1. Technical overview
2. Foreign technology and industry trend
3. Domestic technology and industry trends
3. Patent trend

Chapter 3 R&D Implementation Contents and Results

I. Contents and Results

1. Development of mass culture technology for non-cultivated polar microalgae
 - 1.1 Antarctic and Arctic microalgal species selection and mass culture
2. Whole genomic sequencing of polar microalgae
 - 2.1. Genome mapping of Antarctic microalga KSF00031
 - 2.2. DNA and RNA sequencing of Antarctic microalga KSF00031
 - 2.3. Assembly of Antarctic microalga KSF00031
 - 2.4. Prediction of repeat sequence
 - 2.5. Gene annotation of contig of Antarctic microalga KSF00031
 - 2.6. Genetic mapping of the chloroplast genome and mitochondrial genome from KSF0031
 - 2.7. Orthologous group
 - 2.8. Comparative analysis of genes related to lipid metabolism pathways through KEGG analysis
3. Cluster analysis of secondary metabolites from polar microalgae
 - 3.1. Cluster prediction of secondary metabolite biosynthesis gene
 - 3.2. Discovery of additional biosynthetic genes of secondary metabolites
4. AI-based function prediction
 - 4.1. ORF analysis
 - 4.2. Functional prediction

Chapter 4 Degree of R&D Goal Achievement and Degree of Contribution to Outside Research Institute

I. Annual goal achievement

II. research results

Chapter 5 Application Plans of R&D Results

I. Future research direction and achievement utilization plan

Chapter 6 References

목 차

제 1 장 서론

제1절 연구 목적 및 연구범위

1. 연구 목적
2. 연구 범위
3. 연구목표 및 연구내용
4. 연구의 필요성

제 2 장 국내외 기술개발 현황

제1절 국내외 기술동향 분석

1. 기술개요
2. 국외 기술, 산업동향
3. 국내 기술, 산업동향
4. 특허 동향

제 3 장 연구개발수행 내용 및 결과

제1절 연구개발 수행 내용 및 결과

1. 난배양성 극지 균주 대량배양 기술 개발
(가) 남극·북극 미세조류 종 선별 및 대량배양
2. 극지미세조류의 전장 유전체 염기서열 분석
(가) 남극유래 조류 KSF00031의 유전체 지도 작성
(나) 남극유래 조류 KSF00031의 DNA와 RNA 시퀀싱
(다) 남극 미세조류 KSF00031의 어셈블리
(라) Repeat sequence 예측
(마) 남극유래 조류 KSF00031 contig의 gene annotation
(바) KSF00031의 Chloroplast genome과 mitochondrial genome의 유전자 지도 작성
(사) Orthologous group
(아) KEGG 분석을 통한 지질대사 경로 관련유전자의 비교 분석
3. 극지 미세조류 이차대사산물의 클러스터 분석
(가) 이차대사산물 생합성 유전자 클러스터 예측
(나) 이차대사산물 생합성 유전자 추가 발굴
4. AI 기반 기능 예측
(가) ORF 분석
(나) 기능성 예측

제 4 장 연구개발목표 달성도 및 대외기여도

제1절 연차별 목표달성도

제2절 연구결과

제 5 장 연구개발결과의 활용계획

제1절 향후 연구방향 및 성과활용 계획

제 6 장 연구개발과정에서 수집한 해외과학기술정보

제 7 장 참고문헌



제1장 서론

제1절 연구목적 및 연구범위

1. 연구목적

- 난배양성 극지 미세조류 바이오매스를 안정적으로 확보하기 위해 대량배양 최적화프로토콜을 확립하고 유용물질을 활용하기 위해 극지 미세조류 유래 화합물 프로파일 분석 및 기능성 평가 및 전장 유전체 분석하는 것이다.

2. 연구범위

- 유용물질 탐색을 위한 극지 미세조류 후보군 선별
 - 성장속도가 빠르고 20L 배양 가능 종 탐색 (성장포화 1개월 이내 건중량 1g 확보 충족)
- 유용물질 탐색을 위한 극지 미세조류의 대량생산
 - 선별된 극지 미세조류 바이오매스 대량확보 기술 개발
 - 배양 조건 개선(CO₂, 온도, 광도, 배지 등) 및 광생물 배양기 구축
- 유용물질 탐색을 위한 모델 극지 미세조류 1종 선별
 - 전장 유전체 분석으로 ORF 또는 코딩 단백질 정보 확보
- AI기반 기능 예측

3. 연차별 연구목표 및 연구내용

구분	년도	연구목표	연구내용	평가기준
성과목표	2019.02 ~2019.12 (11개월)	극지 미세조류 대량확보	후보 균주들의 배양 최적화 테스트	배양 프로토콜 건수
			10균주 바이오매스 확보	배양 건조체 확보 건수
		선별종의 전장유전체 분석	ORF 또는 코딩 단백질 정보 확보	전장유전체 완료

4. 연구의 필요성

- 전세계가 생물자원의 권리를 확보하기 위한 경쟁이 치열하나 이미 신규 자원 고갈로 지구상 미개척지로 눈을 돌리고 있는 추세이다. 후발주자들은 남북극 같은 극한지에서 새로운 종의 발굴과 기능성 탐색을 통해 유용 자원의 권리 확보 및 관련 시장 장악을 위한 집중적 노력이 필요한 시점이다. 극한 환경에서 생물들이 만들어내는 천연 방어물질들은 고효능 신물질 가능성이 높아 학계와 산업계에 폭발적 관심과 대안으로 떠오르고 있다. 그러나 극지 생물연구는 현재까지 주로 세균을 중심으로 연구되어 왔고 세균 다음으로 다양성이 높은 진핵 미세조류에 대한 연구는 초기 단계이다. 극지미세조류 발굴과 활용은 세계적 추세인 지속가능한 자원(sustainable resource)에 부합하며 신규자원개발과 환경문제에 대한 대응책을 제시할 수 있다. 극지 미세조류의 낮은 성장속도와 저온성장 등 까다로운 배양 조건을

해결하고 충분한 바이오매스를 확보할 수 있다면 극지생물자원 활용 신산업 분야를 개척할 것으로 확신한다. 한국과학기술연구원과 협력 연구를 통해 신규 물질 발견 가능성을 높임으로써 연구 활성화와 산업적 접근이 가능하다.

가. 기존연구의 한계

(1) 신규 물질 확보의 어려움 (한국과학기술연구원)

(가) 우리 주변의 생명체로부터의 물질 연구는 포화 상태에 이르러 신규 물질을 발견할 가능성이 낮은 상황이다.

(나) 신규 물질 확보를 위하여 극한 환경에서 서식하는 생명체를 확보하고자 하나 접근성이 용이하지 않아 생물자원 확보에 어려움을 겪고 있다.

(2) 극지 미세조류 연구의 어려움 (극지연구소)

(가) 어려운 접근성과 관련 전문가 부족으로 연구가 미흡하여 전 세계적으로도 극지 미세조류 시료 확보와 과학적 정보가 부족하다.

(나) 극한지 특성상 배양이 어려워 산업적 활용을 위한 충분한 바이오매스 확보가 어렵다.

나. 협력연구의 필요성

(1) 극지연구소가 확보한 극지 미세조류에 대한 연구 활성화와 산업적 접근이 필요한 시점이며 발굴, 배양부터 활성물질 분리, 활성평가까지 체계적인 다학제 협력이 가능해진다.

(2) 한국과학기술연구원은 극지연구소에서 확보하고 있는 극지 환경의 생물자원을 분양 받아 연구함으로써 신규 물질 발견 가능성을 높일 수 있다.

다. 기술적 측면

○ 남극과 북극 탐사를 통해 지난 10년 이상 미세조류 균주 100여종을 확보하였고, 2도씨 저온 배양실에서 배양 중에 있으며 현재 이를 활용한 연구는 초기 단계이다. 극지 미세조류를 산업화 소재로 활용하기 위해서는 충분한 양의 바이오매스 확보가 가장 난제이다. 극지 미세조류의 바이오매스를 확보하기 위해 미세조류 후보 균주에 대한 생리 특성 연구데이터가 필요하고, 이를 기반으로 대량배양시스템에 적용해야 한다. 배지조성, 온도, 광도, 이산화탄소 농도 등 배양환경을 조절하여 각 균주별 최적 성장조건을 확립하여 스크리닝을 위해 최소 20L~200L를 확보해야 한다.

라. 경제·산업적 측면

○ 미세조류를 이용한 제품은 건강식품이 주류를 이루며, 클로렐라와 스피룰리나가 대표적인 미세조류 상품으로 다양한 종수에 비해 산업화된 미세조류는 극히 일부 종들이다. 미세조류 추출물 베타카로틴, 아스타잔틴 등이 고부가가치 천연색소로 식품 첨가제나 착색제로 이용되고 있다. 또한 DHA와 EPA등의 필수불포화지방산 등이 영양보충제로 각광을 받고 있다. 미세조류가 생성하는 생리활성물질과 의약품 원료물질에 대한 연구가 진행되고 있으며 이를 활용해 항암, 항염, 항균, 항바이러스 활성을 지닌 물질들이 생산 가능한 것으로 보고되었다. 극지 미세조류 유래 추출물의 항염, 항암, 항균 활성에 대한 연구는 시작단계에 있으며 이에 대한 연구 성과를 국제저널에 보고하는 수준이다. 극지 생물들이 가지는 특이 2차대사산물들의 개발과 분자, 생화학적 기초 연구를 통해 새로운 분자기작과 약리 기전을 규명함으로써 새로운 유용물질 개발에 방향을 제시할 수 있다.

마. 과학적 측면

- 극지 미세조류 유래 추출물을 생산하기 위해 바이오매스를 확보하는 연구는 미세조류를 증대된 규모로 배양하고 수확하는 Low technology 측면에서 중요한 기반을 제공한다. 극지 미세조류에 대한 대량생산연구는 미미한 실정으로 배양기술 개발로 충분한 바이오매스를 확보, 활용할 수 있다면 향후 다양한 저온 균주에 적용 가능한 배양 프로토콜을 제공하고 지속적인 생물자원을 제공에 기여할 수 있다. 극지 미세조류 유래 추출물 및 신기능 천연유용물질 등을 확인함으로써 미확인 신규화합물을 발굴하여 극지 미세조류 유래 유용물질의 가치를 높이고 기초 연구를 넘어 활용 연구로 성과를 이끌어 낼 수 있다.

바. 사회·문화적 측면

- 협력연구를 통해 극지 미세조류에 대한 연구인력과 자원을 효율적으로 활용할 수 있으며 극지연구소가 확보한 극지 미세조류 연구 인프라와 사업영역을 확장시킬 것으로 기대한다. 극지미세조류 발굴과 활용은 세계적 추세인 지속가능한 자원(sustainable resource)에 부합하며 신규자원개발과 환경문제에 대한 대응책을 제시할 수 있다. 극지역에 대한 대국민 이미지를 극대화시키고 극지 생명체 응용연구에 대한 도전정신, 기업가 정신 확산으로 국가 경쟁력을 증대시킬 것이다.



제2장 국내외 기술개발 현황

제1절 기술동향 분석

1. 기술개요

가. 미세조류

- 미세조류(microalgae, 微細藻類)란 바다나 강에 서식하는 현미경적 크기의 단세포 광합성 생물로써, 자연오일, 단백질, 탄수화물로 구성되어 있으며, 지구 전역에 광범위하게 분포하고 있는 생물이다. 약 35억년전 지구상에 출현하여 이산화탄소를 흡수하고 영양염류(N,P)와 태양광을 받아 산소와 물을 부산물로 방출했던 최초의 생물체이다(Chlorella, Schizochytrium, Aphanizomenon, Nostoc 등).
- 전 세계적으로는 약 25,000여종, 우리나라에는 약 1,300여종이 존재하며 세포가 작으면서도 하루에서 2~3회 증식해 좁은 공간에서도 무제한 생산이 가능한 경제적 잠재력이 매우 높은 생물자원이다.
- 기초과학 및 기후변화 연구재료, 대체에너지, 먹이생물, 식품 및 의약품의 원료, 건강보조식품 등 다양한 분야에서 활용되고 있다.
- 현재 미세조류는 다양한 기술 분야에 사용 중이며 농업 분야에서는 토양개량제 및 생물비료로, 대체 에너지 분야에서는 바이오에탄올, 디젤 등을 생산하기 위한 수단으로 사료 및 첨가제 분야에서는 동물, 어분 사료 또는 사료 첨가제 등으로, 식품/의약품 분야에는 일반식품, 건강보조식품, 의약품 등에 색소, 지방산 등을 이용하고 있으며 기타 폐수처리, 환경정화, 건축재료, 방사능 완화 등 다양한 분야에서 사용되고 있다.

2. 국외 기술, 산업동향

- 해외의 경우, 지리적 제약과 희귀성으로 극지 미세조류의 학문적, 산업적 접근은 아직 초기 단계이다. 미국은 막대한 자원을 투자해 다수의 미세조류를 포함한 생물 게놈 프로젝트(Genomes to Life, US DOE, 2001-2020)를 수행하고 있으며, 2003년 ‘게놈시대의 극지생물학 프론티어’ 보고서를 통해 극지 유전체 연구에 대한 적극적인 지원과 관심을 보이고 있다. 북극권 국가들인 핀란드, 노르웨이 기업들이 북극 식물, 해조류를 이용한 향장, daily care 상품들을 시판하고 있다 (<http://www.skyniceland.com>; <http://www.lumenelab.com>; <http://www.oilgae.com> 등). 말레이시아 Algatech는 미세조류 Haematococcus 유래 아스타잔틴을 하와이, 인도, 이스라엘 지역에서 대량생산하고 있으며, 호주 BASF는 Dunaliella 유래 베타카로틴, 미국 Martek는 Cryptocodinium과 Shizochytrium로부터 DHA, Cyanotech사는 Spirulina를 생산중에 있다. 미국 MARY KAY과 독일 Heinrich U.는 해조류 유래 항노화 화장품 개발 및 기술 특허를 가지고 있다. 프랑스 Polaar과 Biotherm, 일본 Shiseido는 미세조류 유래 고보습, 항노화 화장품에 대한 특허 및 제품을 출시하였다. 미세조류 유래 색소물질은 아스타잔틴은 활성산소 저감율이 베타카로틴과 비타민E 보다 각각 10배 및 100배 이상 높은 것으로 보고하였다. 미국 Biotech, Cyanotech, Omegatech같은 기업들은 대학, 정부연구기관과 협력체계를 구축하여 해양미세조류 유래 기능성물질을 개발하고 사업화하여 매출을 올리고 있다. 국제 클라미도모나스 Conference (<http://www.chlamy2010.org/>)가 격

년제로 개최되고 있으며, DNA서열, RNA서열, 단백질 서열 등 유전체 정보 Database가 구축되어 (Chlamy DB (<http://www.chlamy.org/chlamydb.html>), Kazusa (<http://est.kazusa.or.jp/en/plant/chlamy/EST/index.html>), NCBI) 연구자들에게 공개되었다. 최근 들어 전 세계적으로 미세조류 및 해조류의 물질 추출, 배양, 활용, 반응기 기술 분야는 폭발적인 성장세를 나타내고 있으나 비극지 미세조류들이 주대상이었다.

3. 국내 기술, 산업동향

- 국내 또한 극지미세조류 연구와 활용은 제한적인데 현재까지 발표된 관련 성과들은 극지연구소와의 협업 또는 균주 공여 등으로 이루어진 예가 대부분이다. 경북대학교 윤호성 교수팀(차세대바이오매스 연구단)은 토착미세조류 균주들을 이용해 바이오연료 feedstock 생산 및 유용물질 산업화 사업을 수행 중이며 균주 중 일부는 극지연구소와 협력연구 수행 중 채집한 극지미세조류들을 보유하고 있다. 한국생명공학연구원 정원중 박사팀 (식물시스템공학연구센터)은 극지연구소, 충남대와 함께 고농도 지질을 함유한 북극유래 클로렐라 ArM29B를 발굴하여 바이오에너지 연구개발 전문기업인 에이스하이텍과 기술이전 계약을 체결하였고, 한양대 생명과학과 진언선 교수(식물생명공학연구실)는 남극 미세조류에서 결빙방지 단백질을 발견해 보고하였으며, 현재 극지 미세조류에서 발굴한 저온활성 프로모터 형질전환체를 제작하는 연구를 극지연구소 위탁과제로 수행 중이다. 극지연구소에 수행 중인 연구 결과로 주름개선, 항염, 보습 효과와 항암활성을 보이는 균주들을 발굴하여 국제저널에 보고하였다. 비극지 미세조류 활용 예로는 해양바이오소재연구단은 2016년 스피롤리나를 이용하여 활성산소를 낮추고 상처 세포를 활성화하여 피부재생을 높이는 나노소재를 개발, 낙동강 생명자원관은 미세조류 시조카이트리움 속 PB-31 균주 추출물의 항염, 항산화 등 효과를 확인하여 파이크일바이오텍코리아(주)과의 기술이전 협약을 체결하였다.



4. 특허 동향

- 전세계적 웰빙 트렌드로 건강증진과 노화방지 제품에 대한 수요가 증대되고 있으며 이에 따라 천연원료를 이용한 기능성제품들이 개발되고 있음. 특히 미세조류유래 천연 원료를 이용한 산업이 새롭게 부각되고 있다. 미세조류 유래 천연물질 및 화합물의 항염, 항암, 주름개선, 미백, 항노화 연구가 활발하게 진행되어 기술특허 등이 다수 등록되었으나 제품화로 이어진 경우는 미비한 실정이다. 향후 신규원료 물질에 대한 요구가 증대될 경우 국내외 대기업 기술제휴 등을 통해 관련 시장 경쟁이 치열해질 것으로 예상된다. 극지 미세조류의 경우 서식지의 특수성으로 바이오매스의 확보가 어려웠으나 극지 연구소의 보유종과 배양기술을 활용하면 다수의 신기능 물질에 대한 발견이 기대된다. 최근 특허 동향으로는 고려대에서 미세조류 유래 당을 통해 베타락탐계 항생물질인 CPC(세팔로스포린C)를 생산하는 연구를 수행중이며, 한국해양과학기술원에서 미세조류 *Chlamydomonas hedleyi*로부터 UV보호물질 생산기술 특허 등록하였으며, 한국생명공학연구원에서는 미세조류 *Arthrospira* sp.로부터 β -carotene, Phycocyanin, GLA 등의 생산기술 특허 등록하였다. 극지연구소는 항염, 항암활성을 지니는 극지미세조류 유래 추출물에 대한 균주 특허를 등록하였다.

제3장 연구수행 내용 및 결과

제1절 연구개발 수행 내용 및 결과

1. 난배양성 극지 균주 대량배양 기술 개발

(가) 남극·북극 미세조류의 종 선별 및 대량배양

- 남극·북극 해양 및 담수로부터 확보한 극지 미세조류들의 종 분류를 통해 대표성을 지닌 15균주를 선별하였다.

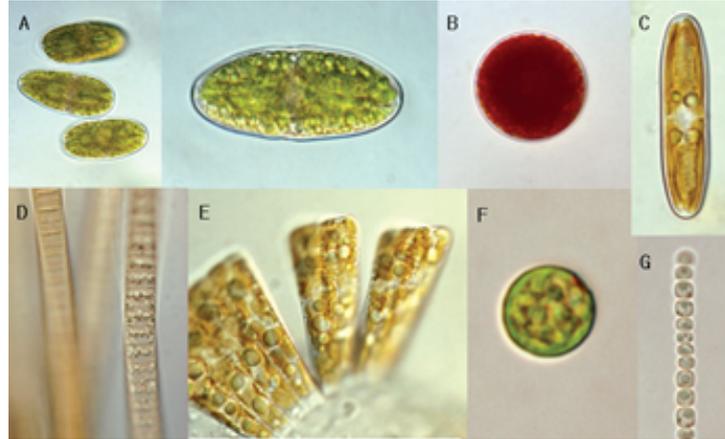


그림 1. 남극·북극 해양 및 담수에서 발견된 미세조류

- 보유중인 극지 미세조류 중 성장속도가 우수한 15종을 선별하여 2.5L 소규모 종배양을 수행한 후 10L로 스케일업하여 2g 이상의 건조체를 확보하였다(표 1).

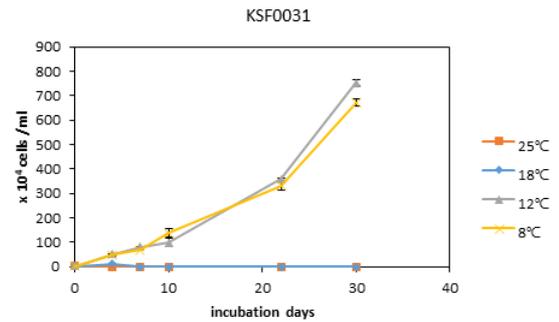
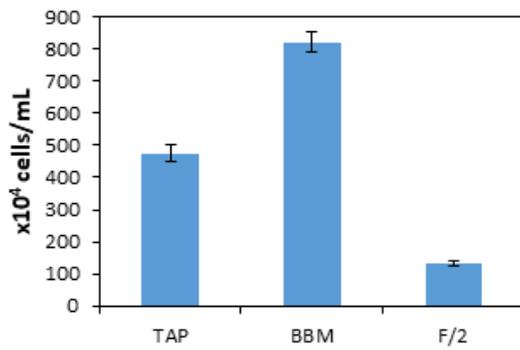
표 1. 광배양기를 활용한 극지 미세조류의 바이오매스 확보

		총수확량 (g)	리터당 생산량
		배양액 12L	g/L
1	KSF0134	4.1478	0.35
2	KSF0123	4.1602	0.35
3	KNF0039	1.6791	0.14
4	KSF0094	2.7997	0.23
5	KSF0082	3.13	0.26
6	KSF0089	3.5182	0.29
7	KSF0208	4.2738	0.36
8	KSF0172	2.2477	0.19
9	KSF0203	5.2544	0.44
10	KSF0202	3.3687	0.28
11	KSF0110	2.3268	0.19
12	KSF0086	4.845	0.40
13	KSF0092	5.6804	0.47
14	KSF0001	3.9915	0.33
15	KSF0031	5.7806	0.48

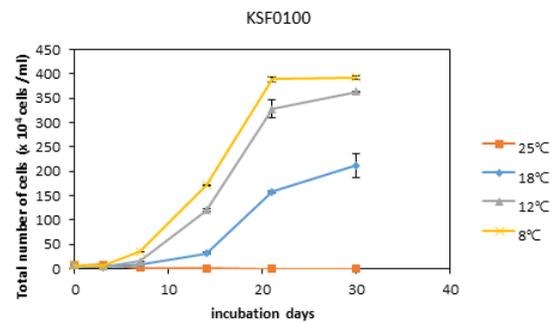
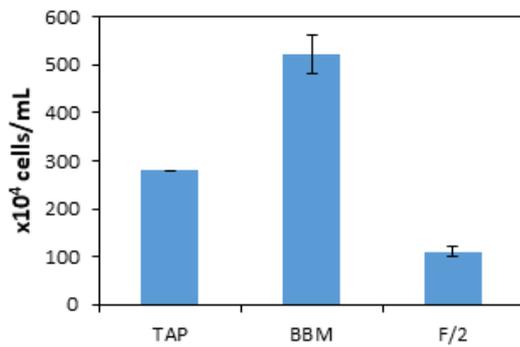
- 15종의 건조중량 확보와 활성효능이 있는 미세조류를 탐색하였고 5종의 미세조류를 선

별하였다. 최적 배양조건을 탐색하기 위해 온도와 배지 테스트를 진행하였다. 배양 온도는 8, 12, 20, 25도로 각각 설정하였고, 배지는 TAP, BG-11, F/2를 사용하여 담수와 해수배지에서 성장성을 평가하였다.

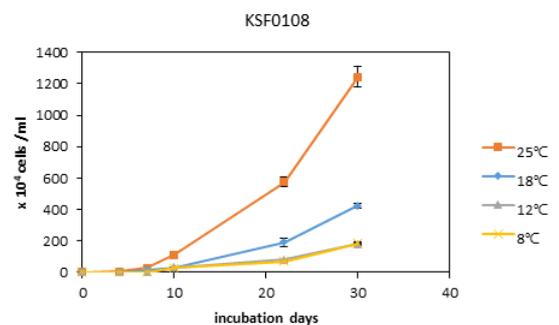
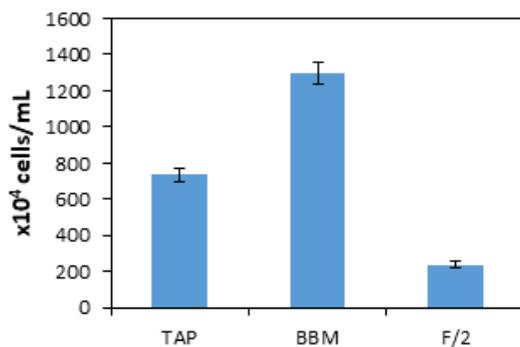
KSF0031



KSF0100



KSF0108



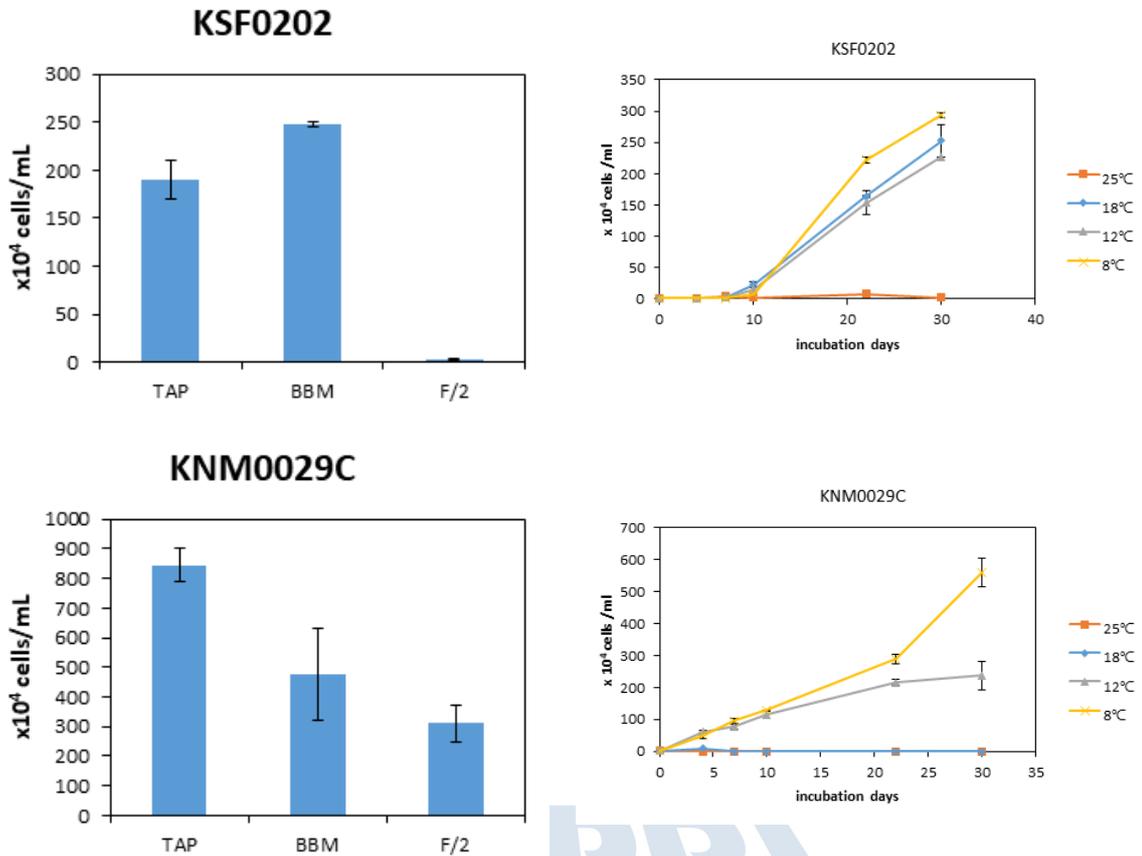


그림 2. 극지 미세조류 온도 및 배지 테스트

- 극지 미세조류 건조체 및 천연추출물을 확보하고 생리활성검증을 위해 균주 KSF0031를 총 30g/200L DW(건조체) 이상 확보하였다.
- KNM0029C의 대량 배양조건을 확립하기 위해 Plackett-Burman design 통계 분석을 하였다(표 2). Box-Behnken design 실험 후 2차 배지를 설계하였다(표 3). Positive effect factors: (+)는 BG-11 배지 기준농도의 2배, (-)는 (+)의 1/10 Negative effect factors: (+)는 BG-11 배지 기준농도, (-)는 (+)의 1/10 으로 한다(표 2). 다른 factor (NaNO₃, CaCl₂ · 2H₂O, citric acid, EDTA · Na₂, Na₂CO₃, Trace)는 원래 농도로 투입 (표 3). 설계된 배지에 KNM0029C를 접종하고 2주후 세포수를 측정하였다. 이를 기반으로 KNM0029C배양을 위한 배지최적화를 완성하였다(표 4).

표 2. Plackett-Burman design 통계 분석

Variable	Medium component	+ value (g/L)	- value (g/L)	Effect	t-statistic	P-value
X ₁	Tris base	0	2.42	564.6	17.19	0.000
X ₂	NH ₄ Cl	0	0.375	296.7	9.03	0.000
X ₃	MgSO ₄ · 7H ₂ O	0	0.1	211.3	6.43	0.000
X ₄	CaCl ₂ · 2H ₂ O	0	0.05	-13.7	-0.42	0.683
X ₅	Potassium phosphate	0	0.432	111.7	3.40	0.004
X ₆	Trace elements	0	0.0977	75.0	2.28	0.036
X ₇	Acetic acid	0	1.0	570.3	17.36	0.000

표 3. Box-Behnken design 통계분석

Variable	Medium component	- value (g/L)	0 value (g/L)	+ value (g/L)
X ₁	NH ₄ Cl	0.075	0.4125	0.75
X ₂	MgSO ₄ · 7H ₂ O	0.02	0.11	0.2
X ₃	Potassium phosphate	0.0432	0.2376	0.432
X ₄	Trace elements	0.00977	0.053735	0.0977

표 4. 배지 최적화

Component	Original TAP (g/L)	Optimized TAP (g/L)
Tris base	2.42	2.42
NH ₄ Cl	0.375	0.545
MgSO ₄ ·7H ₂ O	0.1	0.155
CaCl ₂	0.05	0.05
K ₂ HPO ₄	0.288	0.029
KH ₂ PO ₄	0.144	0.014
Trace	0.097	0.077
AcOH	1.0 (mL)	1.0 (mL)

- 배지 구성성분 최적화, 온도, 광도 등의 환경조건을 개선하고 저온광생물반응기(low temperature photobioreactor)를 활용하여 단시간내 최대 바이오매스 확보를 위한 배양 프로토콜을 확립하였다(그림 3).



그림 3. 극지연구소에서 보유중인 후보균주와 저온광생물반응기 활용모습

2. 극지미세조류의 전장 유전체 염기서열 분석

(가) 남극유래 조류 KSF00031의 유전체 지도 작성

- Antarctic algae KSF00031의 *de novo* genome sequencing을 위해서, Nanopore sequencing과 Illumina sequencing을 진행하였다. Nanopore sequencing결과의 error correction에는 canu를 이용하였으며, long read의 assembly에는 smartDenovo를 이용하여 contig를 형성하였다. 정확도가 높은 short reads를 이용하여 contig의 base quality를 또 한번 보정하였으며, Pilon이라는 프로그램을 이용하였다 (그림 4).

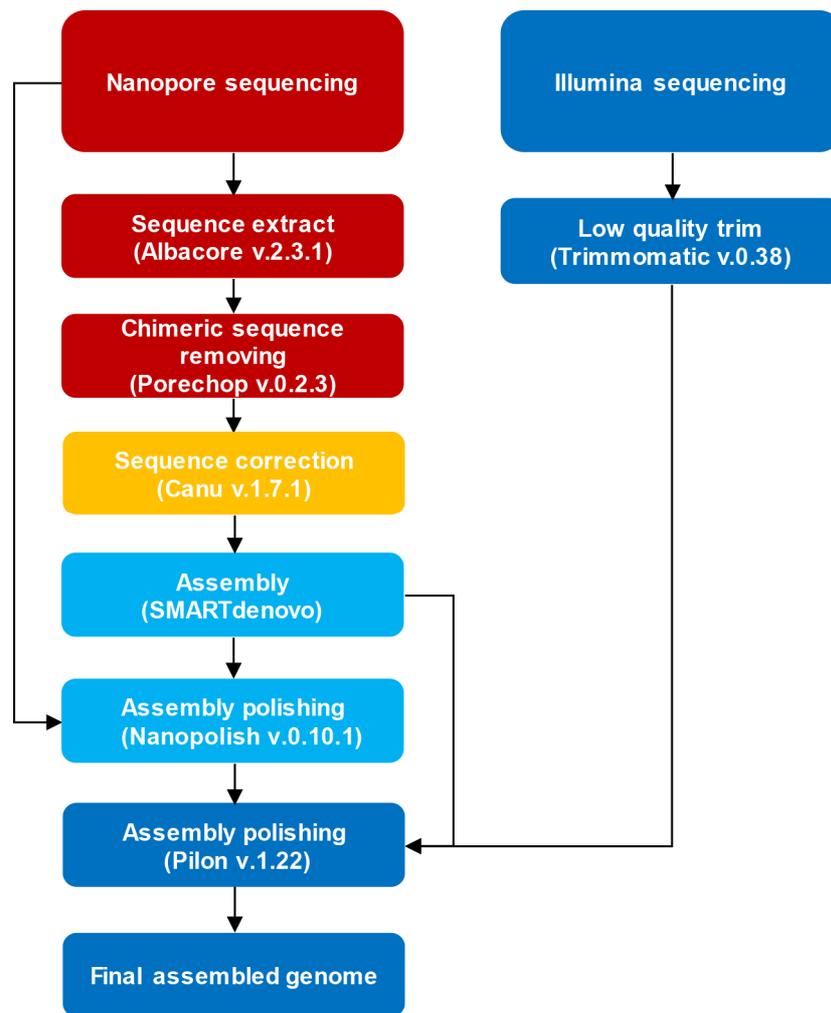


그림 4. 데이터분석 프로세스. Albacore (ver. 2.3.1)와 Canu (ver. 1.7.1)를 사용하여 nanopore sequencing reads를 수행

(나) 남극유래 조류 KSF00031의 DNA와 RNA 시퀀싱

- Algae KSF00031의 genomic DNA를 이용해 총 34,976,686,284 bases의 Illumina sequencing결과를 얻었으며 (표 5), GenomeScope라는 프로그램을 통해 유전체 크기를 예측하였다. 대략 83 Mbase로 추정되어지며 (그림 5 and 표 6), Gene annotation을 위해서 RNA를 이용하여 Illumina sequencing을 추가적으로 진행하였다 (표 7). Contig의 구성을 위한 Nanopore sequencing의 경우 4,617,230,585bases의 결과를 얻었으며, 평균 길이는 4,957base이었다 (표 8).

표 5. Genomic DNA 부터 얻은 Illumina read statistics

No	Sample	Total read bases (bp)	Read	GC(%)	AT(%)	Q20(%)	Q30(%)
1	KSF0031 -DNA	34,976,686,284	231,633,684	62.933	37.07	92.233	84.133

표 6. GenomeScope 를 이용한 Genome 크기 예측

	min	max
Heterozygosity	0.137441%	0.148301%
Genome Haploid Length	82,845,474 bp	82,959,370 bp
Genome Repeat Length	8,023,638 bp	8,034,669 bp
Genome Unique Length	74,821,836 bp	74,924,701 bp
Model Fit	94.8236%	95.8512%
Read Error Rate	0.516846%	0.516846%

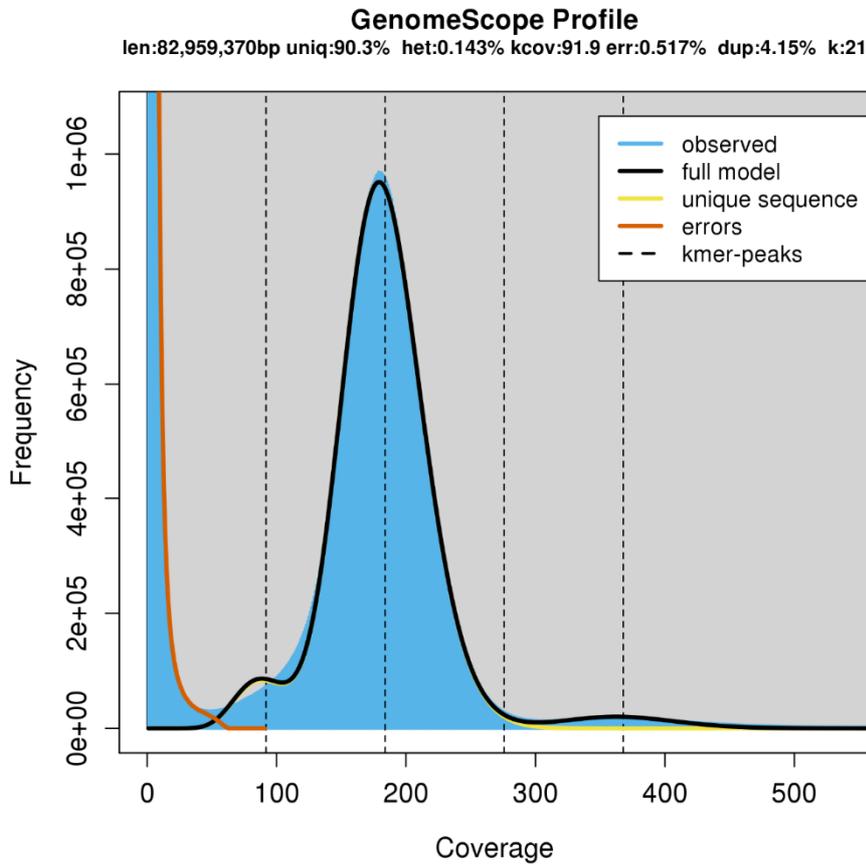


그림 5. GenomeScope 를 이용한 Genome 크기 예측

표 7. Total RNA 에서 얻은 Illumina read statistics

No	Sample	Total read bases (bp)	Read	GC(%)	AT(%)	Q20(%)	Q30(%)
1	11_KSF0031-0a	6,675,054,660	44,205,660	65.09	34.91	97.45	93.64
2	12_KSF0031-0b	5,732,672,116	37,964,716	65.14	34.86	97.54	93.84
3	13_KSF0031-0c	5,448,532,698	36,082,998	65.05	34.95	97.59	93.9
4	14_KSF0031-4a	6,321,442,256	41,863,856	64.96	35.04	97.74	94.31
5	15_KSF0031-4b	6,259,120,932	41,451,132	64.92	35.08	97.58	93.84
6	16_KSF0031-4c	8,460,861,596	56,032,196	64.92	35.08	97.47	93.61
7	17_KSF0031-12a	5,697,595,118	37,732,418	64.87	35.13	97.33	93.31
8	18_KSF0031-12b	5,687,860,752	37,667,952	64.97	35.03	97.57	93.8
9	19_KSF0031-12c	5,332,025,628	35,311,428	64.53	35.47	97.91	94.51

표 8. Nanopore read statistics

	Raw data	After trimming	Corrected read
Total read number	928,069	925,493	202,068
Total read bases (bp)	4,617,230,585	4,581,728,125	2,509,314,860
Mean read length (bp)	4,957	4,951	12,418
Max length (bp)	89,456	89,408	72,353
Read length N50 (bp)	9,910	9,924	14,366
Number above 5 kbp / total length (bp) / percentage of the total reads (%)	294,471/3,431,469,233/74	292,555 / 3,409,495,441/74	51/100
Number above 10 kbp / total length (bp) / percentage of the total reads (%)	134,773 / 2,289,939,383 / 50	133,930 / 2,275,614,496 / 50	76/96
Number above 20 kbp / total length (bp) / percentage of the total reads (%)	32,627 / 882,508,886 / 19	32,424 / 876,967,067 / 19	26,560/711,661,497 /28

kbp = kilo base pairs. The raw data were base-called using Guppy software, and Canu was used to correct the longest reads up to 40× coverage as default.

(다) 남극 미세조류 KSF00031의 어셈블리

- Algae KSF00031의 genomic DNA의 Nanopore Sequencing 결과를 canu를 이용하여 일차적으로 error correction을 진행하였다 (표 9). 그 결과를 이용하여, smartDenovo라는 프로그램을 사용하여, 어셈블리를 진행하였다 (표 10). 어셈블리 결과 103개의 contig에 90Mbases의 유전체로 구성되어 있음을 확인하였으며, Contig의 N50 크기는 대략 1.9Mbase정도였다. Pilon이라는 프로그램을 이용하여, 낮은 base quality를 가지는 long reads로 구성된 contig의 bases quality를 향상시켰다 (표 10). GC content는 63.37%이며, Plastid는 하나의 circular DNA로 119,977 base로 이루어져 있으며, mitochondrial DNA 역시 circular DNA로 65,047 base로 이루어져 있음을 확인하였다 (표 10). 유전체의 완성도를 가늠하기 위해서 BUSCO를 이용하였으며 (표 11), Chlorophyta *odb10*의 데이터베이스를 이용한 경우 초기의 assembly결과에서 genome polishing을 진행함에 따라서 Genome completeness가 증가하는 것을 확인할 수 있었다 (표 11). 4번이상에서는 더 이상 증가 되지 않는 것을 확인하였으며 (표 11), 총 5번의 genome polishing을 진행한 contig를 가지고 gene annotation을 진행하였다.

㉟ 9. Genome assembly statistics

	canuSMART	polishing
Number of scaffolds	103	103
Number of contigs	103	103
Total scaffold sequence (bp)	90344128	91247668
Total contig sequence (bp)	90344128	91247668
Length of N50 scaffold (bp)	1946981	1967715
Length of N50 contig (bp)	1946981	1967715
Max scaffold length (bp)	6225282	6284442
Number of scaffolds > 50 KB	91	91
% main genome in scaffolds > 50 KB	99.61%	99.61%

The nanopore reads were assembled with nanopore reads corrected by Canu using SMARTdenovo. Four polishing were performed using pilon program. Utg175 is chloroplast genome sequence and Utg203 is mitochondrial DNA sequence.

㉟ 10. Genome assembly statistics

	Number of contigs	Total contig sequence (bp)
Genome	101	91062644
Plastid	1	119977
Mitochondrial DNA	1	65047
Total contig sequence (bp)	103	91247668
GC content (%)	63.37	

The nanopore reads were assembled with nanopore reads corrected by Canu using SMARTdenovo. Four polishing were performed using pilon program. Utg175 is chloroplast genome sequence and Utg203 is mitochondrial DNA sequence.

㉟ 11. BUSCO completeness assessments for genomes

Database	Assemblies and genome polishing	Complete BUSCOs	Duplicated BUSCOs	Fragmented BUSCOs	Missing BUSCOs	Total BUSCO groups searched orthologs
Eukaryota <i>odb9</i>	canuSMART	20.8%	0.3%	14.2%	65.0%	303
	canuSMART +pl	81.6%	0.7%	3.6%	14.8%	303
	canuSMART +pl×2	81.6%	0.7%	4.0%	14.4%	303
Embryophyta <i>odb9</i>	canuSMART	1.0%	0.0%	0.8%	98.2%	1440
	canuSMART +pl	22.8%	0.3%	1.7%	75.5%	1440
	canuSMART +pl×2	23.8%	0.3%	1.6%	74.6%	1440
Chlorophyta <i>odb10</i>	canuSMART	46.5%	0.4%	8.5%	45.0%	2,168
	canuSMART +pl	93.3%	1.5%	4.5%	2.2%	2,168
	canuSMART +pl×2	94.3%	1.6%	3.7%	2.0%	2,168
	canuSMART +pl×3	94.5%	1.7%	3.6%	1.9%	2,168
	canuSMART +pl×4	94.5%	1.7%	3.6%	1.9%	2,168

canuSMART +pl×5 94.5% 1.7% 3.6% 1.9% 2,168

The bold characters indicate the best statistics of genome completeness assessment using BUSCO.

(라) Repeat sequence 예측

○ Repeat modeler를 이용하여 *de novo* repeat sequence를 예측하였으며, 결과는 표 12, 13과 같다.

표 12. Major repetitive content and tRNAs

	de novo repeat library		Repeat library	
	bp	%	bp	%
Interspersed repeats	8,551,940	9.37	56,497	0.06
Simple repeats	5,642,726	6.18	5,927,220	6.5
Low complexity tRNA	659,569	0.72	695,624	0.76

The total lengths of the repeats and tRNAs were calculated using RepeatMasker and tRNAscan-SE, respectively, and the number of elements is given in parentheses.

표 13. Statistics of interspersed repeats contents

	<i>de novo</i> repeat library	Repeat library
SINE	0 (0)	4,749 (62)
LINE	3,086,141 (12,335)	40,690 (466)
LTR	230,136 (548)	273 (4)
DNA	325,767 (1,037)	10,286 (133)
Unclassified	4,909,896 (20,443)	499 (4)
Total interspersed repeats	8,551,940	56,497

The total lengths of repeats and tRNAs were calculated using RepeatMasker, and the number of elements is given in parentheses. Long terminal repeats (LTRs) are retrotransposons, and non-LTR retrotransposons comprise long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs).

(마) 남극유래 조류 KSF00031 contig의 gene annotation

○ Algae KSF00031의 annotation을 위해 우선 RNA의 sequencing 결과를 이용하여, Transcriptome assembly를 진행하였다. Trinity를 이용하여 진행하였으며, 결과는 표 14 그리고 표 15과 같다. 이 Transcriptome 정보와 함께, NCBI의 조류의 단백질 정보를 사용하고, MAKER2 프로그램을 이용하여 gene을 prediction하였다. Augustus를 *ab initio* gene prediction program을 선택하여 진행하였으며, BUSCO 프로그램을 사용하여, gene set completeness를 확인하였다 (표 15). Chlorophyta *odb10*의 데이터베이스를 대상으로 대략 90.3%의 완성도를 확인할 수 있었다 (표 16). 총 유전자 수는 12,734로 예측이 되었으며, 유전자를 이루는 평균 길이는 5,604 bases이며, 유전자당 대략 9개의 coding region으로 구성되며, coding sequence의 평균 길이는 1,588 bases임을 확인하였다 (표

17). BLAST, InterPROScan, KEGG Automatic Annotation Server를 이용하여 gene annotation을 진행하였으며 (그림 6 and 표 18), 총 12,734 유전자중 8,284 유전자의 기능을 확인할 수 있었다.

표 14. Transcriptome assembly statistics using trinity

	stats
Number of contigs	230,539
Total contig sequence (bp)	234,732,365
Length of N50 contig (bp)	1729
Max contig length (bp)	35,880

표 15. Contig size distribution of transcriptome assembly

Minimum contig length	Number of contig	Total contig length	Scaffold contig coverage
All	230539	234732365	100.00%
100	230539	234732365	100.00%
250	210179	230153012	100.00%
500	127647	200651182	100.00%
1000	72463	162000033	100.00%
2500	20536	79765441	100.00%
5000	3156	21227378	100.00%
10000	195	2826199	100.00%
25000	22	784706	100.00%

표 16. BUSCO gene set completeness assessments for genomes

Database	Assemblies and genome polishing	Complete BUSCOs	Duplicated BUSCOs	Fragmented BUSCOs	Missing BUSCOs	Total BUSCO groups searched orthologs
Eukaryota <i>odb9</i>	canuSMART +pl×4	92.4%	2.3%	3.0%	4.6%	303
Embryophyta <i>odb9</i>	canuSMART +pl×4	28.2%	1.0%	1.1%	70.7%	1440
Chlorophyta <i>odb10</i>	canuSMART +pl×4	90.3%	1.5%	3.6%	6.1%	2,168

표 17. Summary of MAKER2 annotation

KSF00031		
gene	number ¹	12,734
	length ²	71,366,613 (5,604.4)
CDS	number	120,260 (9.4)
	length	20,320,935 (169.0)
exon	number	127,134 (10.0)
	length	26,689,558 (210.0)
intron	number	114,400 (9.0)
	length	44,677,055 (390.5)

5'-UTR	number	10,896 (0.86)
	length	1,954,051 (179.3)
3'-UTR	number	10,573 (0.83)
	length	4,414,572 (417.5)

CDS = coding sequence; UTR = untranslated region. The numbers and total lengths of the genes, CDSs, exons, introns, and UTRs were calculated from a GFF3 file generated by MAKER2, and the unit averages are given in parentheses.

¹ denotes the number of elements

² denotes the total length of the elements

333,002 protein sequences from NCBI reference sequence for algae and 230,539 contigs assembled using Trinity were used as evidence sequences for MAKER2 annotation pipeline.

Top hit genus

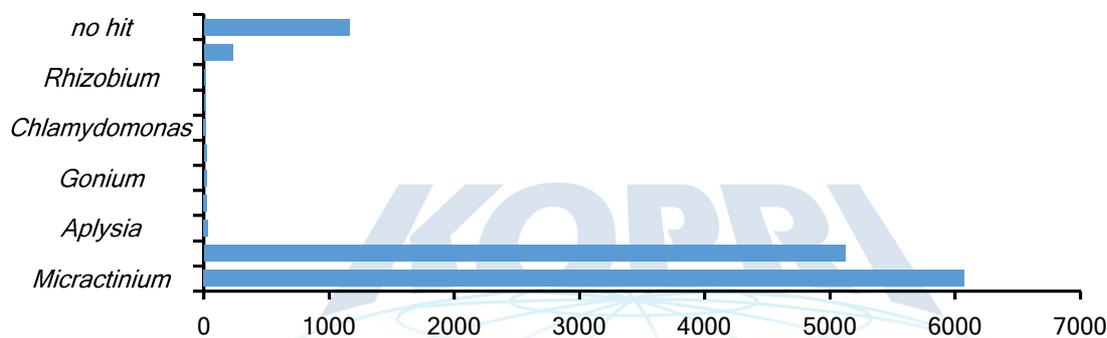


그림 6. Basic local alignment search tool (BLAST) top-hit species distribution with total protein.

표 18. Summary of gene annotation

KSF00031		
gene	number ¹	12,734
	length ²	71,366,613 (5,604.4)
	CDSs assigned to nr database using blastP	11,570
	CDSs assigned to GO	6,251
	CDSs assigned to KEGG	4,144
	Hypothetical protein	4,450

(바) KSF0031의 Chloroplast genome와 mitochondrial genome의 유전자 지도 작성

○ KSF0031의 chloroplast genome의 경우 OGDRAW를 이용하여 annotation 하고 길이는 119,977 bp의 circular map을 작성하였다 (그림 7a). Mitochondrial genome의 경우, MITOS webserver를 통해 얻은 tbl 정보를 tbl2asn (Sequin)을 이용하여 genbank 파일로 변환한 다음 OGDRAW를 이용하여 유전체 그림을 확보하였다 (그림 7b).

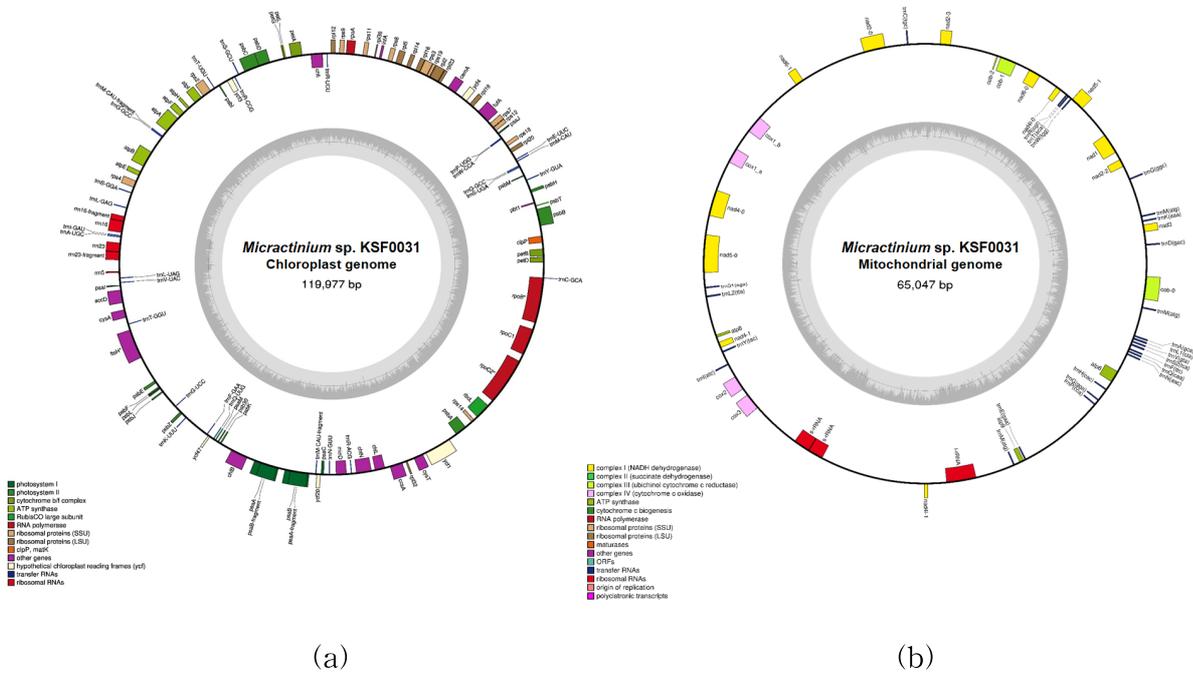


그림 7. Circular map of chloroplast and mitochondrial genome (a) Chloroplast genome (b) Mitochondrial genome

(사) Orthologous group

- KSF0031과 유사한 3종의 algae (*Chlorella variabilis* CHNC64, *Microactinium conductrix* LHPF02, *Chlorella sorokiniana* LHPG02)와 유전자 분석을 진행하였다. OrthoMCL이라는 프로그램을 사용하여 총 4,637 group의 공통 유전자군을 확인 할 수 있었다. 다른 algae 단백질과의 유사도는 *Microactinium conductrix*가 높으나 65% 정도로 상당한 차이를 보이고 있었다 (그림 8).
- 추가적으로 395 유전자 그룹과 2,628개의 KSF0031 특이적인 유전자를 대상으로 특정 GO term이 enrichment되어 있는지 확인하는 분석을 진행하였으며, Protein amino acid phosphorylation, protein de-ubiquitination 그리고 regulation of transcription, DNA-dependent라는 GO term이 KSF0031에서 enrichment되어 있다는 것을 Fisher's exact test ($P < 0.05$) 로 확인할 수 있었다 (그림 9).

Orthologous genes of four algae species

Venn diagram displaying the overlap in orthologous genes of six insect species and the number of unique groups in each species.

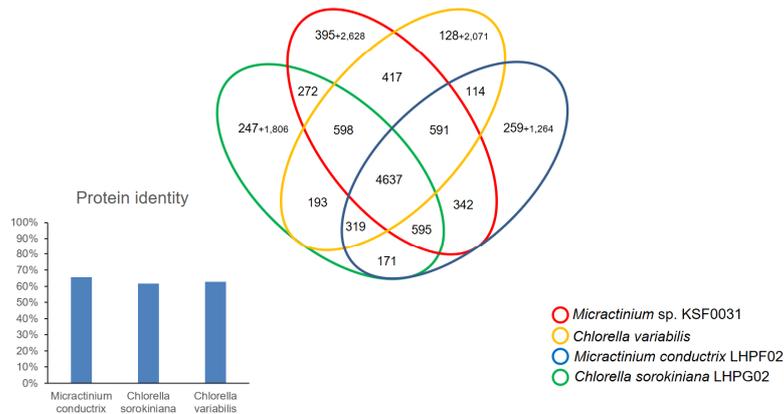
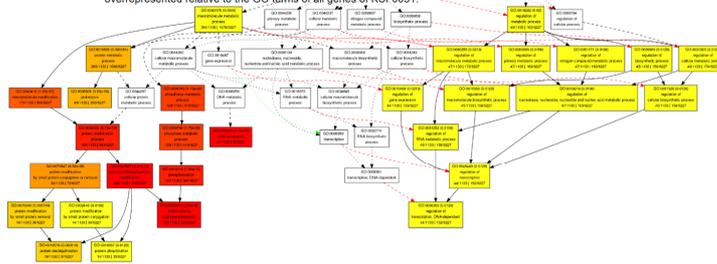


그림 8. Venn diagram showing orthologous protein groups between four algae species

GO enrichment analysis of unique gene in KSF0031

We identify which GO terms of the 395 groups and 2,628 gene that were unique to KSF0031 were statistically overrepresented relative to the GO terms of all genes of KSF0031.



Ontology	Description	Number in input list	Number in BG/Ref	p-value	FDR
Biological Process	protein amino acid phosphorylation	125	346	1.70E-14	2.30E-12
Biological Process	protein deubiquitination	18	37	2.60E-05	0.00052
Biological Process	regulation of transcription, DNA-dependent	44	152	0.00094	0.013

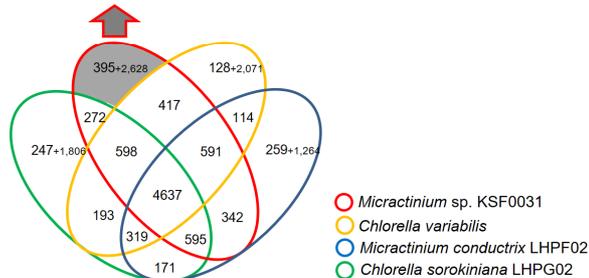


그림 9. GO enrichment analysis of unique gene in KSF0031

(아) KEGG 분석을 통한 지질대사 경로 관련유전자의 비교 분석

○KEGG (Kyoto Encyclopedia of Genes and Genomes)의 정보를 이용하여, 지질대사 경로 관련 유전자의 유무를 비교분석 하였다. KEGG의 organism에 등록되어있는 11개의 green algae 그리고 이전에 orthologous protein의 비교연구에서 사용된 3종의 green algae를 포함해서 총 14개 green algae와 비교하여 지질대사 경로의 유전자의 유무를 확인하였다. ECHS1; enoyl-CoA hydratase [EC:4.2.1.17], alkB1_2, alkM; alkane 1-monoxygenase [EC:1.14.15.3], SGPP1; sphingosine-1-phosphate phosphatase 1

[EC:3.1.3.-], SCP2, SCPX; sterol carrier protein 2 [EC:2.3.1.176], GPAT1_2; glycerol-3-phosphate O-acyltransferase 1/2 [EC:2.3.1.15], THEM4, CTMP; acyl-coenzyme A thioesterase THEM4 [EC:3.1.2.2], CERS1_2_3_4, LASS1_2_3_4; ceramide synthetase [EC:2.3.1.24], MOGAT2, MGAT2; 2-acylglycerol O-acyltransferase 2 [EC:2.3.1.22] 유전자가 KEGG의 green algae 유전체 정보와 비교하여 특이적인 것을 확인할 수 있었으며, 추가적으로 *viaY*; alcohol dehydrogenase [EC:1.1.1.1]와 *adhE*; acetaldehyde dehydrogenase / alcohol dehydrogenase [EC:1.2.1.10 1.1.1.1]를 가지고 있음을 확인할 수 있었다.

metabolism	kegg enzyme	function
Fatty acid metabolism	K00059	bbg, OAR1, 3-oxoacyl [acyl-carrier-protein] reductase [EC:1.1.1.100]
Fatty acid metabolism	K00088	bbd, enoyl [acyl-carrier-protein] reductase [EC:1.1.1.91, 1.1.1.10]
Fatty acid metabolism	K00082	E1.3.3.6, ACO1, ACO3, acyl CoA oxidase [EC:1.3.3.6]
Fatty acid metabolism	K00040	ACADM, acyl-CoA dehydrogenase [EC:1.3.3.7]
Fatty acid metabolism	K00057	SCD, desaturase of CoA dehydratase (Delta 9 dehydratase) [EC:1.14.10.1]
Fatty acid metabolism	K00026	E2.3.1.9, acyl-CoA C-acyl transferase [EC:2.3.1.9]
Fatty acid metabolism	K00045	bbd, MCAT, MCT1, [acyl-carrier-protein] S-malonyl transferase [EC:2.3.1.39]
Fatty acid metabolism	K00048	bbd4, 3-oxoacyl [acyl-carrier-protein] lyase III [EC:2.3.1.180]
Fatty acid metabolism	K00065	FASN, fatty acid synthase, animal type [EC:2.3.1.85]
Fatty acid metabolism	K01074	PPT, palmityl-protein kinase [EC:2.3.1.222]
Fatty acid metabolism	K00045	paaf, acyl-enzyme CoA hydratase [EC:4.2.1.17]
Fatty acid metabolism	K01897	ACSL, long-chain acyl-CoA synthetase [EC:6.2.1.3]
Fatty acid metabolism	K01961	accC, acyl-CoA carboxylase, biofilm carboxylase subunit [EC:6.4.1.26, 3.4.14]
Fatty acid metabolism	K01962	accA, acyl-CoA carboxylase, carboxyl transferase subunit [EC:6.4.1.22, 1.3.15]
Fatty acid metabolism	K01963	accD, acyl-CoA carboxylase, carboxyl transferase subunit [EC:6.4.1.22, 1.3.15]
Fatty acid metabolism	K02160	accB, bcfP, acyl-CoA carboxylase, biofilm carboxyl carrier protein
Fatty acid metabolism	K02272	bbz2, 3-hydroxyacyl [acyl-carrier-protein] dehydratase [EC:4.2.1.59]
Fatty acid metabolism	K03921	FAB2, SSB2, dssA1, acyl [acyl-carrier-protein] dehydratase [EC:1.14.10.21, 14.10.11.1]
Fatty acid metabolism	K07511	ECHE1, enoyl-CoA hydratase [EC:4.2.1.17]
Fatty acid metabolism	K07512	MECR, MRBF1, mitochondrial enoyl [acyl-carrier-protein] reductase / trans-2-enoyl-CoA
Fatty acid metabolism	K07513	ACAA1, acyl-CoA acyl transferase I [EC:2.3.1.16]
Fatty acid metabolism	K08764	SCP2, SCPX, sterol carrier protein 2 [EC:2.3.1.176]
Fatty acid metabolism	K04260	bbf, ODSM, CEM1, 3-oxoacyl [acyl-carrier-protein] lyase III [EC:2.3.1.179]
Fatty acid metabolism	K10226	FADS2, acyl-CoA dehydratase (Delta 6 dehydratase) [EC:1.14.10.3]
Fatty acid metabolism	K10244	ELOVL5 elongation of fatty long chain fatty acids protein 5 [EC:2.3.1.199]
Fatty acid metabolism	K10249	ELOVL4 elongation of fatty long chain fatty acids protein 4 [EC:2.3.1.199]
Fatty acid metabolism	K10251	HSD17B12, KAR, IFAB2, 17beta-oxidation 17-dehydrogenase / very long chain 3-oxoacyl-CoA
Fatty acid metabolism	K11262	FADS2, omega-6 fatty acid dehydratase / acyl lipid omega-6 dehydratase (Delta 12 dehydratase)
Fatty acid metabolism	K10258	TER, TSC1, CER10, very long chain enoyl-CoA reductase [EC:1.3.1.93]
Fatty acid metabolism	K10527	HFP2, enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase [EC:4.2.1.171, 1.1.35, 1.1.1]
Fatty acid metabolism	K10703	HACD, FHS1, PAS2, very long chain 3(omega)-3-hydroxyacyl-CoA dehydratase [EC:4.2.1.134]
Fatty acid metabolism	K11262	ACAC, acyl-CoA carboxylase, biofilm carboxylase [EC:6.4.1.26, 3.4.142, 1.3.15]
Fatty acid metabolism	K13013	ACSBG long-chain-fatty acid-CoA ligase ACSBG [EC:6.2.1.3]
Fatty acid metabolism	K18660	ACSF3, malonyl-CoA: methylmalonyl-CoA synthetase [EC:6.2.1.-]
Fatty acid elongation	K01074	PPT, palmityl-protein kinase [EC:2.3.1.222]
Fatty acid elongation	K07511	ECHE1, enoyl-CoA hydratase [EC:4.2.1.17]
Fatty acid elongation	K07512	MECR, MRBF1, mitochondrial enoyl [acyl-carrier-protein] reductase / trans-2-enoyl-CoA reductase [EC:2.3.1.199]
Fatty acid elongation	K10244	ELOVL5 elongation of fatty long chain fatty acids protein 5 [EC:2.3.1.199]
Fatty acid elongation	K10249	ELOVL4 elongation of fatty long chain fatty acids protein 4 [EC:2.3.1.199]
Fatty acid elongation	K10251	HSD17B12, KAR, IFAB2, 17beta-oxidation 17-dehydrogenase / very long chain 3-oxoacyl-CoA reductase [EC:4.2.1.134]
Fatty acid elongation	K10258	TER, TSC1, CER10, very long chain enoyl-CoA reductase [EC:1.3.1.93]
Fatty acid elongation	K10703	HACD, FHS1, PAS2, very long chain 3(omega)-3-hydroxyacyl-CoA dehydratase [EC:4.2.1.134]
Fatty acid elongation	K15397	KCS, 3-oxoacyl-CoA synthase [EC:2.3.1.199]
Fatty acid elongation	K15398	TEH4, CBF, acetyl-coenzyme A biosynthesis, TEH4 [EC:3.1.2.2]
Biosyn. function	K00022	E1.3.3.6, ACO1, ACO3, acyl CoA oxidase [EC:1.3.3.6]
Biosyn. function	K00057	SCD, desaturase of CoA dehydratase (Delta 9 dehydratase) [EC:1.14.10.1]
Biosyn. function	K00082	FAB2, SSB2, dssA1, acyl [acyl-carrier-protein] dehydratase [EC:1.14.10.21, 14.10.11.1]
Biosyn. function	K00088	ACAA1, acyl-CoA acyl transferase I [EC:2.3.1.16]
Biosyn. function	K08764	SCP2, SCPX, sterol carrier protein 2 [EC:2.3.1.176]
Biosyn. function	K10226	FADS2, acyl-CoA dehydratase (Delta 6 dehydratase) [EC:1.14.10.3]
Biosyn. function	K10244	ELOVL5 elongation of fatty long chain fatty acids protein 5 [EC:2.3.1.199]
Biosyn. function	K10249	ELOVL4 elongation of fatty long chain fatty acids protein 4 [EC:2.3.1.199]
Biosyn. function	K10251	HSD17B12, KAR, IFAB2, 17beta-oxidation 17-dehydrogenase / very long chain 3-oxoacyl-CoA reductase [EC:4.2.1.134]
Biosyn. function	K10258	TER, TSC1, CER10, very long chain enoyl-CoA reductase [EC:1.3.1.93]
Biosyn. function	K10527	HFP2, enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase [EC:4.2.1.171, 1.1.35, 1.1.1]
Biosyn. function	K10703	HACD, FHS1, PAS2, very long chain 3(omega)-3-hydroxyacyl-CoA dehydratase [EC:4.2.1.134]
Biosyn. function	K11262	ACAC, acyl-CoA carboxylase, biofilm carboxylase [EC:6.4.1.26, 3.4.142, 1.3.15]
Biosyn. function	K13013	ACSBG long-chain-fatty acid-CoA ligase ACSBG [EC:6.2.1.3]
Biosyn. function	K18660	ACSF3, malonyl-CoA: methylmalonyl-CoA synthetase [EC:6.2.1.-]
Ethanol metabolism	K00093	ADH1, alcohol dehydrogenase [EC:1.1.1.1]
Ethanol metabolism	K01047	PLAG, SPLA2, acetylserine phospholipase A2 [EC:3.1.1.4]
Ethanol metabolism	K01062	PLA2G7, PAF-AH1, platelet activating factor acetylhydrolase [EC:3.1.1.47]
Ethanol metabolism	K01080	PLPP1, 1,2,3-phosphatidylserine phosphatase [EC:3.1.3.4]
Ethanol metabolism	K01115	PLD1, 2-phosphatidylserine phosphatase [EC:3.1.4.4]
Ethanol metabolism	K04628	CGT, UDPG, ceramide galactosyl transferase [EC:2.4.1.47]
Ethanol metabolism	K06123	AVP1, 1-acylglycerol-3-phosphate reductase [EC:1.1.1.101]
Ethanol metabolism	K13510	LPAT1, 2-lysophosphatidylcholine acyl transferase / lysophosphatidylcholine acyl transferase [EC:2.3.1.232, 2.3.1.67]
Ethanol metabolism	K13519	LPAT, MFL1, lysophosphatidyl acyl transferase [EC:2.3.1.51, 2.3.1.232, 3.1.-]
Ethanol metabolism	K14074	TGL4, TAG lipase / acyl-ester hydrolase / phospholipase A2 / PPA acyl transferase [EC:3.1.1.33, 1.1]
Ethanol metabolism	K16795	PAFAH2B, 2-platelet activating factor acetylhydrolase (B subunit) [EC:3.1.1.47]
Fatty acid degradation	K00021	invA, ADHS, acyl-CoA dehydrogenase of highly branched hydroxygens / alcohol dehydrogenase
Fatty acid degradation	K00128	ALDH4, aldehyde dehydrogenase (NAD+)-[EC:1.2.1.3]
Fatty acid degradation	K00232	E1.3.3.6, ACO1, ACO3, acyl CoA oxidase [EC:1.3.3.6]
Fatty acid degradation	K00240	ACADM, acyl-CoA dehydrogenase [EC:1.3.3.7]
Fatty acid degradation	K00026	bbd1, 2-oxoacyl-CoA dehydratase [EC:1.14.10.3]
Fatty acid degradation	K00026	E2.3.1.9, acyl-CoA C-acyl transferase [EC:2.3.1.9]
Fatty acid degradation	K01092	paaf, acyl-enzyme CoA hydratase [EC:4.2.1.17]
Fatty acid degradation	K01897	ACSL, long-chain acyl-CoA synthetase [EC:6.2.1.3]
Fatty acid degradation	K04872	adH, acyl-CoA dehydrogenase / alcohol dehydrogenase [EC:1.2.1.101, 1.1.1.1]
Fatty acid degradation	K07511	ECHE1, enoyl-CoA hydratase [EC:4.2.1.17]
Fatty acid degradation	K07513	ACAA1, acyl-CoA acyl transferase I [EC:2.3.1.16]
Fatty acid degradation	K08527	HFP2, enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase [EC:4.2.1.171, 1.1.35, 1.1.1]
Fatty acid degradation	K13228	ECHE1, long-chain acyl-CoA hydratase [EC:5.3.3.8]
Fatty acid degradation	K13954	yaP, alcohol dehydrogenase [EC:1.1.1.1]
Fatty acid degradation	K13103	ACSBG long-chain-fatty acid-CoA ligase ACSBG [EC:6.2.1.3]
Fatty acid degradation	K20495	CYP10A8, long-chain fatty acid omega-oxidation cytochrome P-450 [EC:1.14.14.80]
Fatty acid biosynthesis	K00088	bbd, enoyl [acyl-carrier-protein] reductase [EC:1.1.1.91, 1.1.1.10]
Fatty acid biosynthesis	K00045	bbd, MCAT, MCT1, [acyl-carrier-protein] S-malonyl transferase [EC:2.3.1.39]
Fatty acid biosynthesis	K00048	bbd4, 3-oxoacyl [acyl-carrier-protein] lyase III [EC:2.3.1.180]
Fatty acid biosynthesis	K00065	FASN, fatty acid synthase, animal type [EC:2.3.1.85]
Fatty acid biosynthesis	K01071	MCH medium chain acyl [acyl-carrier-protein] hydrolase [EC:3.1.2.21]
Fatty acid biosynthesis	K01897	ACSL, long-chain acyl-CoA synthetase [EC:6.2.1.3]
Fatty acid biosynthesis	K01946	ACACB, acyl-CoA carboxylase, biofilm carboxylase 2 [EC:6.4.1.26, 3.4.142, 1.3.15]
Fatty acid biosynthesis	K01961	accC, acyl-CoA carboxylase, biofilm carboxylase [EC:6.4.1.26, 3.4.14]
Fatty acid biosynthesis	K01962	accA, acyl-CoA carboxylase, carboxyl transferase subunit [EC:6.4.1.22, 1.3.15]
Fatty acid biosynthesis	K01963	accD, acyl-CoA carboxylase, carboxyl transferase subunit [EC:6.4.1.22, 1.3.15]
Fatty acid biosynthesis	K02160	accB, bcfP, acyl-CoA carboxylase, biofilm carboxyl carrier protein
Fatty acid biosynthesis	K03921	bbz2, 3-hydroxyacyl [acyl-carrier-protein] dehydratase [EC:4.2.1.59]
Fatty acid biosynthesis	K03921	FAB2, SSB2, dssA1, acyl [acyl-carrier-protein] dehydratase [EC:1.14.10.21, 14.10.11.1]
Fatty acid biosynthesis	K07512	MECR, MRBF1, mitochondrial enoyl [acyl-carrier-protein] reductase / trans-2-enoyl-CoA
Fatty acid biosynthesis	K07513	ACAA1, acyl-CoA acyl transferase I [EC:2.3.1.16]
Fatty acid biosynthesis	K10703	HACD, FHS1, PAS2, very long chain 3(omega)-3-hydroxyacyl-CoA dehydratase [EC:4.2.1.134]
Fatty acid biosynthesis	K11262	ACAC, acyl-CoA carboxylase, biofilm carboxylase [EC:6.4.1.26, 3.4.142, 1.3.15]
Fatty acid biosynthesis	K13013	ACSBG long-chain-fatty acid-CoA ligase ACSBG [EC:6.2.1.3]
Fatty acid biosynthesis	K18660	ACSF3, malonyl-CoA: methylmalonyl-CoA synthetase [EC:6.2.1.-]
Splitting lipid metabolism	K00854	SPL, acetylserine phosphatase [EC:2.3.1.50]
Splitting lipid metabolism	K01080	PLPP1, 1,2,3-phosphatidylserine phosphatase [EC:3.1.3.4]
Splitting lipid metabolism	K01190	lac2, beta-galactosidase [EC:3.2.1.23]

그림 10. KSF0031 specific genes were in red. *Micractinium* sp. KSF0031. Genome information of 13 Green algae were used in this analysis: *Auxenochlorella protothecoides*, *Volvox carteri f nagariensis*, *Ostreococcus tauri*, *Ostreococcus lucimarinus*, *Monoraphidium neglectum*, *Micromonas pusilla*, *Micromonas commode*, *Coccomyxa subellipsoidea*, *Chlorella variabilis*, *Bathycoccus prasinos*, and *Chlamydomonas Reinhardtii*.

3. 극지 미세조류 이차대사산물의 클러스터분석

(가) 이차대사산물 생합성 유전자 클러스터 예측

- 남극 유래 조류 *Micractinium* sp. KSF00031와 녹조류 근연종인 *Chlorella variabilis*의 이차대사산물 생합성 유전자 클러스터를 antiSMASH v4.2.0 (antibiotics and secondary metabolite analysis shell)를 이용하여 예측하였다 (1). 실행조건에서 ‘bacteria’ 모드를 선택했을 때 더 많은 클러스터가 예측되는 것을 알 수 있었으며, 일부 클러스터들은 ‘fungi’ 모드의 결과와 예측된 영역이 겹치기도 하였다 (표 19-20).

표 19. List of secondary metabolite biosynthetic gene clusters of *C. variabilis* predicted by antiSMASH

Cluster	Type	Position	The Most Similar Known Cluster *
mode ‘fungi’			
Cluster 1	Other	scaffold_9:353822-404524	
Cluster 2	Cf_fatty_acid	scaffold_10:326703-348320	
Cluster 3	Cf_putative	scaffold_14:31015-122545	
mode ‘bacteria’			
Cluster 1	Cf_putative	scaffold_4:1847119-1863888	
Cluster 2	Cf_putative	scaffold_7:576546-683002	
Cluster 3	Nrps	scaffold_8:792014-842561	
Cluster 4	Cf_putative	scaffold_9:333375-443332	Ajudazol biosynthetic gene cluster (BGC0000954)
Cluster 5	Cf_fatty_acid	scaffold_10:326703-348320	
Cluster 6	Cf_fatty_acid	scaffold_10:519248-541235	
Cluster 7	Cf_putative	scaffold_11:1146027-1201146	
Cluster 8	T3pks	scaffold_14:835506-875934	
Cluster 9	Cf_fatty_acid- Arylpolyene	scaffold_18:824030-865534	
Cluster 10	Cf_putative	scaffold_19:186849-239131	
Cluster 11	Cf_putative	scaffold_23:269103-341960	
Cluster 12	Terpene	scaffold_25:445056-465769	
Cluster 13	Cf_putative	scaffold_26:3-18900	
Cluster 14	Cf_putative	scaffold_26:260229-342965	Herboxidiene biosynthetic gene cluster (BGC0001065)
Cluster 15	Terpene	scaffold_26:424996-446661	
Cluster 16	Cf_putative	scaffold_32:259543-281086	
Cluster 17	Cf_putative	scaffold_37:153578-211399	

* The most similar homologous gene cluster of MIBiG (Minimum Information about a Biosynthetic Gene cluster) is shown. The corresponding accession is shown in the parenthesis.

표 20. List of secondary metabolite biosynthetic gene clusters of KSF00031 predicted by antiSMASH

Cluster	Type	Position	The Most Similar Known Cluster *
mode ‘fungi’			
Cluster 1	Cf_putative	utg27:1339463-1401513	
Cluster 2	Cf_putative	utg28:3265876-3481622	

Cluster 3	Cf_putative	utg67:501917-598066	
Cluster 4	Cf_fatty_acid	utg67:988848-1010944	
Cluster 5	Other	utg145:70734-121616	
mode 'bacteria'			
Cluster 1	Cf_fatty_acid	utg13:2251762-2272337	
Cluster 2	Terpene	utg13:4020393-4041319	
Cluster 3	Cf_fatty_acid	utg16:149350-169598	
Cluster 4	Cf_fatty_acid	utg16:170089-192959	
Cluster 5	Cf_putative	utg40:1974968-2063733	
Cluster 6	Cf_putative	utg56:220080-377968	Ajudazol biosynthetic gene cluster (BGC0000954)
Cluster 7	Cf_putative	utg61:2498392-2518551	
Cluster 8	Cf_putative	utg67:494700-578250	Ajudazol biosynthetic gene cluster (BGC0000954)
Cluster 9	Terpene	utg67:721305-744366	
Cluster 10	Cf_fatty_acid	utg67:987965-1008363	
Cluster 11	Cf_putative	utg145:17893-216000	Ajudazol biosynthetic gene cluster (BGC0000954)

* The most similar homologous gene cluster of MIBiG (Minimum Information about a Biosynthetic Gene cluster) is shown. The corresponding accession is shown in the parenthesis.

- 곰팡이 유전체 482개의 antiSMASH 분석 결과에서 유전체 당 평균 44.36개의 클러스터가 예측된 것에 비추어 봤을 때, 이차대사산물 생합성 유전자 클러스터 예측이 생물계 별로 특이적이라는 점을 시사한다(Choi, unpublished data). 미생물의 경우 이차대사산물 생합성을 위한 유전자들이 가까운 영역에 모여 있는 경우가 대부분이지만, 식물의 경우 서로 다른 염색체에 분리되어 존재하는 경우가 많다. 예를 들어 버의 경우 diterpene 생합성에 필요한 유전자들이 염색체 2번과 4번에 각각 존재하며, 이런 이유로 'bacteria' 혹은 'fungi' mode로 algae 및 식물의 이차대사산물 생합성 유전자 클러스터 예측하는 것은 충분하지 않으며 새로운 접근이 요구된다. 이런 한계를 극복하기 위하여 단백질 도메인 탐색과 유전자 발현 데이터를 활용할 수 있는 식물 유전체를 위한 생합성 유전자 예측 프로그램이 개발된 바 있다. 이는 유전자 co-expression을 기반으로 멀리 떨어져 있는 유전자들이 하나의 클러스터로 작용할 것이라는 점에 착안한 것이다. 하지만 최근에 유전자 발현을 측정하기 위하여 널리 사용되는 RNA-seq이 보편적으로 사용해 왔던 3회 반복 (biological replications)으로는 생물학적 해석이 제한되는 문제가 대두되었다. 효모 (*Saccharomyces cerevisiae* BY4741)를 대상으로 RNA-seq 실험이 85% 이상의 true significantly differentially expressed genes (corrected P -values or FDRs ≤ 0.05)를 올바르게 얻어내기 위해서는 20회의 생물학적 반복이 필요한 것으로 나타났다. 또한 모든 RNA-seq 실험에 6회 이상의 반복을 권장하며, differentially expressed genes 분석이 중요한 경우에는 최소 12회 반복을 권장하고 있다. 따라서 이를 보완할 수 있는 식물 및 조류 유전체 분석법 개발 연구가 필요하다.

(나) 이차대사산물 생합성 유전자 추가 발굴

- *C. variabilis*, *C. sorokiniana*, *M. conductrix*와 KSF00031의 유전체로부터 예측된 protein-coding 유전자는 각각 9,791개, 10,384개, 10,070개, 12,734개였다. 각각의 단백질 서열들을 상대방의 단백질 서열에 BLASTP 수행결과를 상동성 40% 및 e-value $< 1e^{-5}$ 기준으로 살펴보면, 각각 5,218개, 6,275개, 6,850개(query: KSF00031)만이 기준을 충족하였다. 또한 reciprocal best BLAST hit 중에서 각각 2,261개, 1,810개, 1,767개의 단백질

서열만이 두 개의 분석 대상 유전체에 single-copy 형태로 공통으로 존재하는 것으로 나타났다. 같은 genus에 속하는 유전체 사이에서도 전체적인 단백질 서열상 상동성은 크지 않음을 알 수 있다. 이차대사산물 생합성에 관여하는 유전자를 추가적으로 탐색하기 위하여 InterPro Scan v5.33-72.0를 활용한 단백질 도메인 예측을 수행하고, 이를 기반으로 KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway 및 gene ontology (GO) 분석을 수행하였다 (5). 이차대사산물 생합성에 연관된 것으로 보이는 69개 GO 및 50개 KEGG accession들을 선별하였으며, 전체 단백질 서열의 도메인 프로파일에서 이들 accession들을 탐색하였다 (표 21-22). 탐색 결과 *C. variabilis*, *C. sorokiniana*, *M. conductrix*와 KSF00031의 유전체로부터 21개 KEGG accession 및 5개 GO accession에 속하는 유전자를 각각 72개, 82개, 75개, 70개 찾아낼 수 있었다 (표 22-23). 이들 유전자들은 다양한 이차대사산물 생합성 경로에 연관된 것으로 나타났으며, 일부 예측된 기능(e.g. terpenoid backbone biosynthesis)의 경우 antiSMASH 'bacteria' mode의 결과와 겹치지는 않았다. antiSMASH가 예측하지 않은 scaffold 영역에서도 이차대사산물 생합성에 관련된 것으로 보이는 유전자들이 다수 발견되었으며, 이는 이차대사산물 생산 잠재력을 평가하기 위한 기초적인 기반으로 활용할 수 있을 것이다. 이를 위해서는 탐색할 핵심 catalytic enzyme 정의, 해당 효소가 속해 있는 생합성 경로의 면밀한 curation 등이 선행되어야 예측 파이프라인을 구축하고 자동화 할 수 있을 것이다.



표 21. The number of protein sequences annotated with KEGG and GO accessions involved in biosynthesis of secondary metabolites

Source	Accession	Description	<i>Chlorella variabilis</i>	<i>Chlorella sorokiniana</i>	<i>Micractinium conductrix</i>	<i>Micractinium</i> sp. KSF00031	
KEGG	00232	Caffeine metabolism	1	3	2	1	
	00254	Aflatoxin biosynthesis	1	2	1	1	
	00261	Monobactam biosynthesis	6	7	8	8	
	00281	Geraniol degradation	3	3	4	2	
	00332	Carbapenem biosynthesis	2	2	2	2	
	00401	Novobiocin biosynthesis	4	6	6	4	
	00405	Phenazine biosynthesis	4	3	3	3	
	00521	Streptomycin biosynthesis	5	5	5	5	
	00524	Neomycin, kanamycin and gentamicin biosynthesis	2	2	2	2	
	00900	Terpenoid backbone biosynthesis	19	18	18	20	
	00903	Limonene and pinene degradation	2	2	2	2	
	00905	Brassinosteroid biosynthesis	1	1	0	0	
	00906	Carotenoid biosynthesis	8	7	8	6	
	00908	Zeatin biosynthesis	1	1	1	1	
	00909	Sesquiterpenoid and triterpenoid biosynthesis	1	4	2	2	
	00940	Phenylpropanoid biosynthesis	0	0	0	0	
	00941	Flavonoid biosynthesis	3	6	2	2	
	00950	Isoquinoline alkaloid biosynthesis	4	6	6	4	
	00960	Tropane, piperidine and pyridine alkaloid biosynthesis	5	7	7	5	
	00966	Glucosinolate biosynthesis	2	2	2	2	
	00999	Biosynthesis of secondary metabolites - unclassified	2	3	3	4	
	GO	0008299	isoprenoid biosynthetic process	9	8	8	8
		0008521	acetyl-CoA transmembrane transporter activity	1	0	1	1
		0016104	triterpenoid biosynthetic process	1	1	0	0
		0016114	terpenoid biosynthetic process	4	4	4	4
		0016117	carotenoid biosynthetic process	5	6	5	4
	The number of distinct sequences			72	82	75	70

표 22. List of GO terms involved in biosynthesis of secondary metabolites*

Accession	Description
0000036	acyl carrier activity
0006084	acetyl-CoA metabolic process
0006715	farnesol biosynthetic process
0006718	juvenile hormone biosynthetic process
0008299	isoprenoid biosynthetic process
0008521	acetyl-CoA transmembrane transporter activity
0009686	gibberellin biosynthetic process
0009701	isoflavonoid phytoalexin biosynthetic process
0009709	terpenoid indole alkaloid biosynthetic process
0009716	flavonoid phytoalexin biosynthetic process
0009717	isoflavonoid biosynthetic process
0009813	flavonoid biosynthetic process
0009962	regulation of flavonoid biosynthetic process
0009963	positive regulation of flavonoid biosynthetic process
0009964	negative regulation of flavonoid biosynthetic process
0010263	tricyclic triterpenoid biosynthetic process
0010682	cinnamic acid biosynthetic process involved in flavonoid metabolism
0010686	tetracyclic triterpenoid biosynthetic process
0015876	acetyl-CoA transport
0016099	monoterpenoid biosynthetic process
0016102	diterpenoid biosynthetic process
0016104	triterpenoid biosynthetic process
0016106	sesquiterpenoid biosynthetic process
0016109	tetraterpenoid biosynthetic process
0016112	polyterpenoid biosynthetic process
0016114	terpenoid biosynthetic process
0016117	carotenoid biosynthetic process
0016135	saponin biosynthetic process
0019745	pentacyclic triterpenoid biosynthetic process
0031525	menthol biosynthetic process
0033332	ent-kaurene biosynthetic process
0033385	geranylgeranyl diphosphate metabolic process
0035348	acetyl-CoA transmembrane transport
0035440	tuberculosinol biosynthetic process
0043612	isoprene biosynthetic process
0043646	cephalosporin biosynthetic process
0043693	monoterpene biosynthetic process
0044550	secondary metabolite biosynthetic process
0045338	farnesyl diphosphate metabolic process
0046211	(+)-camphor biosynthetic process
0046246	terpene biosynthetic process
0046248	alpha-pinene biosynthetic process
0046250	limonene biosynthetic process
0051483	terpenoid biosynthetic process, mevalonate-independent
0051484	isopentenyl diphosphate biosynthetic process, methylerythritol 4-phosphate pathway involved in terpenoid biosynthetic process
0051485	terpenoid biosynthetic process, mevalonate-dependent
0051486	isopentenyl diphosphate biosynthetic process, mevalonate pathway involved in terpenoid biosynthetic process
0051502	diterpene phytoalexin biosynthetic process
0051504	diterpene phytoalexin precursor biosynthetic process pathway
0051553	flavone biosynthetic process

0051555	flavonol biosynthetic process
0051556	leucoanthocyanidin metabolic process
0051557	leucoanthocyanidin biosynthetic process
0051559	phlobaphene biosynthetic process
0051762	sesquiterpene biosynthetic process
0051975	lysine biosynthetic process via alpha-aminoadipate and saccharopine
0052696	flavonoid glucuronidation
0062032	cichorine biosynthetic process
1900376	regulation of secondary metabolite biosynthetic process
1900377	negative regulation of secondary metabolite biosynthetic process
1900378	positive regulation of secondary metabolite biosynthetic process
1900384	regulation of flavonol biosynthetic process
1900385	negative regulation of flavonol biosynthetic process
1900386	positive regulation of flavonol biosynthetic process
1900947	regulation of isoprene biosynthetic process
1900948	negative regulation of isoprene biosynthetic process
1900949	positive regulation of isoprene biosynthetic process
1903193	sesquiterpene biosynthetic process
1990480	obsolete geranyl diphosphate synthase

* GO terms were selected by keywords search: ((secondary metabolite || terpen || cartenoi || flavon) && biosynthe)

표 23. List of KEGG accessions involved in biosynthesis of secondary metabolites*

Accession	Description
00231	Puromycin biosynthesis
00232	Caffeine metabolism
00253	Tetracycline biosynthesis
00254	Aflatoxin biosynthesis
00261	Monobactam biosynthesis
00281	Geraniol degradation
00311	Penicillin and cephalosporin biosynthesis
00331	Clavulanic acid biosynthesis
00332	Carbapenem biosynthesis
00333	Prodigiosin biosynthesis
00401	Novobiocin biosynthesis
00402	Benzoxazinoid biosynthesis
00403	Indole diterpene alkaloid biosynthesis
00404	Staurosporine biosynthesis
00405	Phenazine biosynthesis
00521	Streptomycin biosynthesis
00522	Biosynthesis of 12-, 14- and 16-membered macrolides
00523	Polyketide sugar unit biosynthesis
00524	Neomycin, kanamycin and gentamicin biosynthesis
00525	Acarbose and validamycin biosynthesis
00900	Terpenoid backbone biosynthesis
00901	Indole alkaloid biosynthesis
00902	Monoterpenoid biosynthesis
00903	Limonene and pinene degradation
00904	Diterpenoid biosynthesis
00905	Brassinosteroid biosynthesis
00906	Carotenoid biosynthesis
00908	Zeatin biosynthesis
00909	Sesquiterpenoid and triterpenoid biosynthesis

00940	Phenylpropanoid biosynthesis
00941	Flavonoid biosynthesis
00942	Anthocyanin biosynthesis
00943	Isoflavonoid biosynthesis
00944	Flavone and flavonol biosynthesis
00945	Stilbenoid, diarylheptanoid and gingerol biosynthesis
00950	Isoquinoline alkaloid biosynthesis
00960	Tropane, piperidine and pyridine alkaloid biosynthesis
00965	Betalain biosynthesis
00966	Glucosinolate biosynthesis
00981	Insect hormone biosynthesis
00999	Biosynthesis of secondary metabolites - unclassified
01051	Biosynthesis of ansamycins
01052	Type I polyketide structures
01053	Biosynthesis of siderophore group nonribosomal peptides
01054	Nonribosomal peptide structures
01055	Biosynthesis of vancomycin group antibiotics
01056	Biosynthesis of type II polyketide backbone
01057	Biosynthesis of type II polyketide products
01058	Acridone alkaloid biosynthesis
01059	Biosynthesis of enediyne antibiotics

* KEGG accessions under “Metabolism of terpenoids and polyketides (09109)” and “Biosynthesis of other secondary metabolites (09110)” were listed here.

표 24. List of loci with the target KEGG and/or GO accessions in KSF00031

Locus*	Scaffold	GO term	KEGG accessions	Enzymes involved
augustus-utg20-processed-gene-2.9-mRNA-1	utg20		00405	4.1.3.27
augustus-utg21-processed-gene-21.3-mRNA-1	utg21		00906	5.2.1.14
augustus-utg235-processed-gene-0.3-mRNA-1	utg235		00281 00903	4.2.1.17+1.1.1.35 4.2.1.17
augustus-utg258-processed-gene-0.19-mRNA-1	utg258		00401,00950,00960	2.6.1.1
augustus-utg26-processed-gene-0.12-mRNA-1	utg26		00908	2.5.1.75
augustus-utg27-processed-gene-21.11-mRNA-1	utg27		00900	2.7.4.26
augustus-utg27-processed-gene-9.2-mRNA-1	utg27		00900	3.4.24.84
augustus-utg36-processed-gene-8.6-mRNA-1	utg36		00909	1.14.14.17
augustus-utg45-processed-gene-10.7-mRNA-1	utg45		00401	1.3.1.12
augustus-utg45-processed-gene-22.11-mRNA-1	utg45		00254	6.4.1.2
augustus-utg47-processed-gene-13.10-mRNA-1	utg47	0016117	00906	1.3.5.6
augustus-utg56-processed-gene-6.7-mRNA-1	utg56	0016114	00900	2.7.1.148
augustus-utg61-processed-gene-8.11-mRNA-1	utg61		00232	1.7.3.3
augustus-utg7-processed-gene-42.2-mRNA-1	utg7		00405	4.1.3.27
maker-utg117-augustus-gene-4.21-mRNA-1	utg117		00900	2.5.1.87
maker-utg117-augustus-gene-6.46-mRNA-1	utg117	0008299		
maker-utg13-augustus-gene-36.16-mRNA-1	utg13		00521,00524	2.7.1.2
maker-utg13-augustus-gene-47.14-mRNA-1	utg13	0008299		
maker-utg13-augustus-gene-5.28-mRNA-1	utg13		00261	2.7.2.4
maker-utg146-augustus-gene-0.23-mRNA-1	utg146		00900	2.7.7.60
maker-utg16-augustus-gene-23.31-mRNA-1	utg16		00281 00903	4.2.1.17+1.1.1.35 4.2.1.17
maker-utg172-augustus-gene-0.11-mRNA-1	utg172	0008299		
maker-utg172-augustus-gene-4.19-mRNA-1	utg172		00950,00960	1.4.3.21
maker-utg172-augustus-gene-5.29-mRNA-1	utg172		00261	1.17.1.8

maker-utg19-augustus-gene-17.20-mRNA-1	utg19	0016117	00906	5.2.1.13
maker-utg20-augustus-gene-3.28-mRNA-1	utg20		00941	5.5.1.6
maker-utg21-augustus-gene-35.18-mRNA-1	utg21		00261	2.7.7.4
maker-utg21-augustus-gene-36.29-mRNA-1	utg21		00966	2.6.1.42
maker-utg235-augustus-gene-1.20-mRNA-1	utg235		00999	2.3.1.30
maker-utg27-augustus-gene-15.26-mRNA-1	utg27		00900	1.3.1.83
maker-utg27-augustus-gene-19.16-mRNA-1	utg27		00966	2.6.1.42
maker-utg28-augustus-gene-27.33-mRNA-1	utg28		00999	2.3.1.30
maker-utg28-augustus-gene-27.34-mRNA-1	utg28		00999	2.3.1.30
maker-utg28-augustus-gene-31.31-mRNA-1	utg28		00900	2.2.1.7
maker-utg28-augustus-gene-35.20-mRNA-1	utg28		00950,00960	1.4.3.21
maker-utg28-augustus-gene-41.20-mRNA-1	utg28		00261	2.7.2.4
maker-utg28-augustus-gene-6.17-mRNA-1	utg28		00261	2.7.7.4
maker-utg28-augustus-gene-6.21-mRNA-1	utg28		00261	2.7.7.4
maker-utg29-augustus-gene-13.20-mRNA-1	utg29		00941	5.5.1.6
maker-utg29-augustus-gene-4.21-mRNA-1	utg29	0008521		
maker-utg38-augustus-gene-18.32-mRNA-1	utg38		00900	3.4.24.84
maker-utg38-augustus-gene-21.59-mRNA-1	utg38		00900	1.17.7.4
maker-utg40-augustus-gene-15.21-mRNA-1	utg40	0016117	00906	1.3.5.5
maker-utg40-augustus-gene-17.14-mRNA-1	utg40		00332	2.7.2.11
maker-utg40-augustus-gene-8.12-mRNA-1	utg40		00521	3.1.3.25
maker-utg45-augustus-gene-19.29-mRNA-1	utg45		00521	5.5.1.4
maker-utg45-augustus-gene-6.37-mRNA-1	utg45	0008299	00900	2.3.3.10
maker-utg49-augustus-gene-1.18-mRNA-1	utg49		00900	2.5.1.58
maker-utg55-augustus-gene-1.43-mRNA-1	utg55	0008299	00900	2.7.7.60
maker-utg57-augustus-gene-4.18-mRNA-1	utg57		00401,00960	2.6.1.9
maker-utg58-augustus-gene-13.28-mRNA-1	utg58		00521,00524	2.7.1.1
maker-utg58-augustus-gene-5.43-mRNA-1	utg58	0008299	00900	1.1.1.267
maker-utg61-augustus-gene-23.24-mRNA-1	utg61		00401,00950,00960	2.6.1.1
maker-utg61-augustus-gene-24.30-mRNA-1	utg61		00521	3.1.3.25
maker-utg61-augustus-gene-25.23-mRNA-1	utg61	0016117		
maker-utg61-augustus-gene-7.29-mRNA-1	utg61		00900	2.1.1.100
maker-utg63-augustus-gene-6.19-mRNA-1	utg63		00909	2.5.1.21
maker-utg67-augustus-gene-5.22-mRNA-1	utg67		00261	4.3.3.7
maker-utg67-augustus-gene-5.23-mRNA-1	utg67	0016114	00900	2.2.1.7
maker-utg77-augustus-gene-15.11-mRNA-1	utg77		00906	5.2.1.14
maker-utg77-augustus-gene-15.20-mRNA-1	utg77		00261	1.2.1.11
maker-utg77-augustus-gene-5.20-mRNA-1	utg77		00906	1.23.5.1
maker-utg79-augustus-gene-0.12-mRNA-1	utg79	0016114	00900	1.17.7.3
maker-utg79-augustus-gene-2.28-mRNA-1	utg79		00405	4.1.3.27
maker-utg7-augustus-gene-5.15-mRNA-1	utg7		00332	1.2.1.41
maker-utg81-augustus-gene-8.15-mRNA-1	utg81		00900	3.4.24.84
maker-utg81-augustus-gene-8.23-mRNA-1	utg81	0016114	00900	4.6.1.12
maker-utg84-augustus-gene-4.49-mRNA-1	utg84	0008299	00900	2.5.1.10+2.5.1.1
maker-utg8-augustus-gene-11.38-mRNA-1	utg8	0008299	00900	5.3.3.2
maker-utg94-augustus-gene-10.21-mRNA-1	utg94		00999	2.3.1.30

* Locus names were shortened by removing repetitive strings: e.g. maker-utg8_pilon_pilon_pilon_pilon_pilon-augustus-gene-11.38-mRNA-1 to maker-utg8-augustus-gene-11.38-mRNA-1.

4. AI 기반 기능 예측

(가) ORF 분석

○ KOPRI에서 분석한 전장유전체 염기서열 분석 결과를 활용하여 물질 합성에 관련된 유전자 정보 획득하였고, 유전자가 코딩하고 있는 단백질의 구조 예측을 통한 활성을 예측하였다.

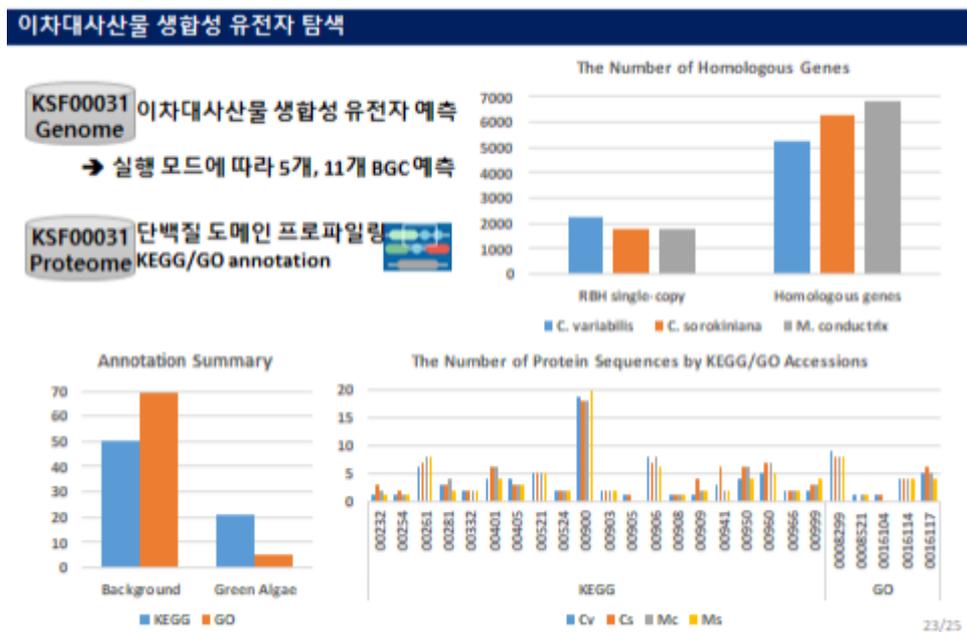


그림 11. KSF0031 유전체 분석을 통한 이차대사산물 유전자 예측

(나) 기능성 예측

- in silico 화합물 DB 구축: 이미 보고된 화합물, 기능성 정보와 본 연구의 결과를 통합하였다. 화합물의 프로파일링을 위해 HPLC, LC-MS, NMR 활용하였다.
- 기능성 예측: 화합물 정보로부터 표적 단백질 예측 및 기능성을 예측하였고, 이를 위해 향산화, 세포 독성, 면역 활성 평가등이 활용되었다.
- 생합성 유전자 확보: ORF 분석을 통하여 천연물 생합성 유전자 확보하였고, in silico 화합물 DB를 활용하여 이차대사산물의 기능을 AI기반 예측하였다.

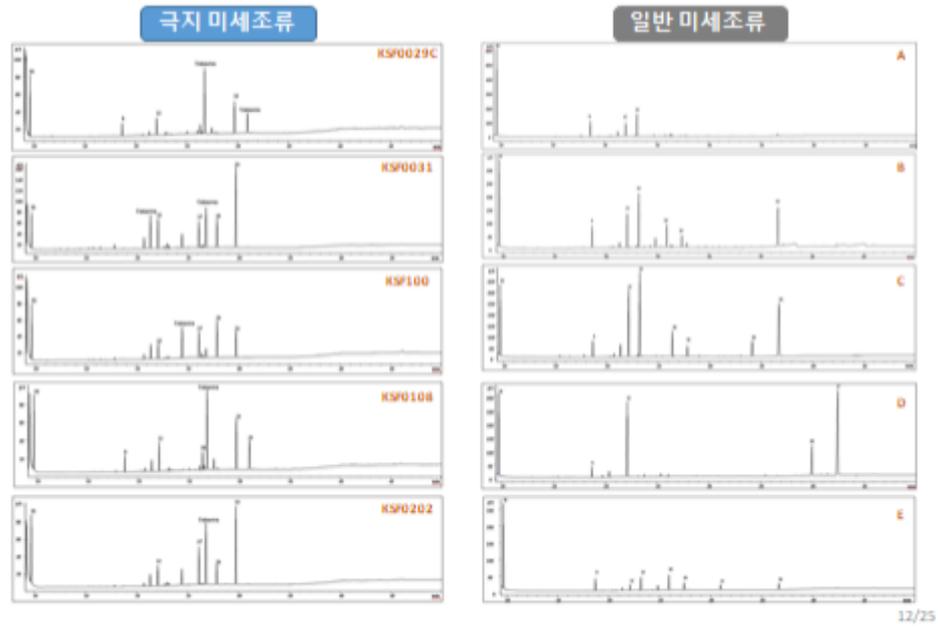


그림 12. 극지 미세조류 5종의 GC를 이용한 지방산 프로파일

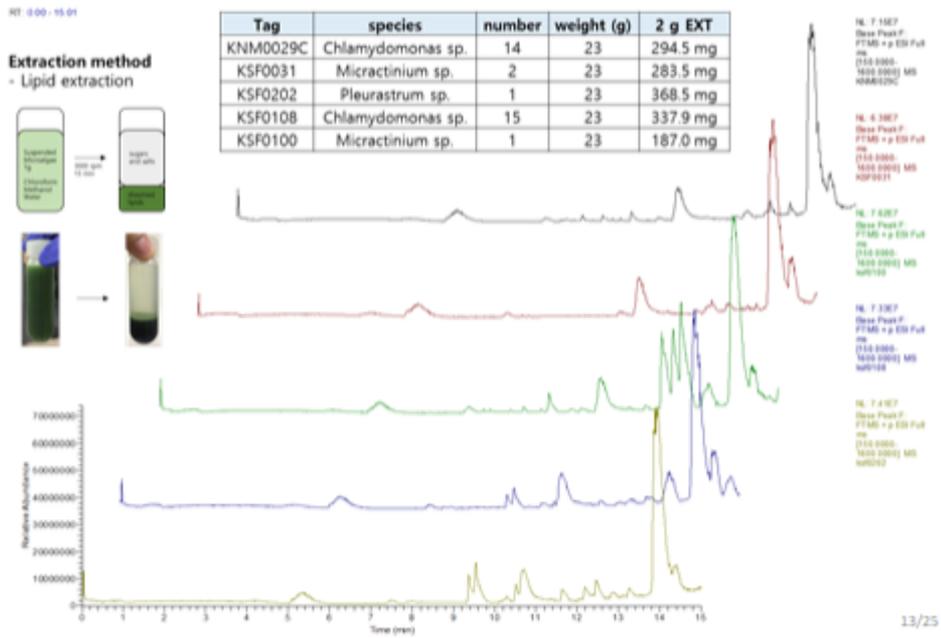


그림 13. 극지 미세조류 5종의 LC-MS를 이용한 지방산 프로파일

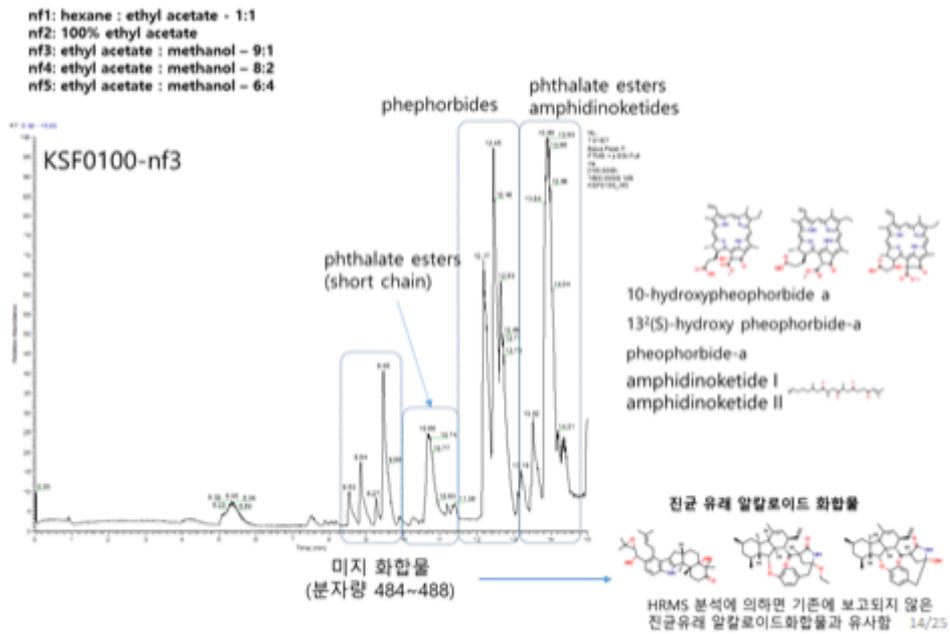
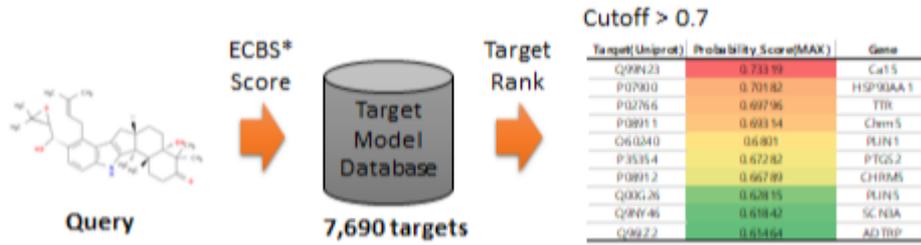


그림 14. KSF0031의 화합물 DB 확보

분자량	분자식	예상화합물	검출 미세조류
235	C10H21O5N		KNM0029C
278	C16H22O4	phthalate	KSF0031, 0100, 0108
350			KNM0029C
351	C20H33O4N	quinoline/pyrrolizidine	KNM0029C
379			KNM0029C
381	C22H39O4N		KNM0029C
390	C24H38O4	phthalate	KNM0029C, KSF0100, 0108
465	C27H47O1N1S2		KNM0029C
465	C28H49O6N		
471	C26 H49O6N		KSF0108
473	C26H51O6N		KNM0029C
484	C25H40O9	glycolipid	KNM0029C
487			KSF0031
488	C25H44O9		KSF0100
489	C31H39O4N	pyrrospirone	KSF0031, 0100
491	C31H41O4N		KSF0031, 0100
493	C31H43O4N	terpene	KNM0029C, KSF0202
495	C31H45O4N	quinone	KSF0031, 0202
517	C33H43N04	pyrrocidine	KSF0031, 0100
519	C33H45O4N	pyrrocidine	KSF0031, 0100, 0108, 0202
520	C30H51O6N		KSF0108
521			KSF0202
541			KSF0202
592	C35H36O5N4	phosphorbide	KSF0100
608	C35H36O6N4	phosphorbide	KSF0100, 0108
624	C35H36O7N4	phosphorbide	KSF0108
638	C35H34O8N4		KSF0108
725	C44H71O7N		KNM0029C

그림 15. 극지 미세조류 이차대사산물의 화합물 정보



*Evolutionary chemical binding similarity, Park et al., NAR, 2019

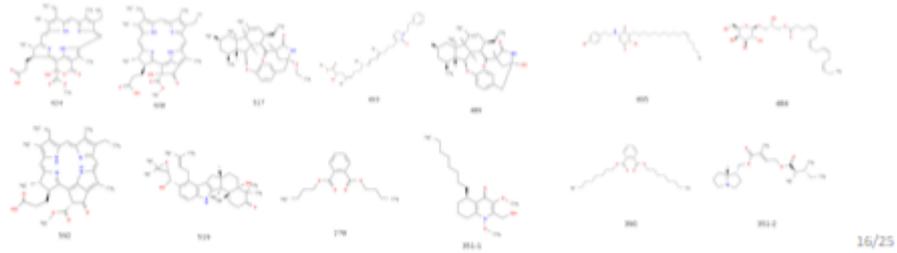


그림 16. 화합물 구조에 기반한 표적 단백질 예측

Chemical	Target(Uniprot)	Probability_Score(MAX)	Probability_Score(Avg)	Protein	Gene	Organism
592	P25101	0.70432	0.2542	Endothelin-1 receptor	EDNRA	Homo sapiens (Human).
592	P28088	0.59997	0.2264	Endothelin B receptor	EDNRB	Bos taurus (Bovine).
592	P25021	0.59143	0.1599	Histamine H2 receptor	HRH2	Homo sapiens (Human).
592	Q15904	0.58182	0.5818	V-type proton ATPase subunit S1	ATP6AP1	Homo sapiens (Human).
592	P51684	0.58153	0.2729	C-C chemokine receptor type 6	CCR6	Homo sapiens (Human).
592	P15502	0.57698	0.577	Elastin	ELN	Homo sapiens (Human).
592	P21918	0.57524	0.2277	D(1B) dopamine receptor	DRD5	Homo sapiens (Human).
592	P18825	0.57263	0.2354	Alpha-2C adrenergic receptor	ADRA2C	Homo sapiens (Human).
592	Q9H773	0.55346	0.5535	dCTP pyrophosphatase 1 [ECO:000305]::0000312 HGNC:		Homo sapiens (Human).
592	P22086	0.5513	0.1475	Alpha-2C adrenergic receptor	Adra2c	Rattus norvegicus (Rat).

Chemical	Target(Uniprot)	Probability_Score(MAX)	Probability_Score(Avg)	Protein	Gene	Organism
519	P35368	0.83999	0.2978	Alpha-1B adrenergic receptor	ADRA1B	Homo sapiens (Human).
519	P28223	0.79846	0.3928	5-hydroxytryptamine receptor 2A	HTR2A	Homo sapiens (Human).
519	Q9HC97	0.76632	0.1651	G-protein coupled receptor 35	GPR35	Homo sapiens (Human).
519	P41144	0.72818	0.3905	Kappa-type opioid receptor	OPRK1	Cavia porcellus (Guinea pig).
519	P34975	0.72713	0.3192	Kappa-type opioid receptor	Oprk1	Rattus norvegicus (Rat).
519	P08173	0.72263	0.3778	Muscarinic acetylcholine receptor M4	CHRM4	Homo sapiens (Human).
519	P43119	0.72006	0.2453	Prostacyclin receptor	PTGIR	Homo sapiens (Human).
519	P48039	0.7132	0.32	Melatonin receptor type 1A	MTNR1A	Homo sapiens (Human).
519	P33533	0.71301	0.3389	Delta-type opioid receptor	Oprd1	Rattus norvegicus (Rat).
519	Q63931	0.70062	0.3261	Cholecystokinin receptor type A	CCKAR	Cavia porcellus (Guinea pig).

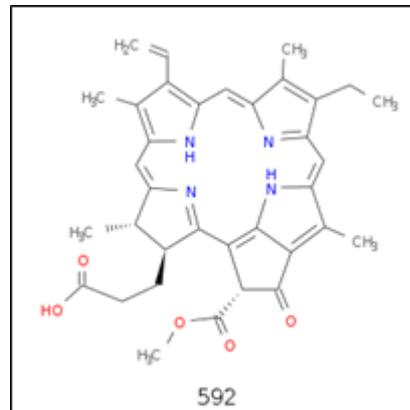
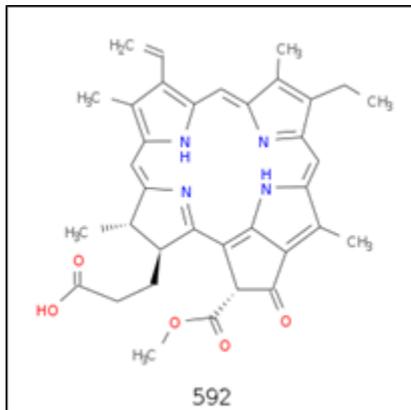


그림 17. 이차대사산물의 기능성 정보 및 표적 예측1

Chemical	Target(Uniprot)	Probability_Score(MAX)	Probability_Score(Avg)	Protein	Gene	Organism
517	P28223	0.83525	0.3933	5-hydroxytryptamine receptor 2A	HTR2A	Homo sapiens (Human).
517	P08909	0.75598	0.3991	5-hydroxytryptamine receptor 2C	Htr2c	Rattus norvegicus (Rat).
517	P18130	0.72375	0.2871	Alpha-1A adrenergic receptor	ADRA1A	Bos taurus (Bovine).
517	Q9QZN9	0.70791	0.409	Cannabinoid receptor 2	Cnr2	Rattus norvegicus (Rat).
517	P35462	0.70739	0.2493	D(3) dopamine receptor	DRD3	Homo sapiens (Human).
517	P16473	0.68474	0.4904	Thyrotropin receptor	TSHR	Homo sapiens (Human).
517	O43193	0.66692	0.244	Motilin receptor	MLNR	Homo sapiens (Human).
517	Q9HBW0	0.6614	0.2059	Lysophosphatidic acid receptor 2	LPAR2	Homo sapiens (Human).
517	P51679	0.6603	0.2013	C-C chemokine receptor type 4	CCR4	Homo sapiens (Human).
517	P79218	0.65858	0.3472	Substance-K receptor	TACR2	Oryctolagus cuniculus (Rabbit).
Chemical	Target(Uniprot)	Probability_Score(MAX)	Probability_Score(Avg)	Protein	Gene	Organism
495	P28223	0.79549	0.3263	5-hydroxytryptamine receptor 2A	HTR2A	Homo sapiens (Human).
495	P35368	0.78658	0.3158	Alpha-1B adrenergic receptor	ADRA1B	Homo sapiens (Human).
495	P23945	0.76513	0.1358	Follicle-stimulating hormone receptor	FSHR	Homo sapiens (Human).
495	P23944	0.74577	0.2771	Alpha-1D adrenergic receptor	Adra1d	Rattus norvegicus (Rat).
495	P41145	0.69327	0.3022	Kappa-type opioid receptor	OPRK1	Homo sapiens (Human).
495	P18090	0.69322	0.3677	Beta-1 adrenergic receptor	Adrb1	Rattus norvegicus (Rat).
495	P18031	0.67825	0.2788	Ine-protein phosphatase non-receptor tyrosine kinase	PTPN1	Homo sapiens (Human).
495	Q13258	0.66607	0.2214	Prostaglandin D2 receptor	PTGDR	Homo sapiens (Human).
495	P28646	0.66283	0.3283	Somatostatin receptor type 1	Sstr1	Rattus norvegicus (Rat).
495	P32239	0.65976	0.2201	Gastrin/cholecystokinin type B receptor	CCKBR	Homo sapiens (Human).

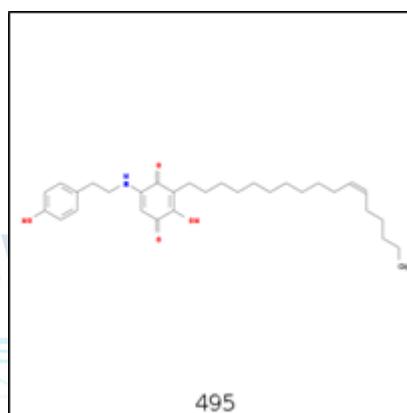
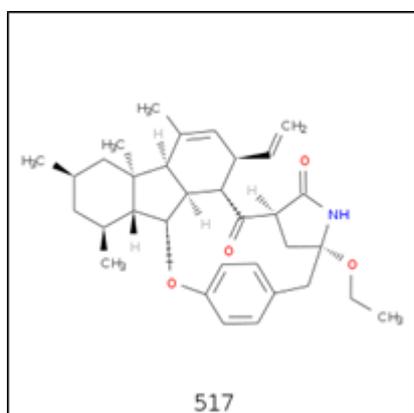
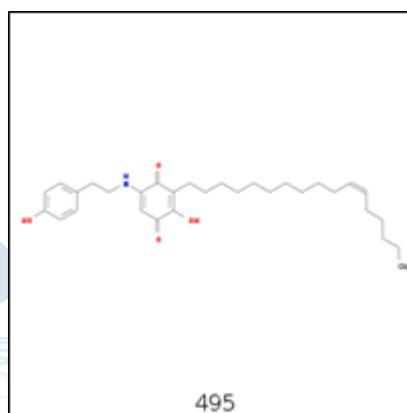
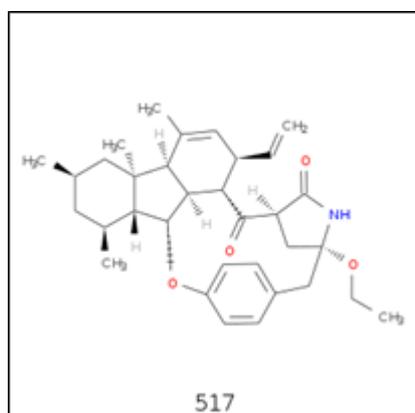


그림 18. 이차대사산물의 기능성 정보 및 표적 예측2

Chemical	Target(Uniprot)	Probability_Score(MAX)	Probability_Score(Avg)	Protein	Gene	Organism
517	P28223	0.83525	0.3933	5-hydroxytryptamine receptor 2A	HTR2A	Homo sapiens (Human).
517	P08909	0.75598	0.3991	5-hydroxytryptamine receptor 2C	Htr2c	Rattus norvegicus (Rat).
517	P18130	0.72375	0.2871	Alpha-1A adrenergic receptor	ADRA1A	Bos taurus (Bovine).
517	Q9QZN9	0.70791	0.409	Cannabinoid receptor 2	Cnr2	Rattus norvegicus (Rat).
517	P35462	0.70739	0.2493	D(3) dopamine receptor	DRD3	Homo sapiens (Human).
517	P16473	0.68474	0.4904	Thyrotropin receptor	TSHR	Homo sapiens (Human).
517	O43193	0.66692	0.244	Motilin receptor	MLNR	Homo sapiens (Human).
517	Q9HBW0	0.6614	0.2059	Lysophosphatidic acid receptor 2	LPAR2	Homo sapiens (Human).
517	P51679	0.6603	0.2013	C-C chemokine receptor type 4	CCR4	Homo sapiens (Human).
517	P79218	0.65858	0.3472	Substance-K receptor	TACR2	Oryctolagus cuniculus (Rabbit).

Chemical	Target(Uniprot)	Probability_Score(MAX)	Probability_Score(Avg)	Protein	Gene	Organism
495	P28223	0.79549	0.3263	5-hydroxytryptamine receptor 2A	HTR2A	Homo sapiens (Human).
495	P35368	0.78658	0.3158	Alpha-1B adrenergic receptor	ADRA1B	Homo sapiens (Human).
495	P23945	0.76513	0.1358	Follicle-stimulating hormone receptor	FSHR	Homo sapiens (Human).
495	P23944	0.74577	0.2771	Alpha-1D adrenergic receptor	Adra1d	Rattus norvegicus (Rat).
495	P41145	0.69327	0.3022	Kappa-type opioid receptor	OPRK1	Homo sapiens (Human).
495	P18090	0.69322	0.3677	Beta-1 adrenergic receptor	Adrb1	Rattus norvegicus (Rat).
495	P18031	0.67825	0.2788	Ine-protein phosphatase non-receptor tyrosine	PTPN1	Homo sapiens (Human).
495	Q13258	0.66607	0.2214	Prostaglandin D2 receptor	PTGDR	Homo sapiens (Human).
495	P28646	0.66283	0.3283	Somatostatin receptor type 1	Sstr1	Rattus norvegicus (Rat).
495	P32239	0.65976	0.2201	Gastrin/cholecystokinin type B receptor	CCKBR	Homo sapiens (Human).



극지연구소

그림 19. 이차대사산물의 기능성 정보 및 표적 예측3

Chemical	Target(Uniprot)	Probability_Score(MAX)	Probability_Score(Avg)	Protein	Gene	Organism
351-1	P30968	0.80496	0.2139	gonadotropin-releasing hormone receptc	GNRHR	Homo sapiens (Human).
351-1	P28223	0.76985	0.3169	5-hydroxytryptamine receptor 2A	HTR2A	Homo sapiens (Human).
351-1	Q96P88	0.73509	0.4531	ve gonadotropin-releasing hormone II re	GNRHR2	Homo sapiens (Human).
351-1	P19328	0.72426	0.2391	Alpha-2B adrenergic receptor	Adra2b	Rattus norvegicus (Rat).
351-1	P41597	0.70254	0.1435	C-C chemokine receptor type 2	CCR2	Homo sapiens (Human).
351-1	P08912	0.69635	0.2446	Muscarinic acetylcholine receptor M5	CHRM5	Homo sapiens (Human).
351-1	P56718	0.67863	0.3676	Orexin receptor type 1	Hcrtr1	Rattus norvegicus (Rat).
351-1	P21462	0.67418	0.1558	fMet-Leu-Phe receptor	FPR1	Homo sapiens (Human).
351-1	P18130	0.66898	0.271	Alpha-1A adrenergic receptor	ADRA1A	Bos taurus (Bovine).
351-1	Q9QZN9	0.66525	0.5285	Cannabinoid receptor 2	Cnr2	Rattus norvegicus (Rat).

Chemical	Target(Uniprot)	Probability_Score(MAX)	Probability_Score(Avg)	Protein	Gene	Organism
278	Q99N23	0.73319	0.4408	Carbonic anhydrase 15	Ca15	Mus musculus (Mouse).
278	P07900	0.70182	0.4287	Heat shock protein HSP 90-alpha	HSP90AA1	Homo sapiens (Human).
278	P02766	0.69796	0.2403	Transthyretin	TTR	Homo sapiens (Human).
278	P08911	0.69314	0.2123	Muscarinic acetylcholine receptor M5	Chrm5	Rattus norvegicus (Rat).
278	O60240	0.6801	0.3536	Perilipin-1	PLIN1	Homo sapiens (Human).
278	P35354	0.67282	0.1366	Prostaglandin G/H synthase 2	PTGS2	Homo sapiens (Human).
278	P08912	0.66789	0.2136	Muscarinic acetylcholine receptor M5	CHRM5	Homo sapiens (Human).
278	Q00G26	0.62815	0.3547	Perilipin-5	PLIN5	Homo sapiens (Human).
278	Q9NV46	0.61842	0.3878	dium channel protein type 3 subunit alp	SCN3A	Homo sapiens (Human).
278	Q96I22	0.61464	0.5964	drogen-dependent TFPI-regulating proti	ADTRP	Homo sapiens (Human).

Chemical	Target(Uniprot)	Probability_Score(MAX)	Probability_Score(Avg)	Protein	Gene	Organism
351-2	P18825	0.99089	0.6424	Alpha-2C adrenergic receptor	ADRA2C	Homo sapiens (Human).
351-2	P18089	0.98326	0.5843	Alpha-2B adrenergic receptor	ADRA2B	Homo sapiens (Human).
351-2	P31388	0.98199	0.5366	5-hydroxytryptamine receptor 6	Htr6	Rattus norvegicus (Rat).
351-2	P41595	0.97194	0.5358	5-hydroxytryptamine receptor 2B	HTR2B	Homo sapiens (Human).
351-2	P08173	0.96651	0.7199	Muscarinic acetylcholine receptor M4	CHRM4	Homo sapiens (Human).
351-2	P10980	0.96048	0.6366	Muscarinic acetylcholine receptor M2	Chrm2	Rattus norvegicus (Rat).
351-2	P08909	0.96025	0.6001	5-hydroxytryptamine receptor 2C	Htr2c	Rattus norvegicus (Rat).
351-2	P25021	0.95766	0.5038	Histamine H2 receptor	HRH2	Homo sapiens (Human).
351-2	P28223	0.95614	0.6706	5-hydroxytryptamine receptor 2A	HTR2A	Homo sapiens (Human).
351-2	Q9QYN8	0.95578	0.7497	Histamine H3 receptor	Hrh3	Rattus norvegicus (Rat).

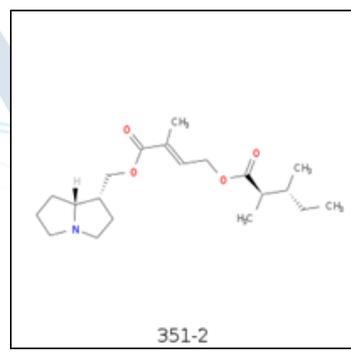
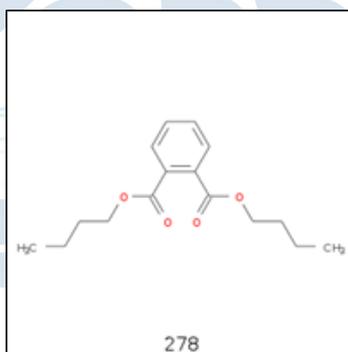
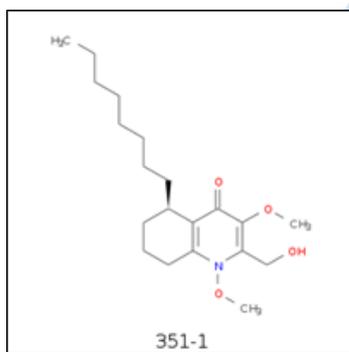


그림 20. 이차대사산물의 기능성 정보 및 표적 예측4

제4장 연구개발목표 달성도 및 대외기여도

제1절 연차별 목표달성도

총연구기간내 연차별 목표 대비 달성율(%)					
구분	연차별 달성내용				연차별 계획대비 연구실적 달성율(B) (%)
	세부연구목표	연구내용	가중 치 (A)	달성실적	
1년차 (2019)	1.난배양성 극지 균주 대량배양 기술 개발	1-1.후보 균주들의 배 양 최적화 테스트	25	- 활성효능을 지닌 5종 미세조류 배양 조건 확립 - 천연물확보를 위한 KSF0031균주 30g 이상 건조체 확보 - KNM0029C균주의 배지최적화 확립	100
		1-2. 10균주 바이오매 스 확보	25	- 극지미세조류 15종을 선별 및 2g 이상의 건조체 확보	100
	2.선별종의 전장 유전체 분석	2-1.ORF 또는 코딩 단백질 정보 확보	50	- KSF0031균주의 전장유전 체 분석 완료 - KSF0031균주의 이차대사 산물 생합성 유전체 분석 완 료 - AI기반 기능성 정보 및 표 적예측 완료	100
	계		1.0		100

제2절 연구결과

○ 본 과제는 극지 미세조류로부터 산업화 소재 발굴을 위해 KOPRI 연구자들이 극지에서 채집하여 유지해온 미세조류들을 활용하여 과학적, 산업적 가치를 발굴하는데 목적이 있다. 신규 물질 확보를 위하여 극한 환경에서 서식하는 미세조류의 바이오매스 대량 확보 가능성을 검토하고 유용물질 대량 생산가능성 및 화합물 정보, 기능성 정보를 확인하는데 연구 초점을 두었다. 선행 연구에서 확인된 항암, 항염, 항산화 효능이 있지만 배양이 어려운 균주들의 바이오매스를 확보를 위해 최적배양 방법을 확립하였다. 이를 통해 얻어진 바이오매스를 활용하여 추출물을 분석하였다. 화합물 분석을 위해 HPLC, LC-MS, NMR 등의 기기분석을 통하여 화합물 프로파일 정보를 획득하였다. 기능성 평가를 위해 항산화, 세포 독성, 면역 활성화 등의 활성을 세포주 기반으로 평가하였다. 화합물 프로파일 분석, 기능성 평가, 배양 최적화 결과, 균주의 신규성 등을 종합하여 본 연구에서 집중할 미세조류 1종 선발하였고, 전장 유전체 염기서열 분석으로 ORF 정보를 획득하였다. ORF 정보를 활용하여 *in silico* 화합물 DB로부터 AI 기반으로 코딩 단백질의 3차 구조 모델링을 통해 잠재적 기능을 예측하였고, 이는 향후 실용화를 위한 활성을 지닌 이차대산산물 생산 후보균주 및 단일 물질을 도출할 가능성을 제시하였다.

○ 주관기관 극지연구소

- 극지연구소에서 보유중인 미세조류 중 성장속도가 우수한 15종을 선별하여 2.5L 규모 종배양을 수행함
- 유용물질 확보를 위한 극지 미세조류의 최적 배양배지 제작 및 대량배양 (광생물반응기 활용, 총 건중량 30g 이상 확보) 프로토콜을 확립함
- 극지 미세조류 대량배양을 위해 최적온도, 다양한 농도의 이산화탄소 및 광량 조건을 시험 적용함으로써 200L 배양을 수행함
- 배양 완료후 원심분리와 동결건조 과정을 거쳐 저온에서 수분을 최대한 제거한 바이오매스 확보
- KIST에서 극지 미세조류 시료들로부터 추출물 제조, 분획, 유효성분의 분리, 화합물 프로파일 분석 및 기능성 평가를 통하여 1종 선별
- 극지연구소에서 선별 미세조류의 전장 유전체 염기서열 분석

○ 협동기관 한국과학기술연구원

- preparative MPLC, preparative HPLC 등 다수의 크로마토그래피 기법을 통한 물질 분리
- NMR 및 MS 등을 활용한 분리된 생리활성 화합물의 구조결정
- 추출물의 주요 작용기의 변형을 통한 활용 가능성 연구 및 대량 합성 추진
- 극지 미세조류 유래 화합물의 프로파일 분석 및 기능성 평가
- 상기 통합적 평가를 기반으로 선발한 미세조류의 전장 유전체 염기서열 분석을 통해 ORF 데이터를 획득하고 AI 기술을 활용하여 기능성 물질의 잠재적 기능을 예측하여 산업화 확장성 검토

- in silico 화합물을 Database로부터 예측하여 코딩 단백질의 아미노산 서열을 결정하고 단백질의 3차 구조 모델링을 예측함
- 단백질의 구조와 활성부위를 규명하여 기전연구를 수행, 향후 단백질 발현시스템을 이용하여 생산하고 산업화에 활용



제5장 연구개발결과의 활용계획

제1절 향후 연구방향 및 성과활용 계획

- 미세조류 연구가 다각화 되면서 생리활성물질, 의약품 원료물질 같은 고부가 가치를 창출하는 연구가 본격화되고 있다. 특히 극지 미세조류 연구는 새로운 개척분야로써 항암, 항세균, 항진균, 항바이러스, 신경활성과 같은 신규 생리활성물질들의 생산 가능할 것으로 예측된다. 최근 극지 미세조류 유래 천연물질의 주름개선, 항염, 보습 효과를 보고하였고 정상세포에 독성을 나타내지 않는 특성을 보여 세포보호 물질로써 사업화가 진행 중이다.
- 현 사업과제를 통해 신규 물질들의 활용성이 입증되고, 분자 기전 등이 분석된다면 건강식품, 의약품등의 시장 진입이 용이해진다. 극지의 청정 이미지가 합쳐져 현재 전세계 웰빙 트렌드에 맞는 고부가 상품으로 독보적인 분야 구축이 가능하다.
- 극지 미세조류 균주자원 및 배양 기술확보를 목표로 하며 향후 상업적 제품화 개발로 이어지도록 언론과 매체를 통한 홍보활동을 강화해야 한다.
- 극지 생물들이 극한의 환경에 적응하기 위해 생산하는 신규 이차대사산물들의 탐색과 기능개선, 분자생물학적 기초, 응용 연구를 통해 새로운 분자기작과 약리 기전을 규명하고 신기능 유용물질 개발의 방향 제시하여야 한다.
- 후보 물질의 용도에 부합하는 식품의약품안전처 허가를 받기 위한 일련의 연구(특히 확보)를 추진할 필요가 있다.
- 극지 미세조류 유래 유용 물질을 활용한 고부가 가치 제품 개발로 미래원천 기술을 확보함으로써 새로운 먹거리창출과, 일자리 증대로 인한 경제활성화, 연구 개발 투자 증대, 신물질 활용 추가 제품개발 등의 시장 선순환 고리를 형성할 수 있다.
- 극지 생물들이 가지는 특이 이차대사산물들의 개발과 분자, 생화학적 기초 연구를 통해 새로운 분자기작과 약리 기전을 규명함으로써 새로운 유용물질 개발에 방향 제시할 것이다.
- 극지 미세조류는 낮은 온도뿐만 아니라, 계절에 따른 광량 변화가 크고, 자외선이 강하며, 염도의 변화 폭이 큰 환경에 서식하기 때문에 이에 적응하기 위한 다양한 활성 물질 생산이 추정되어, 새로운 기능성 물질을 발견할 확률이 높을 것으로 기대된다.
- 미세조류 다양성에 비해 활용되고 있는 균주는 제한적인데 연구가 미비한 극지 미세조류의 다학제간 활용으로 새로운 바이오매스 경제체제로 패러다임을 바꾸는 돌파구가 될 것이다.

제6장 연구과정에서 수집한 해외과학기술정보

○ 해당사항 없음



제7장 참고문헌

1. Blin, K., Wolf, T., Chevrette, M.G., Lu, X., Schwalen, C.J., Kautsar, S.A., Suarez Duran, H.G., de Los Santos, E.L.C., Kim, H.U., Nave, M. *et al.* (2017) antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res*, 45, W36-W41.
2. Osbourn, A. (2010) Secondary metabolic gene clusters: evolutionary toolkits for chemical innovation. *Trends Genet*, 26, 449-457.
3. Kautsar, S.A., Suarez Duran, H.G., Blin, K., Osbourn, A. and Medema, M.H. (2017) plantiSMASH: automated identification, annotation and expression analysis of plant biosynthetic gene clusters. *Nucleic Acids Res*, 45, W55-W63.
4. Schurch, N.J., Schofield, P., Gierlinski, M., Cole, C., Sherstnev, A., Singh, V., Wrobel, N., Gharbi, K., Simpson, G.G., Owen-Hughes, T. *et al.* (2016) How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *Rna*, 22, 839-851.
5. Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G. *et al.* (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics*, 30, 1236-1240.