

극지생물 유전체 정보 생산 및 분석

de novo Genome analysis of Polar organisms



극지연구소

제 출 문

극지연구소장 귀하

본 보고서를 “포스트 극지유전체 프로젝트 : 극지 유용유전자 발굴을 위한 기능유전체 연구” 과제의 위탁연구 “극지생물 유전체 정보 생산 및 분석” 과제의 최종보고서로 제출합니다.



(본과제) 총괄연구책임자 : 김진형

위탁연구기관명 : 고려대학교

위탁연구책임자 : 박 현

위탁참여연구원 : 최은경

“ : 김진무

“ : 김장연

“ : 이승재

보고서초록

위탁연구과제명	극지생물 유전체 정보 생산 및 분석				
위탁연구책임자	박 현	해당단계 참여연구원수	5	해당단계 연구비	150,000,000원
연구기관명 및 소속부서명	고려대학교 생명공학부		참여기업명		
국제공동연구	상대국명 : 상대국연구기관명 :				
요약 (연구결과를 중심으로 개조식 500자 이내)				보고서 면수	17
<p>1. 극지 생물의 유전체 해독</p> <ul style="list-style-type: none"> ○ 남극 어류 scaly rockcod 유전체 서열 분석 <ul style="list-style-type: none"> - 근육 조직으로부터 DNA 추출 - 70Gb의 유전자 서열분석 수행 ○ 남극코끼리 해표 유전체 분석 <ul style="list-style-type: none"> - 혈액으로부터 DNA 추출 - 240Gb의 유전자 서열분석 수행 <p>2. 극지 생물의 유전체 표준지도 구축</p> <ul style="list-style-type: none"> ○ 남극 어류 scaly rockcod 유전체 조립 및 주석 작업 <ul style="list-style-type: none"> - 생물정보학기법을 이용한 유전체 조립과정을 통하여 23개의 chromosome 구축, - 주석작업을 통한 24,525개의 유전자 확인 ○ 남극 코끼리 해표 유전체 조립 및 유전자 주석 작업 <ul style="list-style-type: none"> - 생물정보학기법을 이용한 유전체 조립과정을 통하여 18개의 chromosome 구축, - 주석작업을 통한 20,457개의 유전자 확인 - 남극 코끼리 해표와 California sea lion의 크로모솜 비교 수행 - 남극 코끼리 해표와 다른 해양동물 포유류와의 상동 유전자 비교 수행 					
색 인 어 (각 5개 이상)	한 글	남극어류, 남극코끼리해표, 유전체, 유전자 서열분석, 크로모솜, 주석			
	영 어	fish, southern elephant seal, Antarctic area, genome, sequencing, chromosome, annotation			

요 약 문

I. 제목 : 극지생물 유전체 정보 생산 및 분석

II. 연구개발의 목적 및 필요성

- 극지 생물은 극지의 환경에 적응하기 위해 다양한 형태로 진화
- 이는 형태 형성에 관여하는 유전자들의 기능과 연관 있음
- 이 유전자들의 구조 및 기능 진화에 대한 분석을 통해 생물 진화의 연구수행
- 매우 낮은 온도에서도 생존을 이어나가는 원인을 찾아서 이를 다양한 상업적 가치가 있는 자원을 개발할 수 있음

III. 연구개발의 내용 및 범위

1. 극지생물의 선도 유전체 해독

- 남극어류 (scaly rockcod)의 유전체 서열분석
- 남극 코끼리 해표의 유전체 서열분석

2. 극지생물의 유전체 표준지도 구축

- 남극어류 (scaly rockcod)의 유전체 조립 및 주석 작업
- 남극 코끼리 해표의 유전체 조립 및 주석 작업

IV. 연구개발결과

1. 극지생물의 선도 유전체 해독

- 남극어류 (scaly rockcod) : 70Gb의 유전체 서열분석
- 남극 코끼리 해표 : 240Gb의 유전체 서열분석

2. 극지생물의 유전체 표준지도 구축

- 남극어류 (scaly rockcod) : 23개의 크로모솜 구축 및 24,524개의 유전자확인
- 남극 코끼리 해표의 : 18개의 크로모솜 구축 및 20,457개의 유전자확인

V. 연구개발결과의 활용계획

- 극지 생명과학 연구의 근간인 유전체 정보를 확보함으로써, 극지생물의 적응과 진화연구의 토대 마련
- 극지생물로부터 유용유전자원 확보를 위한 유전체 정보 기본정보 제공
- 극지 적응 핵심 유전자의 수생물 및 작물 내 형질전환 도입 기술 확보를 위한 유전자 정보 제공
- 진화의 과학적 근거 제시를 위한 정보 제공

S U M M A R Y

(영 문 요 약 문)

I . Title : *de novo* Genome analysis of Polar organisms

II . Purpose and Necessity of R&D

- polar organisms evolve into various forms to adapt to the polar environment
- this is related to the function of genes involved in morphogenesis
- research on the evolution of organisms through the analysis of the structure and functional evolution of these genes
- it is possible to develop a variety of commercially valuable resources by finding the causes that contribute to survive even at the very low temperature

III . Contents and Extent of R&D

1. *de novo* genome sequencing

- scaly rockcod
- southern elephant seal

2. genome assembly with chromosome level

- scaly rockcod
- southern elephant seal

IV . R&D Results

1. *de novo* genome sequencing

- scaly rockcod : 70Gb
- southern elephant seal : 240Gb

2. genome assembly with chromosome level

- scaly rockcod : 23 chromosomes and 24,524 genes

V . Application Plans of R&D Results

- by securing genomic information, it lays the foundation for adaptation and evolution research of polar organisms
- provides basic genome information for securing useful genetic resources from polar organisms
- provides gene information for securing the technology for introducing transformation into aquatic organisms and crops of core genes for polar adaptation
- provision of information to present the scientific basis for evolution

목 차

제 1 장. 연구 개발의 목적 및 필요성

제 2 장. 국내외 기술개발 현황

제 3 장. 연구개발수행 내용 및 결과

1. 선도유전체 해독 연구 과정

2. 선도유전체 해독 연구 결과

가. 극지생물의 유전체 해독

(1) 남극어류 (scaly rockcod)의 유전체 해독

(2) 남극 코끼리 해표 유전체 해독

나. 극지생물의 유전체 표준지도 구축

(1) 남극어류 (scaly rockcod)의 유전체 조립

(2) 남극 코끼리 해표의 유전체 조립

제 4 장. 연구개발목표 달성도 및 대외기여도

1. 연구개발 최종목표

2. 연구개발 달성도

3. 경제적 파급효과

제 5 장. 연구개발결과의 활용계획

제 6 장. 참고문헌

제 1 장. 연구 개발의 목적 및 필요성

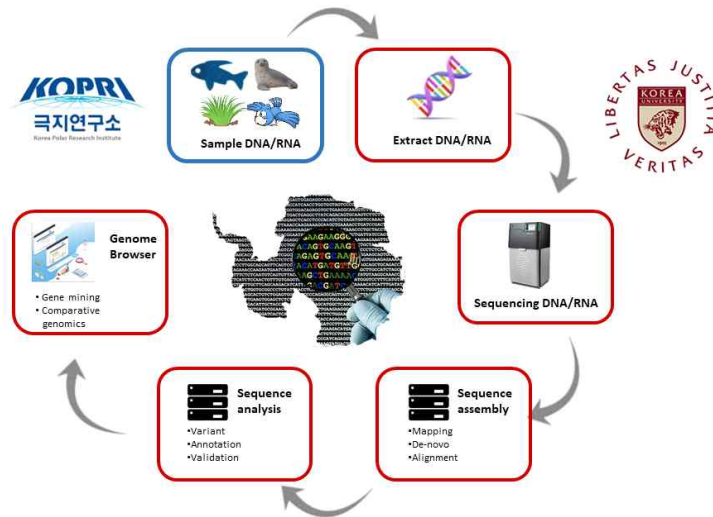
- 서로 유기적으로 연결되어 상호작용하며 환경에 적응하는 일반적인 생물들이 서식하는 온대, 열대 등의 지역과 달리 극지방은 서식 환경이 매우 추운 지역이며, 다른 지역들과 유기적으로 연결되어 있지 않은 일종의 고립된 지역임
- 이렇게 외부와 차단 된 지역에서 기온 및 수온이 타 지역과 달리 매우 낮은 극한의 환경은 극지 환경에 적응 하기 위한 진화가 필수적임
- 극지 생물은 극지의 환경에 적응하기 위해 다양한 형태로 진화하였으며, 이는 형태형성에 관여하는 유전자들의 기능에 의해 조절됨. 이들 유전자들의 구조진화 및 기능 진화에 대한 분석을 통해 극지 생물 진화의 연결고리를 탐구하는 것은 학문적으로 매우 중요하고, 흥미로운 연구 주제 임
- 진화의 고리를 파악함과 동시에 매우 낮은 온도에서도 생존을 이어나갈 수 있는 원인을 찾을 수 있다면 이를 응용하여 다양한 산업적 가치가 있는 자원을 개발할 수 있음
- 특히, 극지생물 특이적 유전자, 단백질, 대사산물 등은 수산업, 축산업, 농업, 원예기술 혁신 뿐만 아니라 미용, 의료 등 다양한 산업 분야로의 활용이 가능한 미래 자원으로써 그 활용 가치가 매우 큼
- 또한, 특정 생물의 게놈 정보를 발표 후 꾸준히 그 정보를 활용하며 비교분석을 통한 적응과 진화 양상 예측 등의 기능유전체를 바탕으로 한 유용유전자의 발굴은 다양한 분야에의 활용의 파급효과가 매우 높음
- 이와 같은 연구를 위해서는 극지 생물의 유전체 정보의 해독을 통한 확보는 무엇보다도 선행되어야 하며, 특히 고품질의 유전체 정보는 필수적인 요인임.
- 본 연구에서 제시된 극지 생물의 유전체 정보는 극지연구소의 본과제의 극지 유용유전자 발굴을 위한 기능유전체 연구의 기반이 되는 정보를 제공함

제 2 장. 국내외 기술개발 현황

- 샘플링 방법, 시퀀싱 및 생물 정보학의 최근 획기적인 발전을 고려할 때 극지방 해양 환경이 최근 과학적 주목을 받고 있음
- 지구 기후 변화와 관련하여 극지방은 가장 먼저 영향을 받는 고립된 지역이므로, 이 지역의 생물에 대한 다양한 연구 변화가 많은 나라에서 활발하게 진행되고 있음
- 극지방 생물은 '차가운 적응'이라고 정의된 생물로서, 초 냉각 액체 물의 영하의 온도에서 영구적으로 추운 환경에서 번성하며 이러한 극한 조건에서 생존하고 번식할 수 있는 생리학적 및 생화학 적 능력을 진화시킨 생물입니다.
- 극지방의 해양 생물에 대한 더 나은 지식을 바탕으로 새로운 생물 활성 대사 산물과 잠재적인 상업적 응용이 가능한 단백질 / 효소 측면에서 이전에는 접근할 수 없었던 생물 산물이 밝혀지고 있음
- 해양 생물 공학의 세계 시장은 2015년에 41억 달러로 추산되며 2020년에는 48억 달러, 2025년에는 64억 달러에 이를 것으로 예상 됨
- 또한, 최근 유전체 연구 방법의 발달로 국제적인 요구 수준이 염색체 수준의 어셈블리로 높아짐에 따라, 논문 게재가 점점 어려워지고 있음
- 유전체 어셈블리의 국제적 요구가 염색체 수준의 어셈블리로 상향되면서, 최상위등급 논문의 게재가 어려워지고 있음

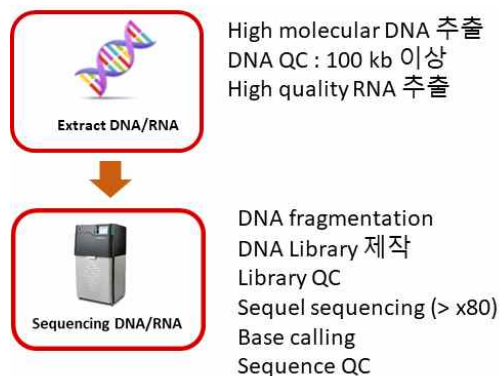
제 3 장. 연구개발수행 내용 및 결과

1. 선도유전체 해독 연구 과정



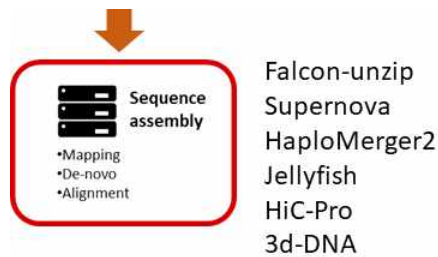
1) 유전체 분석

선도게놈(*de novo* assembly)은 유전체 지도가 없는 생물 중에서 유전체 초안 지도를 작성하기 위한 assembly 방법임. 2000년대 초반 차세대해독기술(next generation sequencing, NGS) 기술로 해독(sequencing) 가격이 낮아지면서 짧은 mate-pair read와 paired-end read를 연결하는 *de novo* assembly 방법이 개발 되었음. 하지만, 짧은 서열의 한계 때문에 유전체의 많은 영역이 un-assembled 된 상태로 남겨지고, 유전자를 coding 하는 영역도 partial로 assemble 되는 한계를 보임 이러한 한계를 극복하기 위하여, long read를 이용하여 genome을 분석하는 방법이 개발 되었으며, 이전보다 quality 높은 scaffold를 만들어 내고 있음. 따라서 고품질의 유전체 해독을 위하여 long-reads의 DNA 서열 분석을 통하여 유전체의 gap이 존재하지않는 contig 조립



2) 유전체 크로모솜 조립

종, 식물과 같은 고등생물의 수개의 크로모솜으로 유전체가 구성되어있음. 생물의 적응과 진화 메카니즘을 규명하기 위해서는 유전자의 배열과 조합의 gene syteny 분석을 통하여서 시스템적인 이해가 가능함. 따라서 고품질로 조립된 유전체 contig를 유전자좌가/컨티그간 근접도를 Hi-C 분석을 통한 구조적인 정보를 이용하여 염색체 수준으로 조립

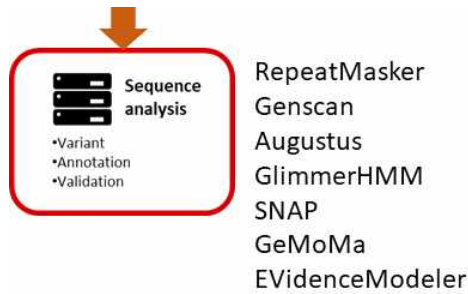


3) 전사체(Transcriptome) 분석

전사체(Transcriptome) 분석은 세포나 조직에서 발현되는 RNA로부터 총 전사물(transcript)을 분석하여, 생물체의 모든 형질 특성, 성장 시기, 환경적 자극에 따른 유전자 발현의 차이를 분석하는 omics 기법임. 이 분석은 표현 형질의 특성에 따른 유전자의 발현 차이나 조절기전을 이해하기 위해 이용 됨. Transcriptome 분석은 샘플의 total RNA를 추출 후, mRNA만 정제하여 Pacbio의 Isoseq 분석을 통하여 full length cDNA의 정보를 해독

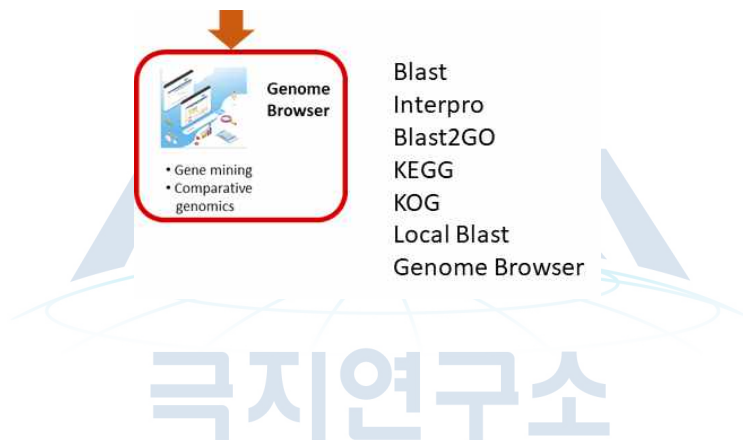
4) 게놈의 유전자 annotation

조립된 게놈과 isoseq으로부터 얻는 전사체 정보를 통하여 게놈내의 유전자 정보를 찾는 과정을 annotation이라함. annotation은 게놈서열로부터 de novo 예측을 통하여 우선적으로 유전자 부위를 예측한후 동일샘플의 전사체 정보와 homology 분석, 유사종의 단백질 서열정보와 homology 분석을 통하여 유전자 정보를 해독을 진행



5) 유전자 기능분석

게놈으로부터 확인된 유전자의 DNA서열과 protein서열을 통하여 ncbi의 nr database에 blast를 통하여 유전자를 확인함. 확인된 유전자의 gene ontology 분석, KEGG 분석, domain 분석을 통하여 유전자 기능을 유추함



2. 선도 유전체 해독 연구결과

가. 극지어류의 유전체 해독

(1) 남극 어류 (scaly rockcod)의 유전체 해독

- 남극 어류(scaly rockcod)의 근육조직을 유전체가 보존된 상태를 유지하며 채취 함
- 채취한 근육 조직으로부터 NGS 실험이 가능한 수준의 다량의 genomic DNA 추출 함
- Phenol을 이용하여 고순도의 genomic DNA를 얻기 위한 정제 과정을 거친 후 Nanodrop 및 Qubit으로 농도 및 순도를 측정 함
- Tapestation을 이용한 genomic DNA QC를 통해 고순도 gDNA 확인 함 (그림 1)

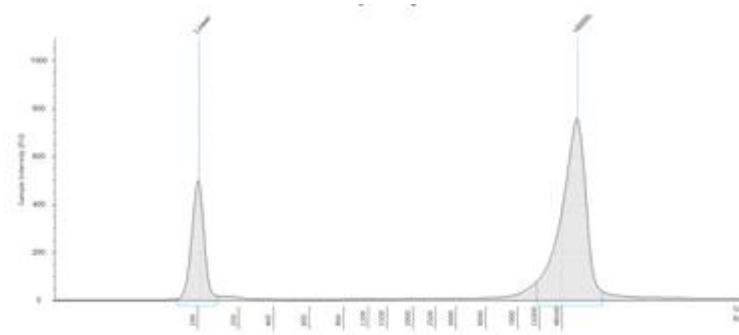


그림 1. 추출한 남극 어류(scaly rockcod)의 Genomic DNA 품질

- 확보한 고순도의 genomic DNA를 이용하여 유전자 서열 분석을 시행 함
- 게놈 내의 유전자 정보를 찾기 위한 주석 과정 (annotation)을 위해 total RNA를 추출
- 추출한 total RNA를 이용하여 RNAseq을 진행 함
- 확보된 scaly rockcod gDNA를 Pacbio sequel을 이용하여 남극어류 유전체의 70X 수준에 해당하는 70G의 Raw Data를 생산 함 (표 1)

표 1. 생산된 scaly rockcod의 DNA Raw data

Library type	Number of cells	Number of reads	Total read bases (bp)
PacBio	4	6,258,640	77,938,427,833

(2). 남극 코끼리 해표의 유전체 해독

- 남극 코끼리 해표의 혈액으로부터 DNA를 추출하여 10× Genomics Chromium technology 를 이용하여 유전체 서열 분석 수행 (표 5)

표 5. 생산된 DNA Raw data

	Number of Reads	Total Read Bases
Genome	1,606,582,076	242,593,893,476
Transcriptome	80,958,400	19,430,016,000

나. 극지어류의 유전체 표준지도 구축

(1) 남극어류(scaly rockcod)의 유전체 조립

- 생산된 DNA 서열 데이터를 이용하여 cluster system을 이용하여 falcon-unzip assembler로 유전체 조립을 수행 함 (표 2)

표 2. 남극 어류(scaly rockcod) 유전체 조립 결과

Assembly	Contigs
Number	1,482
Total size of assembly (bp)	944,447,341
Longest contig (bp)	17,998,618
N50 contigs length (bp)	1,726,674
Number of scaffolds > 9 Mb	4

- 조립된 유전체의 완성도는 BUSCO 분석을 통하여 core 유전자의 coverage 로서 확인함 (표 3)

표 3. 유전체 조립완성도 평가를 위한 Benchmarking Universal Single-Copy Orthologs (BUSCO) 분석

Actinopterygii_odb10	Number	Percentage (%)
Complete BUSCOs (C)	3,291	90.4
Complete and single-copy BUSCOs (S)	3,222	88.5
Complete and duplicated BUSCOs (D)	69	1.9
Fragmented BUSCOs (F)	96	2.6
Missing BUSCOs (M)	253	7.0
Total BUSCO groups searched	3,640	

- 조립된 유전체와 RNA isoseq 서열, 유사 어류종의 단백질 서열 등을 이용하여 유전자 annotation 분석 수행 (표 4)

표 4. 남극어류(scaly rockcod)의 유전자 확인 결과

Annotation database	Annotated number	Percentage (%)
No. Genes	24,525	
nr Annotation	23,016	93.8

(2) 남극 코끼리 해표의 유전체 조립

- 남극 코끼리 해표의 DNA 서열을 기반으로 Supernova assembler를 이용하여 2.4 Gb의 유전체 서열 조립 (표 6)

표 6. 남극 코끼리 해표 유전체 조립 결과

Assembly	Supernova 2.0
Number of scaffolds	1115
Total size of scaffolds	2,417,339,903
Longest scaffolds	111,625,095
N50 scaffolds length	54,232,831
Number of scaffolds >10M	54
Gap (%)	0.65

- 유전체 서열 기반으로 Transposable elements 분석 수행 (표 7)

표 7. 남극 코끼리 해표의 Transposable elements 분석 결과

	Number of Elements	Length (bp)	Percentage of Sequence (%)
SINEs:	888,301	148,028,385	6.06
MIRs	254,250	35,222,085	1.44
LINEs:	1,624,363	540,626,278	22.12
LINE1	1,473,069	508,962,965	20.83
LINE2	150,067	31,135,376	1.27
L3/CR1	634	81,165	0.00
LTR elements:	268,713	74,825,384	3.06
ERV1	81,307	29,432,080	1.20
ERV1-MaLRs	134,246	29,421,479	1.20
ERV_classI	52,451	15,768,764	0.65
ERV_classII	370	158,839	0.01
DNA elements:	261,769	44,140,597	1.81
hAT-Charlie	146,299	24,083,632	0.99
TcMar-Tigger	43,576	10,220,361	0.42
Unclassified:	20,630	5,747,808	0.24
Total interspersed repeats:		813,368,452	33.28
Small RNA:	637,161	113,111,897	4.63
Satellites:	1867	85,913	0.00
Simple repeats:	618,437	26,612,557	1.09
Low complexity:	99,311	5,218,008	0.21

- 조립된 유전체를 이용하여 maker pipeline을 통한 유전자 annotation 분석 수행 (표 8)

표 8. 남극코끼리 해표의 유전자 확인 결과와 기능분석 결과

Annotation Database	Annotated Number	Percentage (%)
No. of genes	20,457	
nr annotation	19,901	97.3
GO annotation	3109	15.2
KEGG annotation	11,501	56.2
KOG annotation	15,498	75.8
Pfam annotation	16,733	81.8
Swissprot annotation	19,653	96.1
TrEMBL annotation	17,600	86.0
	Count	Length Sum (bp)
Exon	168,375	30,497,777
CDS	167,919	28,718,340

- 남극 코끼리 해표와 California sea lion 의 크로모솨 비교 수형 (그림 3)

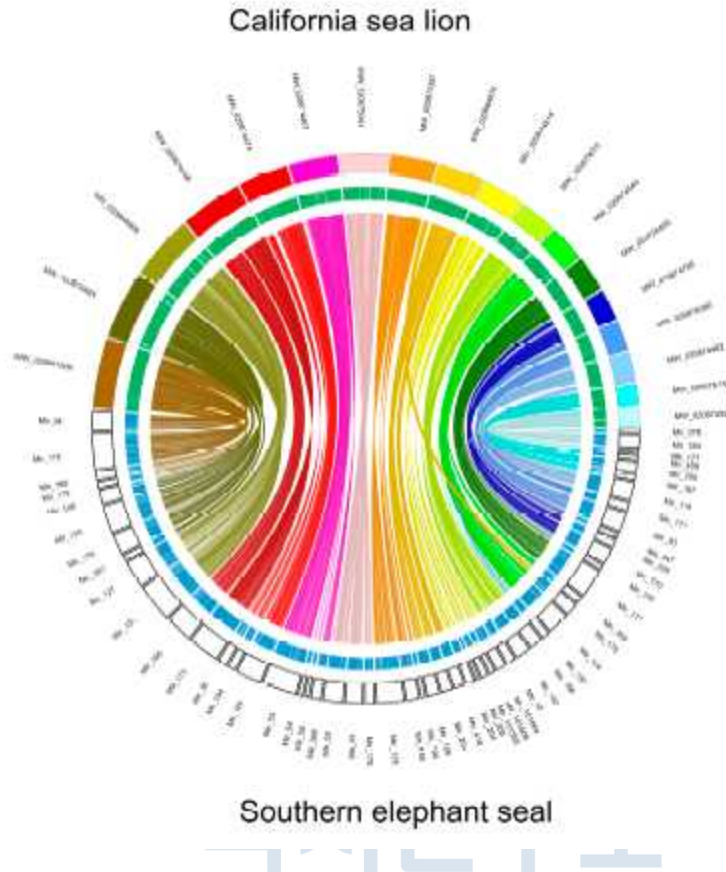


그림 3. 남극 코끼리 해표와 California sea lion의 크로모솨 비교 결과

- 다른 해양포유류와 남극 코끼리 해표와의 상동유전자 분석 수행 (그림 4)

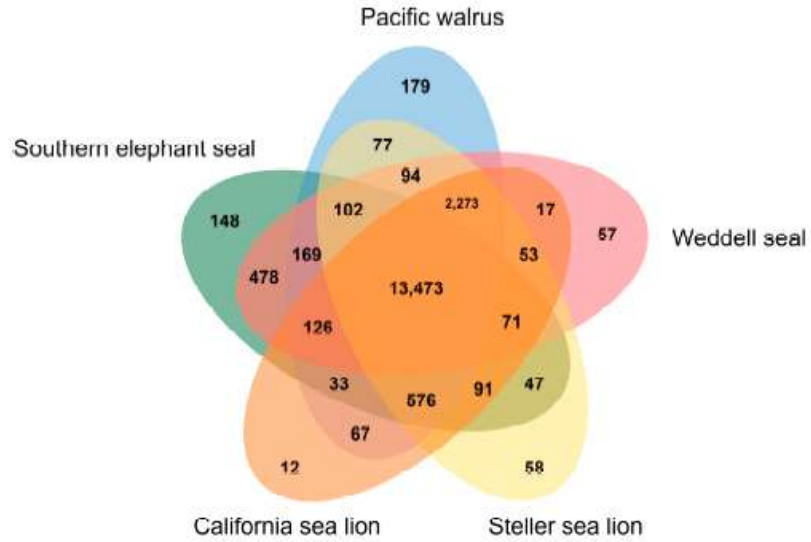
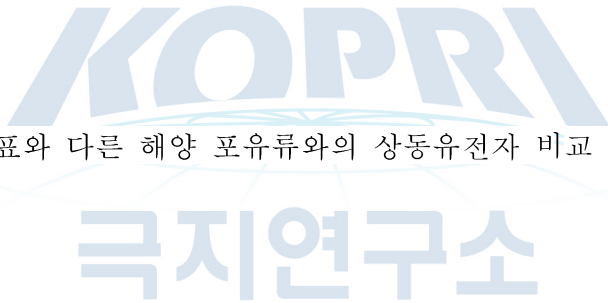


그림 4. 남극 코끼리 해표와 다른 해양 포유류와의 상동유전자 비교 결과



제 4장. 연구개발목표 달성도 및 대외기여도

1. 연구개발 최종 목표

가. 극지생물 선도 유전체 해독

- 남극 어류 (scaly rockcod)의 유전체 해독 및 유전체 분석
- 남극 코끼리 해표의 유전체 해독 및 유전체 분석

2. 연구개발 달성도

성과목표	세부목표		달성 주요내용	달성도(%)
극지 어류의 선도유전체 해독	1-1	극지어류의 유전체 해독	<ul style="list-style-type: none"> - 남극 어류 scaly rockcod 유전체 서열 분석 (70 Gb의 유전자 서열분석 수행) - 남극코끼리 해표 유전체 분석 	100
	1-2	극지어류의 유전체 표준지도 구축	<ul style="list-style-type: none"> - 남극 어류 scaly rockcod 유전체 조립 및 주석 (생물정보학기법을 이용한 유전체 조립과정을 통하여 23개의 chromosome 구축, 주석작업을 통한 24,525개의 유전자 확인) - 남극 코끼리 해표 유전체 조립 및 유전자 주석 작업 	100

3. 경제적 파급효과

- 극지 생물 신소재 개발을 위한 기초연구로 21세기 세계 생명공학 분야에서 선도적인 역할 담당
- 세계 바이오산업 시장 규모는 약 330조 원 (2013년 기준)로 연평균 10.8%의 고성장세를 보이고 있으며, 지속적으로 증가될 전망. 이에 미래 글로벌 시장의 선점을 위해서는 극지생물을 이용한 다양한 생명공학기술의 연구개발이 필요하며, 미래경제력 창조의 기초 제공
- 선도 유전체 해독과 기증 분석을 위한 BT, IT 등 융합학문을 통하여 4차 산업 촉진하므로 국가 경쟁력 제고

제 5 장 연구개발결과의 활용계획

- 극지 생명과학 연구의 근간인 유전체 정보를 확보함으로써 극지생물의 적응과 진화 연구의 토대 마련
- 극지 생물로부터 유용유전자원확보를 위한 유전체 정보 기본정보제공
- 극지 적응 핵심 유전자의 수생물 및 작물 내 형질전환 도입 기술 확보를 위한 유전자 정보 제공
- 극지 특이적인 생명현상 규명을 통한 지구 생명의 생성과 적응, 진화의 과학적 근거 제시를 위한 생명정보 제공



제 6 장 참고문헌

1. Jinmu Kim, Sung-Min Jang, Eunkyung Choi, Euna Jo, Seung Jae Lee, Sun Hee Kim, Young Min Chi, Jin-Hyoung Kim, Hyun Park. 2020. The complete mitochondrial genome of the Eaton's skate, *Bathyraja eatonii* (Rajiformes, Arhynchobatidae). *Mitochondrial DNA Part B: Resources*. in press
2. Eunkyung Choi, Seong Hee Park, Seung Jae Lee, Euna Jo, Jinmu Kim, Jeong-Hoon Kim, Jin-Hyoung Kim, Jong Seok Lim, Hyun Park. 2020. The complete mitochondrial genome of Patagonian moray cod, *Muraenolepis orangiensis* Vaillant, 1888 (Gadiformes, Muraenolepididae) *Mitochondrial DNA Part B: Resources*. 5(3), 2707-2708.
3. Seung Jae Lee, Tae-Eul Im, Euna Jo, Eunkyung Choi, Young Min Chi, Jin-Hyoung Kim, Jeong-Hoon Kim Hyun Park 2020. The complete mitochondrial genome of *Macrourus whitsoni* (Gadiformes, Macrouridae). *Mitochondrial DNA Part B: Resources*. 5(3). 2326-2327.
4. Euna Jo, Yll Hwan Cho, Seung Jae Lee, Eunkyung Choi, Jeong-Hoon Kim, Young Min Chi, Jin-Hyoung Kim, Hyun Park. 2020. The complete mitochondrial genome of the Antarctic marbled rockcod, *Notothenia rossii* (Perciformes, Nototheniidae). *Mitochondrial DNA Part B: Resources*. 5(3). 2421-2422.
5. Bo-Mi Kim, Yoon Jin Lee, Jeong-Hoon Kim, Jin-Hyoung Kim, Seunghyun Kang, Euna Jo, Seung Jae Lee, Jun Hyuck Lee, Young Min Chi, Hyun Park. 2020. The genome assembly and annotation of the southern elephant seal *Mirounga leonina*. *Genes*. 11. 160.
6. Bruno, S., Coppola, D., di Prisco, G., Giordano, D., & Verde, C. (2019). Enzymes from marine polar regions and their biotechnological applications. *Marine drugs*, 17(10), 544.

주 의

1. 이 보고서는 극지연구소 위탁과제 연구결과보고서입니다.
2. 이 보고서 내용을 발표할 때에는 반드시 극지연구소에서 위탁연구과제로 수행한 연구결과임을 밝혀야 합니다.
3. 국가과학기술 기밀유지에 필요한 내용은 대외적으로 발표 또는 공개하여서는 안됩니다.