

극지어류 유전자군의 독립유전정보 분석을 통한
진화기작 해석

Understanding of the adaptation mechanism of
Antarctic fish to extreme environment through
analysis of comparative genome evolution



인천대학교 산학협력단

제 출 문

극지연구소장 귀하

본 보고서를 “극지유전체 101 프로젝트: 극지생물 유전체 정보 분석 및 활용 기반 구축” 과제의 위탁연구 “극지어류 유전자군의 독립유전정보 분석을 통한 진화 기작 해석” 과제의 최종보고서로 제출합니다.



총괄연구책임자 : 김 진 형
위탁연구기관명 : 인천대학교 산학협력단
위탁연구책임자 : 이 재 성
위탁참여연구원 : 엄 혜 진
“ : 남 상 은

보고서 초록

위탁연구과제명	극지어류 유전자군의 독립유전정보 분석을 통한 진화기작 해석				
위탁연구책임자	이 재 성	해당단계 참여연구원수	2	해당단계 연구비	100,000,000
연구기관명 및 소속부서명	인천대학교 해양학과		참여기업명	해당없음	
국제공동연구	해당없음				
요약				보고서면수	37
<p>기존의 남극 적응성에 관련된 어류유전체 진화 이론에 대한 재검증과 신이론 발굴을 목표로 현재까지 남극어류의 극한 환경에 대한 유전체 적응성 이론에 널리 활용되는 예들인 antifreezing glycoprotein의 게놈 내 분화 및 반복, lack of hemoglobin expression and oxygen-binding protein myoglobin과 small fragment of γ-globin gene family, 그리고 mitochondrial metabolic change와 ATP 생합성간의 관계 등에 주목하여 연구를 진행함. 상기 예들은 다양한 남극어류에서 꾸준히 제시되고 있으나, 게놈 정보 기반에 의한 분석이 아닌 특정 유전자에 대한 target-specific approach를 이용하여 밝혀진 정보들이기에, 확보된 남극어류 게놈 정보를 이용하여 이들 적응요소들이 게놈 내 insertion, deletion, 또는 expansion 및 contraction 여부를 정확히 분석해냄으로 이론을 재점검하며 특이점을 분석해냄으로 교과서와 같은 이론 정립 체계를 구축하였음. 또한 남극어류 각 종에 대한 계통학적 진화양상을 추적하였음. 남극어류는 극한 환경내에서도 다양한 종이 서식하고 있기 때문에 상기 게놈 적응 요인 뿐 아니라 일반적으로 어류 게놈에서 다루어지는 게놈 진화 타겟 gene family 비교 분석을 통해 남극어류 각 종이 지니고 있는 적응 요인이 종(種; species) 또는 속(屬; genus) 특이성인지, 보다 높은 상위단계인 과(科; family) 또는 목(目; order) 단위에서 분지되어 유지된 것이거나 새롭게 획득한 것임을 분석하여 남극어류에 대한 진화양상을 분석하였음. 어류의 게놈진화에 중요한 hox gene family와 극저온 저항과 적응에 관여하는 antifreezing protein, transient receptor potential(TRP), olfactory receptor superfamily, heat shock protein(Hsp) superfamily, zona pellucida(ZP) gene family, opsin gene family, hemoglobin coding gene cluster 발굴 및 비교분석을 수행하였음. 특히 극한 온도에 대한 분자적응능력 중 중요한 요소인 체 내 얼음 생성방지 및 극저온 조절에 대한 연구는 현재까지 주로 미생물을 대상으로 연구가 이루어져 왔음. 상대적으로 작은 게놈 크기와 적은 수의 유전자를 지닌 미생물을 대상으로 상기 유전요인들을 조사하는 것은 그리 어렵지 않지만, 게놈 크기가 크며 유전자의 개수와 isoform과 같은 gene variant가 높은 어류에서는 이러한 연구가 거의 이루어지지 않음. 이러한 니체를 활용하여 막 구성에 주요한 역할을 하는 gene family와 지질 대사에 관여하는 대규모 gene family를 분석하였음</p>					
색 인 어	한 글	남극어류, 게놈, 적응, 유전자군, 유전자진화			
	영 어	Antartic fish, genome, environmental adaptation, gene family, gene evolution			

요 약 문

I. 제 목

극지어류 유전자군의 독립유전정보 분석을 통한 진화기작 해석

II. 연구개발의 목적 및 필요성

- 기 분석된 극지어류 유전 진화 재검증을 통한 분자적응 기작 확립 필요
- 극지어류 유전체 활용 극대화를 통한 선진 유전체 분석 플랫폼 구축
- 극지어류 유전체 활용을 통한 세계적인 연구분석능력 배양

III. 연구개발의 내용 및 범위

- 극지어류 계놈 내 독립적 유전자 진화 분석을 통한 적응 방식의 이해
- 남극 어류 계놈 내 주요 유전자군 정보 확립
- 극지어류와 아열대 서식 어류 계놈 상호 비교 분석
- 남극 어류 계놈 내 주요 유전자군 정보 확립

IV. 연구개발결과

- 경골어류 계놈 진화 대비 남극어류 계놈 진화에 관여하는 주요 유전자군 분석 완료
- 남극어류 내 특이 유전자군 재검증 및 분류 완료
- 남극 어류 계놈 내 특정 유전자 및 유전자군의 변화가 환경적응성에 미치는 영향 분석 완료
- 유전자군 기능예측을 통한 환경적응 능력과의 상관관계 분석 완료

V. 연구개발결과의 활용계획

- 극지어류 유전체 정보를 통해 확보한 분자진화 기작을 이용하여 학술적으로 극대화 추진 및 최정상급 논문 발표
- 어류 비교 계놈 전문 분석 연구진 양성
- 나고야의정서 대비 극지어류 계놈 정보 선재 확보
- 남극 어류 계놈 내 특이 단백질 생성 유전자 발굴 및 상업화 추진

S U M M A R Y

I. Title

Understanding of the adaptation mechanism of Antarctic fish to extreme environment through analysis of comparative genome evolution

II. Purpose and Necessity of R&D

- Revisiting on the analyzed genome resource for establishment of molecular evolution and adaptation of Antarctic fish
- Development of genome analysis platform in Antarctic fish genomes
- Training genome analyzing technique for global of human resources

III. Contents and Extent of R&D

- Understanding of Antarctic fish-specific molecular evolution and adaptation mechanisms
- Establishment of Antarctic fish genome database
- Comparative analysis of genome evolution between Antarctic and subtropic fish

IV. R&D Results

- Completion of analysis of Antarctic fish-specific genome evolution
- Establishment of Antarctic fish genome database
- Comparative analysis of genome evolution between Antarctic and subtropic fish

V. Application Plans of R&D Results

- Publication of high quality manuscript on top-tier journal
- Training high quality human resource for genomic analysis
- Commercialization of useful Antarctic fish-specific proteins

목 차

제 1 장 서론

제 2 장 국내외 기술개발 현황

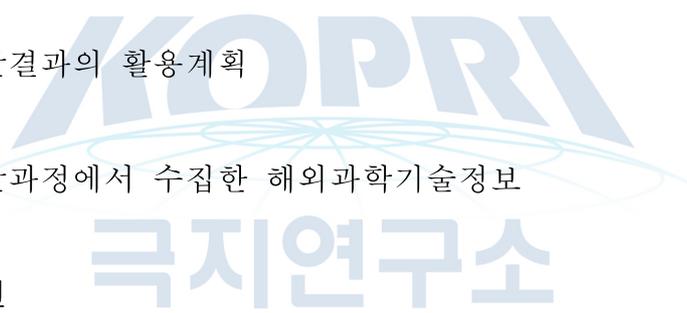
제 3 장 연구개발수행 내용 및 결과

제 4장 연구개발목표 달성도 및 대외기여도

제 5 장 연구개발결과의 활용계획

제 6 장 연구개발과정에서 수집한 해외과학기술정보

제 7 장 참고문헌



제 1 장 서론

제 1 절 연구개발의 목적

가. 연구개발 최종 목표

극저어류 유전자군의 독립유전정보 분석을 통한 진화기작 해석

나. 연구개발 세부 목표

성과목표	세부목표		연구수행내역
전체 어류 계통 내 주요 유전자군 비교분석을 통한 남극어류 특이적 계통 진화 분석	1-1	남극어류 계통 내 경골어류 진화를 이끌어온 주요 유전자군 정보 확보	<ul style="list-style-type: none"> - 경골어류 계통 진화에 관여하는 주요 유전자군 정보 분석 - 남극어류 계통 정보와의 비교 분석
	1-2	남극 어류 계통 내 주요 유전자군 정보 확립	<ul style="list-style-type: none"> - 상기 남극어류 내 유전자군 분류 - 대상유전자군의 진화속도 비교분석
	1-3	남극어류 유전자군의 어류 내 계통학적 보존유전 또는 독립유전정보 분석	<ul style="list-style-type: none"> - 남극어류 유전자군의 계통학적 비교 분석 - 주요 유전자군의 보존 또는 독립진화여부 분석
남극 환경 적응에 있어 남극어류 특이적 계통 진화의 역할 규명	2-1	남극어류 계통 특이적 진화양상 제시	<ul style="list-style-type: none"> - 주요 유전자군의 계통 내 synteny 비교 분석 - 남극 어류 특이적 유전자군의 insertion 및 deletion 여부 분석
	2-2	남극어류 특이 계통 진화가 환경적응에 미치는 영향 분석	<ul style="list-style-type: none"> - 계통 내 특정 유전자 또는 유전자군의 변화가 환경적응성에 미치는 영향 예측 - 유전자군 기능예측을 통한 환경적응 능력과의 상관관계 분석
	2-3	남극어류 유전자군의 어류 내 계통학적 보존유전 또는 독립유전정보 분석	<ul style="list-style-type: none"> - 현재까지 계통 정보가 밝혀진 어류들을 대상으로 남극어류 특이 유전자군 진화 분석 - 남극어류 특이 유전자군의 진화속도 분석

제 2 절 연구개발의 필요성 및 범위

- 본 연구에서는 게놈 정보가 발표된 남극어류들(남극대구, 아이스피쉬 및 드래곤피쉬)의 게놈 데이터베이스 내에서 극한환경적응과 연관된 게놈과 유전자 진화 요소 및 패턴을 찾는 것이 주된 목적임. 특히 기존에 발표된 분자 기반 적응 메커니즘은 전체 게놈 정보의 부재로 인해 단편적인 예상 메커니즘들이기에, 본 연구를 통해 기발표된 게놈 정보의 후속연구에의 활용을 극대화하여 독보적인 남극어류 유전체 기능분석에 있어 중요한 국제적 입지를 세우며, 산업화 소재로 활용될 수 있는 기능 유전 물질 발굴, 그리고 추가로 발굴되는 남극어류의 정보를 안정적으로 활용할 수 있는 플랫폼과 연구인력 양성이 부수적인 목적임
- 기존에 발표된 게놈 정보들을 최신 assembly methodology 적용을 통해 기발표된 정보의 확장을 극대화하며, genome resequencing이나 BAC library sequencing 등의 절차를 생략함으로써 유전체 확보에 필요한 예산을 절감할 수 있고, 역으로 선행발표와 후속연구의 연속성이 유지될 수 있는 안정적인 연구 플랫폼을 구축할 수 있음. 특히 극지연구소와 인천대학교는 지리적으로도 매우 가깝기 때문에 유전체 관련 연구진의 수시 토의 및 디스커션 공유가 쉬우며, 연구 자원의 공유 역시 쉽게 이루어져 큰 시너지 효과를 얻을 수 있을 것이라 사료됨
- 추가적으로 극지연구소에서는 다양한 극지생물의 게놈 정보 발표를 준비 중에 있으나 이미 발표된 정보의 후속연구 처리 또는 관련 분야로의 활용이 예산 및 연구인력 부족으로 진행되기 힘든 상황임. 이에 따라 전 세계적으로 매우 탁월한 연구능력을 갖춘 극지유전체 연구팀으로 나아가기 위해 후속 유전체 연구가 반드시 뒷받침되어야 하며, 이를 통해 경골어류 내 독보적인 극지어류 유전전문성을 전 세계적으로 인정받을 수 있을 것이라 사료됨. 특히 나고야의정서와 관련하여 최근 공해 등에서 확보한 유전자원의 선점은 매우 중요한 이슈가 되었기에, 단 시간 내 유전정보를 학술결과로 발표함으로써 우리나라가 극지생물의 유전체 정보 확보 및 활용에 있어 우위를 점할 필요가 있음. 극지생물은 극한환경에 적응시 다양한 유전물질을 활용하여 왔기 때문에 유전체 정보 내 특이 기능성 유전물질을 발굴 시 산업적으로 개발하여 경제적인 부수물 창조에도 중요한 필요성이 있음

제 2 장 국내외 기술개발 현황

제 1 절 국내 기술개발 현황

- 삼면이 바다로 둘러싸인 우리나라의 장점을 극대화하여 수산물과 특히 해양생물의 유전체 정보를 확보할 수 있는 이점을 보유하고 있으며, 일부 해역을 북한, 중국, 일본 등과 공유하며 해양 유래 천해의 자원을 활용할 수 있는 지리적 장점을 지니고 있음. 그러나 독자적인 유전체 기반 수환경 모델을 개발하여 세계화에 성공한 일본과 유전체 분야에 엄청난 예산을 투자하여 최근 기술력과 논문의 질적 수준이 크게 증가한 중국 사이에서 점차 밀리고 있는 실정임. 특히 경제·산업적으로 매우 중요한 생물들과 수산자원들의 유전체 정보 확보에 대해 이미 많은 예산 투입으로 연구들이 상당수 진행되고 있기에 점차 주권을 확보할 기회가 줄어들 것이라 사료됨. 현재 우리나라는 해양수산부 및 해양생물자원관 등에서 몇몇 어류의 게놈 분석이 시도되고 있으나 진행 속도는 매우 더딘 편임. 최근 해양수산부 포스트게놈 다부처 유전체사업을 통해 대규모 연구비가 투입이 되고 있으며, 해조류 등에서 최정상급 논문이 출판되고 있으나 주요 수산물이나 특히 해양생물로부터의 유전체 정보의 확보와 발표는 더디게 진행되고 있는 실정임
- 현재까지 국내에서는 경골어류의 게놈 정보를 최정상급 국제전문학술지에 발표한 선례가 많지 않으며, 현 상황에서 특정 어류의 게놈 정보를 발표하더라도 기존의 탁월한 연구능력을 보유한 연구진들과의 연구능력 차이에 대한 갭은 줄이기 어려운 실정임. 따라서 특정 지역에 서식하며 그곳에서 독특한 진화양상으로 적응한 생물 또는 자국 내 토착종을 이용하여 장기적으로 꾸준한 연구를 하여야 하는 실정에 처하고 있으며, 극지연구소의 극지유전체사업단에서 이러한 니체를 활용하여 최정상급 진장유전체 분석 시스템 기반 하 꾸준하게 남극에 서식하는 남극대구(Shin et al., 2014)나 아이스피시(Kim et al., 2019) 등의 경골어류의 게놈 정보를 분야 내 최정상급 학술지에 발표하고 있음. 특히 최근 진장 유전체는 short-read sequencer가 아닌 long-read sequencer로 유전체를 분석해야만 최정상급 국제전문학술지에 논문을 투고할 수 있기에 sequencing platform은 세계적인 수준이나 이에 대한 후속 연구 및 응용 연구, 산업화 추진은 예산 및 인력 부족 등의 문제로

진행되기 힘든 상황임. 극지연구소는 이러한 국제적 연구 동향을 파악 후 long-read sequencer 중 고품질 유전체를 읽을 수 있는 PacBio RSII sequencer(그림 1)를 도입하여 전장 유전체 확보에 활용하고 있으며, 향후 몇 년 간 시퀀싱 품질과 양적인 측면에서

Platform Features



Feature	HiSeq2500 - Highoutput	HiSeq2500 - Rapid mode	MiSeq	PacBio RSII
Number of reads	150-180M/lane	100-150M/lane	12-15M (v2) 20-25M (v3)	50-80K/SMRT cell
Read length	2 x 100 bp	2 x 150 bp	2 x 300 bp (v3)	~ 10-20 kb
Yield per lane (PF data)	up to 35 Gb	up to 45Gb	up to 15 Gb	up to 0.4 Gb
Instrument Time	~12-14 days	~2 days	~2 days	~2 hours
Pricing per Gb	\$59 (PE100)	\$53 (PE150)	\$108 (PE300)	\$697

그림 1. 시판되는 게놈분석용 시퀀서들의 비교 분석
세계적인 연구 수준의 우위를 지닐 수 있으리라 판단됨

- 해외의 선도연구진의 사례나 영장류 게놈 분석 사업 등에서 확인할 수 있는 것처럼 게놈 정보는 단순히 DNA 서열을 읽은 뒤 논문과 데이터베이스 구축 등으로 발표하면 끝나는 것이 아닌 유전정보의 활용성에 따른 의학, 헬스, 웰빙, 산업, 경제적 이익 창출 등을 크게 이끌어낼 수 있는 중요한 정보임. 특히 나고야의정서(Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization to the Convention on Biological Diversity) 등 전 세계적으로 자국 내 유전정보의 이익을 극대화시키기 위한 노력에 편승하여 특정 지역에(예: 극한 환경) 서식하는 생물종이나 자국 내 토착종의 유전정보를 이용한 독자적인 모델화에 성공하는 사례가 크게 증가하고 있음
- 우리나라는 남극조약에 가입하여 과학 연구를 할 수 있는 권리를 지닌 몇 안 되는 나라 중 하나이며, 이와 같은 장점을 통해 남극생물 특이성 연구를 수행할 수 있는 장점을 지니고 있음. 특히 남극에 서식하는 어류들의 유전체 정보는 2000년대 중반부터 남극생물 전문과학자들에 의해 제시되어 왔으나 큰 가시적인 성과는 나타나지 않은 가운데 극지연구소 극지유전체사업단에서 최근 5년간 남극대구(Shin et al., 2014)나 아이스피시(Kim et al., 2019) 게놈 분석을 수행하며 이 정보들을 분야 내 최정상급 저널들인 Nature Ecology and Evolution에 논문을 출판하여 극지서식 경골어류 유전체 분석에 있어 탁월한 능력을 전 세계적으로 알린바 있음

제 2 절 국외 기술개발 현황

- 전 세계적으로 어류의 게놈 정보와 연속성 있는 유전체 기반 후속 연구에의 활용은 몇몇 주요 어류 유전학·진화학자들과 그들의 연구네트워크 형태의 컨소시엄들에 의해 큰 발전을 거듭하고 있음. 특히 특정 어류의 게놈 정보를 발표 후 꾸준하게 그 정보를 활용하며 비교분석을 통한 경골어류의 진화 양상 예측과 기능유전체를 바탕으로 한 다양한 분야에의 활용을 통해 독자적인 연구집단과 후속세대가 양성되고 있음. 최근에는 중국 BGI를 중심으로 대규모 시퀀싱을 저렴하게 서비스로 공급하고 있으며, 다수의 중국 연구진들이 황해에 서식하는 어류, 패류, 갑각류 등의 생물들 전장 유전체 정보를 심도 있는 분석 없이 공개함으로써 우위를 선점하며 해양생물 게놈생물학계의 질을 떨어뜨리고 있는 실정임

- Zebrafish의 경우 2000년 초반 전장 유전체의 정보가 밝혀진 후 현재까지 10개의 버전으로 게놈 정보가 계속해서 업데이트되고 있음(그림 2; Howe et al., 2013). 이에 따라 전 세계 어류생물학자들의 모델생물로의 입지는 확실하게 굳어졌으며, 특히 인간 질병 모델로 그 활용도가 점차 커지고 있음. 일본에서는 토착 송사리를

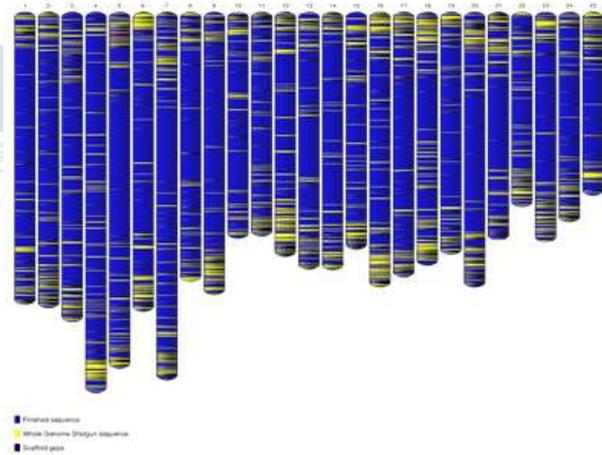


그림 2. Zebrafish genome project와 결과물인 염색체 수준의 genome database

이용하여 게놈 정보를 발표 후 꾸준한 업데이트를 통해 zebrafish에 이어 전 세계적으로 널리 사용되는 모델생물종화를 성공하였음. 이들 어류 외에도 다음의 예와 같이 여러 나라에서 각국의 이익에 맞는 어류들의 게놈이 발표되며 후속 연구를 통한 선진 기술 개발 및 국제적인 과학력 향상에 집중하고 있음.

예) African Turquoise killifish(Valenzano et al., 2015): 노화 연구, Atlantic cod(Star et al., 2011), Atlantic salmon(Lien et al., 2016) 및 European seabass(Tine et al., 2014): 수산양식 및 번역 연구, cavefish(McGaugh et al., 2014): 안구 질환 및 체내 감각 연구, coelacanth(Amemiya et al., 2013) 및 spotted gar(Braasch et al., 2016): 경골어류 진화, fugu(Aparicio et al., 2002),

lamprey(Smith et al., 2018) 및 rainbowtrout(Berthelot et al., 2014): 척추동물 계놈진화, stickleback(Jones et al., 2012): 어류의 담수 및 해수 적응 연구, *Xiphophorus* sp.(Schartl et al., 2013): 암 및 흑색종 연구(Texas University에 큰 규모의 *Xiphophorus* 질병 연구 센터를 설립)

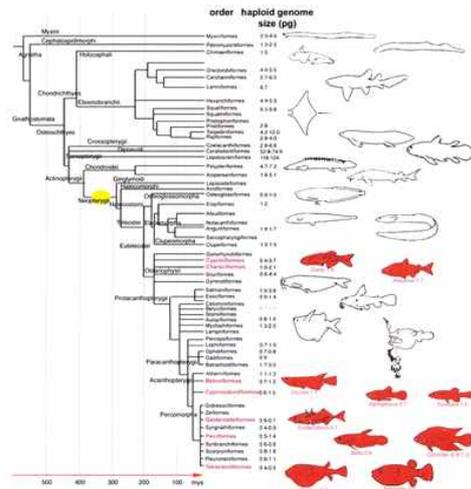


그림 3. 어류 계놈 분석 동향 및 분석된 어류들의 계통학적 관계

- 최근 중국 내에서 의학, 양식 또는 상업적으로 활용이 가능한 어류의 계놈을 발표하며 다양한 첨단 응용기술을

자국화하려는 데 큰 연구비와 인력을 투자하고 있음. 특히 grass carp(Wang et al., 2015)과 common carp(Xu et al., 2014) 계놈은 분야 내 최정상급 저널에 실린 후 꾸준히 수산양식 분야에서의 세계적 모델종으로 제시하고 있음. Mudskipper(You et al., 2014)의 계놈 정보를 이용한 어류 갯별 환경 적응 능력과 물-육지 간 생물의 이진 현상을 설명함으로 갯별 적응 특화 어류 모델종으로 개발하는 노력을 쏟고 있음. 최근에는 해마(Lin et al., 2016)와 Asian arowana(Austin et al., 2015)의 계놈 정보를 밝힘으로 경골어류 계놈 진화와 더불어

관상업계에서의 대규모 상업화를 추진 중임. 특히 중국연구진에 의해 최근 발표된 아시아권 수산양식 주요종인 넙치(Shao et al., 2017)의 계놈 정보는 중국을 넘어서 아시아 내 수산양식 분야 내 분자육종을 통한 매우 큰 경제적인 이익을 확보하는데 중요한 역할을 할 것이라 예상됨

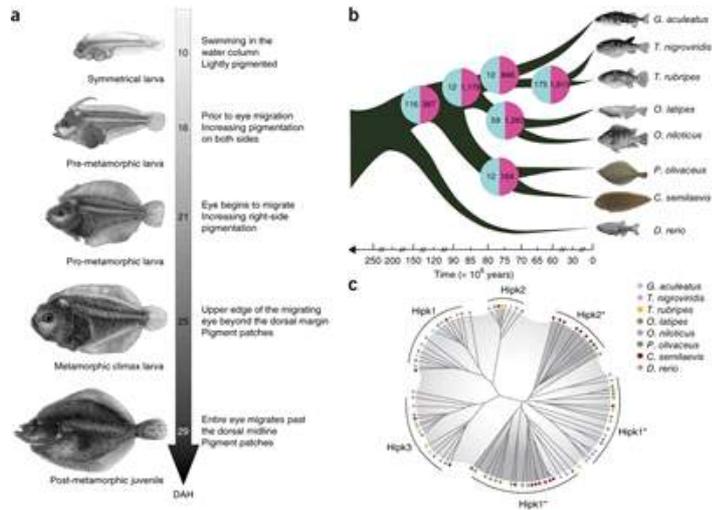


그림 4. 대표적 수산양식어류인 넙치의 계놈 분석 결과 요약

제 3 장 연구개발수행 내용 및 결과

제 1 절 이론적 및 실험적 접근방법

- 남극어류 게놈 정보는 연구비 효율성을 극대화시키기 위해 극지연구소 극지유전체사업단의 공조로 자체 서버 및 인트라넷 등을 공유하여 추가 분석비 없이 확보된 정보를 연구에 활용하였음. 분석에 사용된 bioinformatics는 기본적으로 웹 또는 프로그램화된 freeware를 활용하였으며, 기보유한 툴들을 최대한 활용하였음. 대부분의 bioinformatics 방법론은 최신 버전을 사용하였으며, 기타 필요한 분석들은 자체적 또는 게놈분석 연구네트워크를 최대한 활용하여 데이터를 확보하였음.

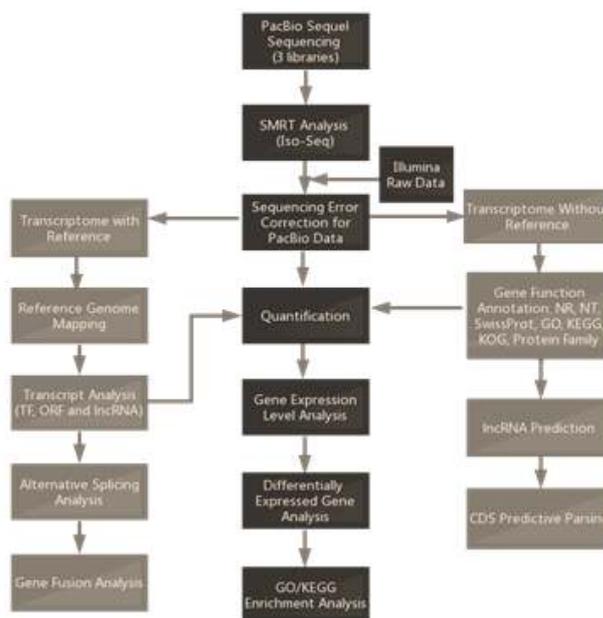


그림 5. 남극어류 게놈 re-assembly에 활용된 pipeline 예시

- Genome assembly는 5개 pipeline 이상을 변형하며 수행하였으며(그림 5), 최종적으로 2년간의 연구단계를 통해 기 발표된 극지어류 게놈 자체 분석 platform을 성공적으로 구축하였음

- *De novo* assembly는 FALCON-Unzip assembler (ver. 0.4, Falcon, Resource Identification Portal identification number (RRID):SCR_016089; Chin et al., 2016)를 사용하였으며, length_cutoff = 12,000, length_cutoff_pr = 10,000 파라미터를 이용하였음. 염색체 수준의 assembly를 위해 Hi-C 분석(Belton et al., 2012)을 수행하였으며, 최종 library는 Illumina Novaseq platform with 150 bp paired-end strategy 기반 시퀀싱을 수행하였음. Raw read는 HiRise로 mapping을 수행하였으며(Putnam et al., 2016), scaffold 1:1 매치는 pairwise aligner Large-Scale Genome Alignment Tool (LASTZ) version 1.10 (Harris, 2007)을 이용하였음. Assembly 완성도는 Benchmarking Universal Single-Copy

Orthologs (BUSCO) ver. 3.0 (RRID:SCR_015008)을 기반으로 Actinopterygii database (actinopterygii odb9 constructed from 20 fish species consisting of 4,584 orthologs; Simão, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015) 내 core gene set에 직접 mapping을 수행하였음

- 게놈 내 유전자 유무와 발현값을 확인하기 위해 전사체 sequencing을 수행하였음. 어류 조직들(혈액, 뇌, 눈, 부레, 아가미, 생식소, 신장, 간, 근육, 피부, 위 등)로부터 RNeasy Mini Kit (Qiagen)을 이용하여 total RNA를 추출하였으며, Agilent Bioanalyzer (Agilent Technologies)를 이용하여 추출한 RNA의 품질과 정량을 확인하였음. 약 2 ug의 total RNA 샘플을 SMRT cell v3 based on P6-C4 chemistry을 이용한 sequencing에 사용하였으며, PacBio Sequel system(Pacific Biosciences)에서 수행하였음. Iso-sequencing의 경우 SMRT Link (version 6.0.0)를 사용하였으며, 필터링, 품질 수준 분석 등을 수행하였음
- *De novo* repeat library의 경우 repeatmodeler version 1.0.3(RRID:SCR_015027; Bao & Eddy, 2002)와 함께 RECON version 1.08(Price, Jones, & Pevzner, 2005)과 repeatscout version 1.0.5 (RRID:SCR_014653; Price et al., 2005)를 이용해 확보하였음. Tandem Repeats Finder version 4.09을 이용해 consensus sequences 분석과 반복서열 각각의 classification을 예측하였음(Benson, 1999). 전체 transposable element의 나이와 분포는 Kimura distances(Kimura, 1980)를 이용해 분석하였음. Genome annotation은 maker version 2.28 (maker, RRID:SCR_005309)를 이용해 분석하였으며(Holt & Yandell, 2011), Ab initio gene prediction을 snap(snap, RRID:SCR_002127)과 Augustus(Augustus: Gene Prediction, RRID:SCR_008417)를 이용해 분석하였음. 단백질 정보는 12종의 기 발표된 어류 게놈들(*Astyanax mexicanus*, *Danio rerio*, *Gadus morhua*, *Gasterosteus aculeatus*, *Oreochromis niloticus*, *Oryzias latipes*, *Poecilia formosa*, *Takifugu rubripes*, *Tetraodon nigroviridis*, *Dicentrarchus labrax*, *Astyanax mexicanus*, *Xiphophorus maculatus*)을 사용하였으며, 최종적으로 gene prediction을 수행하였음. 단백질 서열들은 nonredundant database(E < 10⁻¹⁰)와 blastp version 2.2.29 기반 UniProtKB/Swissprot database에 직접적으로 align시켰으며, interproscan 5를 이용하여 blast2go-based gene ontology (GO) analysis(version 4.1.9; Gotz et al., 2008)를 수행하였음. Kyoto Encyclopedia of Genes and Genomes(KEGG) database를 이용해 유전자들의 기

능 예측을 수행하였으며, non-coding RNAs 예측은 infernal software package version 1.1(infernal, RRID:SCR_011809; Nawrocki, Kolbe, & Eddy, 2009)과 rfam(rfam, RRID:SCR_007891; Gardner et al., 2010) database를 이용해 수행하였음. tRNA와 이들의 이차 구조는 trnscan-se version 1.4(trnscan-se, RRID:SCR_010835; Lowe & Eddy, 1997)를 이용해 분석하였음

- 게놈 내 유전자의 open reading frame(ORF), exon/intron 분석 및 isoform genomic structure 분석은 Gbrowser 등을 활용하였으며, conserved or unique domain, motif 및 이들의 genome 내 배열과 분포도 분석 등은 연구진이 기보유한 일련의 bioinformatics 분석법을 활용하였음. 유전자 또는 gene family 간 계통도 분석은 PAUP 또는 MrBayes 등의 일련의 분석법을 활용하여 gene family 간 특이적 진화양상을 예측하였음. 12종의 게놈이 발표된 경골어류 정보를 이용하여 orthomcl(Li, Stoeckert, & Roos, 2003)와 orthomcl pipeline with application of the Markov Clustering Algorithm (MCL; Fischer et al., 2011)를 이용해 orthologous gene clusters를 분류하였으며, maker annotation pipeline (Holt & Yandell, 2011)을 이용해 최종적으로 유전자를 annotation하였음. 단백질 코돈 정렬은 Probabilistic Alignment Kit(prank; Löytynoja & Goldman, 2005)를 사용하였으며, Gblocks(-t=c -e=-gb1 -b4=5 -d=y; Castresana, 2000)을 이용해 코돈 내 갭을 예측하였음. 어류 게놈 내 divergence time 분석은 mega 6 software (Tamura, Stecher, Peterson, Filipowski, & Kumar, 2013)를 사용하여 generalised time-reversible model 기반 maximum-likelihood tree(RRID:SCR_006086; Stamatakis, 2014) 1,000 bootstrap 반복을 통해 생성하였음
- 어류 게놈 내 synteny 및 genomic structure 비교분석은 Ensemble 등 public database 또는 어류 게놈을 분석한 개별 연구진의 자체 database 서버를 활용하였으며, 극지연구소 극지유전체사업단 자체적으로 개발한 게놈 비교분석용 알고리즘을 활용한 분석 platform을 본 연구에 활용하였음. 앞서 분석된 orthologous gene clusters는 12종의 게놈 분석이 완료된 어류에 직접적으로 align시켰으며, positively selected genes(PSGs)을 예측하였음. 프로그램 prank를 이용해 alignment 품질을 분석하였으며, Gblocks를 이용해 저품질 서열(aligned sequences with <50% identity and shorter than 150 bp)을 제거하였음. dN, dS, 그리고 ω 값들은 paml package (RRID:SCR_014932)(Yang, 2007) 내

codeml program를 이용해 분석하였으며, Likelihood Ratio Tests (LRTs) 기반 basic and branchsite models을 이용해 PSGs를 추출하였음. 기능과 pathways 분류는 Fisher's exact test(cutoff: $p \leq .05$) 통계 기반 blast2go enrichment test (Conesa et al., 2005)를 통해 수행하였음

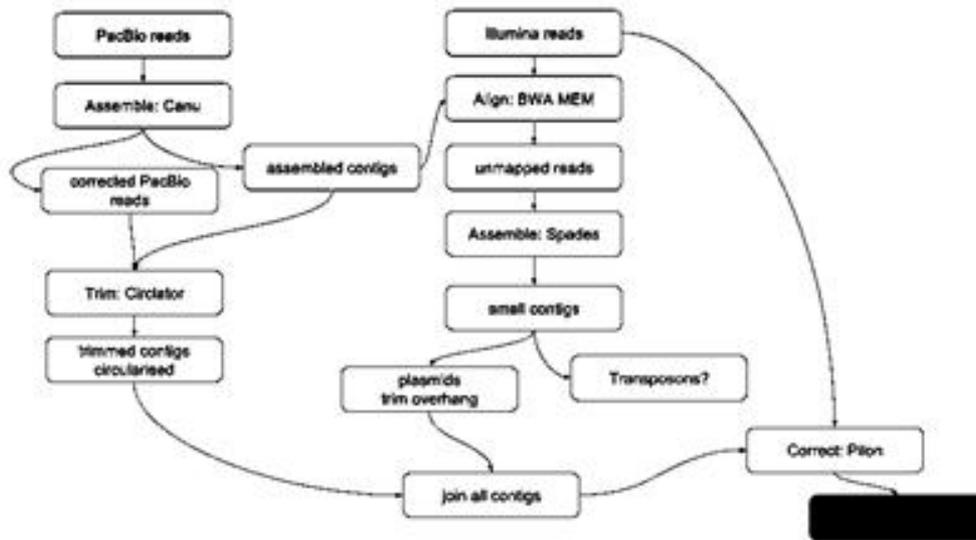


그림 6. PacBio raw read를 이용해 최종 유전자 annotation까지 진행한 workflow

극지연구소

- 게놈 비교분석에 있어 특정 염기서열의 오류 또는 불확실성으로 인해 어류 특이적 진화양상 및 gene family의 계통학적 추적이 상이한 양상으로 나올 가능성이 매우 클 것으로 예상하였기에 정확한 서열 분석이 전제로 주어져야 후속 연구의 진실성을 계속해서 유지할 수 있음. 이에 따라 추가로 sequencing error, unfilled gap, N 등의 미확인서열 또는 ORF와 exon/intron 분석은 기본적으로 PCR을 이용해 확인하였으며, 본 연구진은 상기 기초 bioinformatics tool과 분자생물학 기법을 응용할 기반이 잘 갖춘 상태임
- 최종적으로 확보한 결과들은 연구책임자의 국내외 어류 게놈 및 유전체 분석 관련 연구네트워크를 최대한 활용하여 극지어류로부터 게놈 특이성 및 유전체 진화와 환경 적응과의 관계를 밝힘으로 우수한 연구결과 양산, 후학 양성 및 논문출판을 통해 극지어류 게놈 연구에 있어 국내 연구진의 실력 향상에 최선으로 노력할 계획임

제 2 절 연구 내용

- 본 연구에서는 2단계의 연구 동안 계놈 정보가 발표된 남극어류들의 계놈을 이용하여 다음의 접근법으로 국내연구진의 남극어류 계놈분석에 있어 전문성과 탁월함을 알려 극지유전체 분석에 있어 탄탄한 입지를 구축하였음. 전체 연구내용의 요약은 다음 표와 같음

성과목표	세부목표	연구수행방법 (이론적·실험적 접근방법)	구체적인 내용
전체 어류 계놈 내 주요 유전자군 비교분석을 통한 남극어류 특이적 계놈 진화 분석	경골어류 진화를 이끌어온 주요 유전자군 정보 데이터베이스화	Bioinformatics	어류 유전체 데이터베이스 접근 및 전체 유전체 정보 다운로드
		Database 구축	어류 유전체 데이터베이스 구축
	남극 어류 계놈 내 주요 유전자군 정보 확립	Bioinformatics	어류 유전체 데이터베이스 내 유전자 검색 및 대상 유전자군에 속하는 유전자 리스트 확보
		Bioinformatics 및 매뉴얼 분석	유전자군에 속해있는 유전자들의 유사도 및 진화도 비교 분석
	남극어류 유전자군의 어류 내 계통학적 보존유전 또는 독립유전정보 분석	Bioinformatics 및 매뉴얼 분석	계놈 정보가 공개되어 있는 어류 유전체 정보 내 특정 유전자군의 비교 분석을 통한 남극어류 특이적 진화양상 분석

- 기존의 남극 적응성에 관련된 어류유전체 진화 이론에 대한 재검증: 현재까지 남극어류의 극한 환경에 대한 유전체 적응성 이론에 널리 활용되는 예는 antifreezing glycoprotein의 계놈 내 분화 및 반복, lack of hemoglobin expression and oxygen-binding protein myoglobin과 small fragment of -globin gene family, 그리고 mitochondrial metabolic change와 ATP 생합성간의 관계 등이 있음. 이러한 예들은 다양한 남극어류(약 15종)에서 밝혀졌으나,

게놈 정보 기반에 의한 분석이 아닌 특정 유전자에 대한 target-specific approach를 이용하여 밝혀진 정보임. 이에 따라 확보된 남극어류 게놈 정보를 이용하여 이들 적응요소들이 게놈 내 insertion, deletion, 또는 expansion 및 contraction 여부를 정확히 분석해 냄으로써 기존 이론을 재점검하며 특이점을 분석해냄으로써 교과서와 같은 이론 정립 체계를 구축함. 구체적으로 hox gene cluster, antifreezing protein, transient receptor potential(TRP), olfactory receptor superfamily, heat shock protein(Hsp) superfamily, zona pellucida(ZP) gene family, opsin gene family, hemoglobin coding gene cluster 등을 집중적으로 분석함

- 남극어류 각 종에 대한 계통학적 진화양상 추적: 남극어류는 극한 환경내에서도 다양한 종이 서식하고 있음. 상기 게놈 적응 요인 뿐 아니라 일반적으로 어류 게놈에서 다루어지는 게놈 진화 타겟 gene family 비교 분석을 통해 남극어류 각 종이 지니고 있는 적응 요인이 종(種; species) 또는 속(屬; genus) 특이성인지, 보다 높은 상위단계인 과(科; family) 또는 목(目; order) 단위에서 분지되어 유지된 것이거나 새롭게 획득한 것임을 분석하여 남극어류에 대한 진화양상을 추적하여 논문으로 최종 투고함

극지연구소

제 3 절 연구 결과

1. Hox gene cluster 분석

◦ Hox 유전자는 생물의 체형을 형성하는데 매우 중요한 유전자로 하등생물부터 척추동물까지 모두 그 기능을 보유함. 경골어류 genome duplication의 마커이자 진화 및 발생에 중요한 유전자들이 모여있는 곳이 hox gene cluster이며, 사지동물(tetrapod) 내에서도 유전자 발현 및 조절 패턴이 비슷할 것으로 알려짐. 이들은 고등생물로 올라갈수록 특정 염색체 내 유전자들이 밀집되는 경향을 나타내며, 생물마다 유전체 복제시 일부가 소실되는 경향이 나타나 이를 마커로 활용할 수 있음. 포유류는 총 4 세트를 지니고 있으나, 경골어류에서 증폭이 일어났으며 전체 유전자가 네 군데 혹은 그 이상의 염색체에 밀집되어 있음. 원구류나 연골어류에서는 정확한 증폭이 확인되지 않아 경골어류내에서는 전체 클러스터의 synteny 비교를 통해 계통학적 연관도 등을 측정하는 중요한 지표임. 극지어류 3종의 계놈 내에서 conserve한 hox gene cluster를 발굴하였으며, 3종의 남극어류들 모두 genome duplication 이후 안정적으로 유지된 것을 확인함(그림 7, 8)

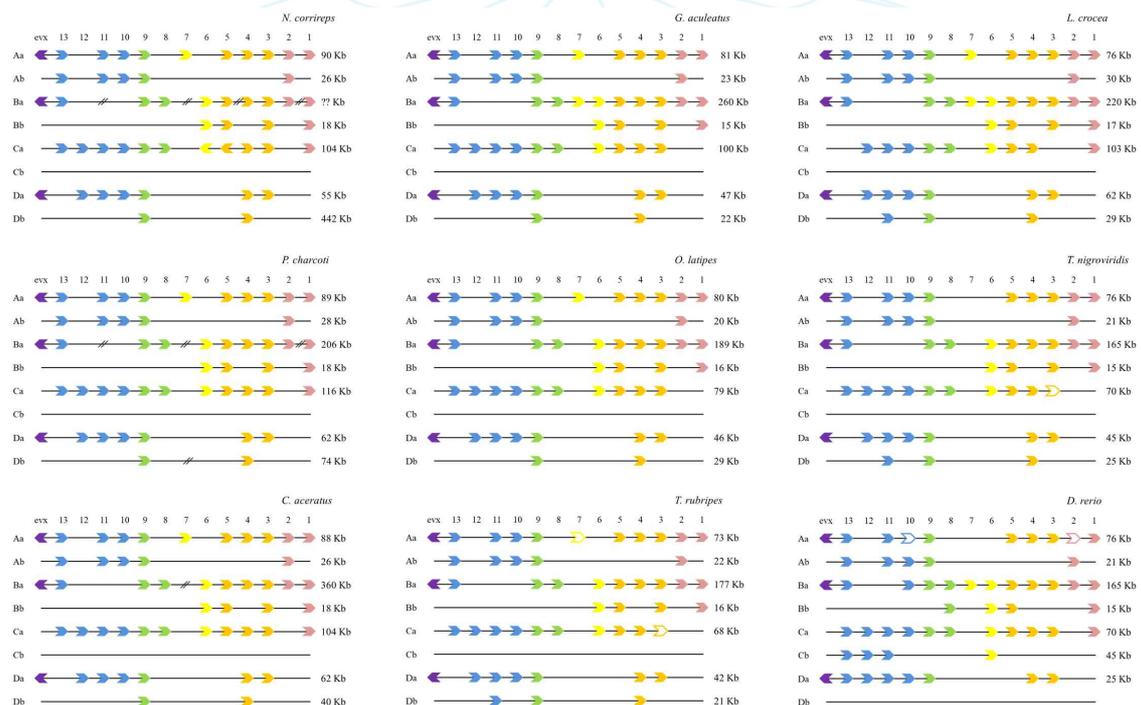


그림 7. 극지어류 3종을 포함한 경골어류들 계놈 내 분포하는 hox gene cluster 비교분석 결과

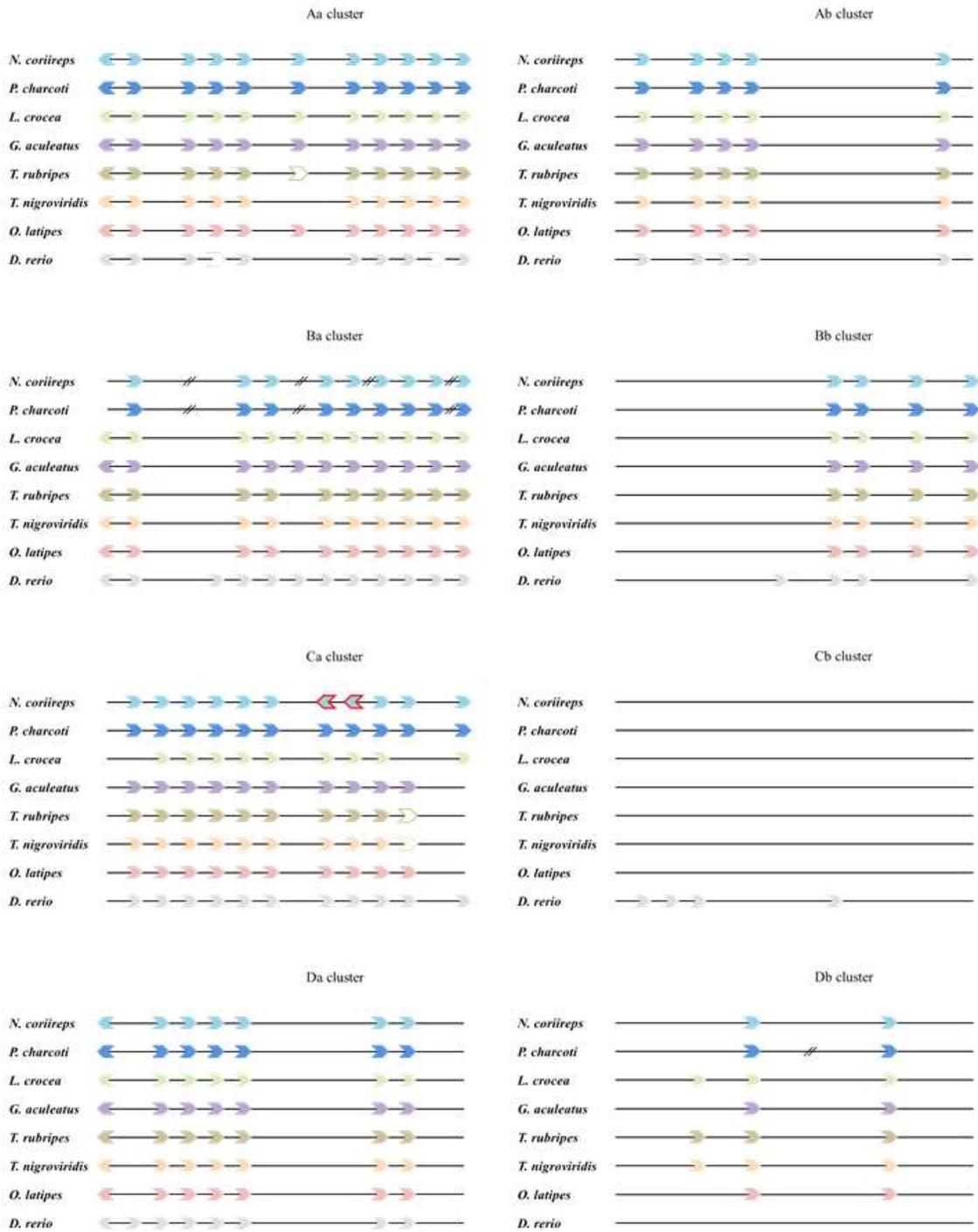


그림 8. 남극어류 3종을 포함한 경골어류 내 hox gene cluster 비교 분석 결과

2. Antifreezing protein 분석

- 남극에 서식하는 어류들은 결빙방지단백질을 합성하여 체 내 얼음생성을 억제하며, 보다 구체적으로 용질의 종류로 인한 어는점 내림 현상을 활용하는데, 물의 에너지와 얼음결정의 표면의 물 층의 에너지가 동일하게 되는 열적 평형 상태를 유도함으로써 얼음 결정을 커지지 않게 하는 것으로 알려짐. 특정 AFP의 경우 어류 내 synteny에서 유사하게 위치한 것을 확인하였으며(그림 9), 남극 빙어의 경우 이 유전자가 매우 큰 수로 증폭된 것으로 확인하였고(그림 10), 일부 열대성 어류에서는 유전자가 소실되거나 pseudogene 형태로 최소량만 보유한 것을 확인하였음

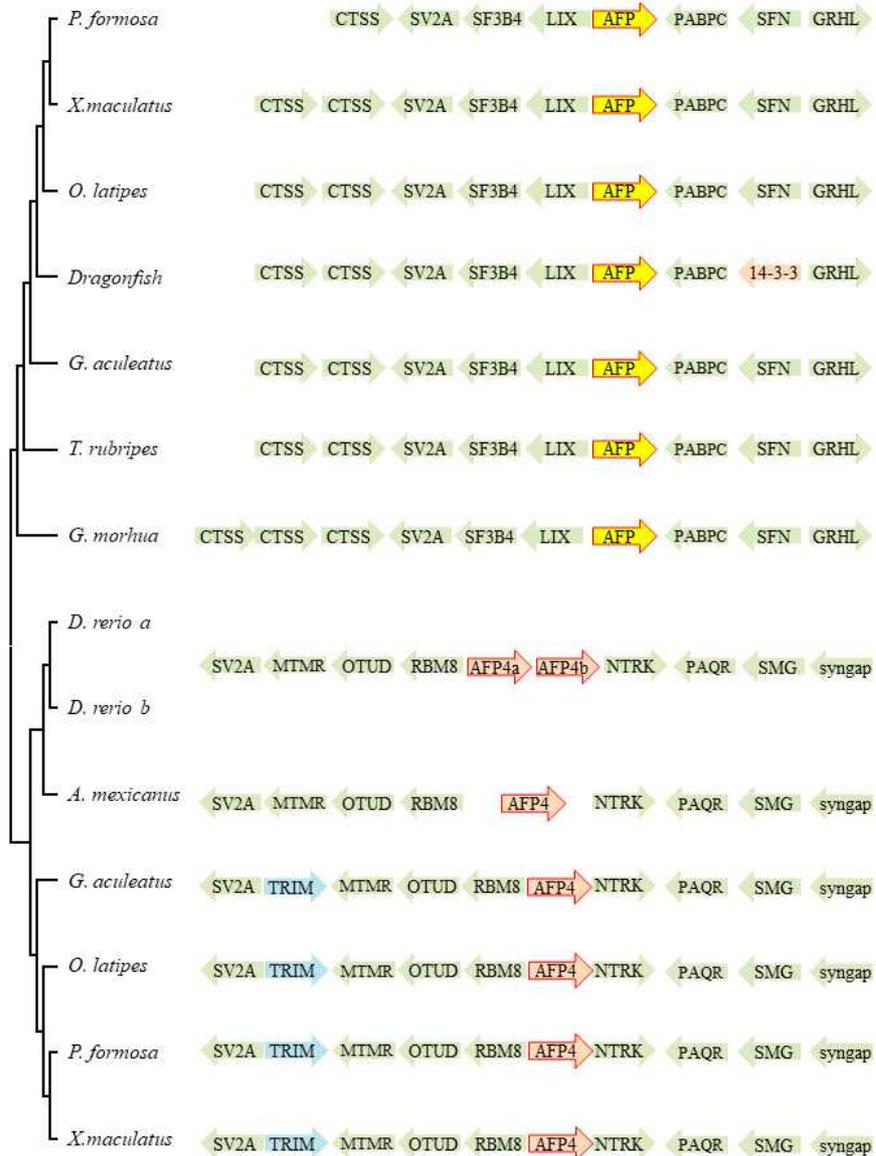


그림 9. 특정 AFP 유전자의 경골어류 내 위치 분석 결과

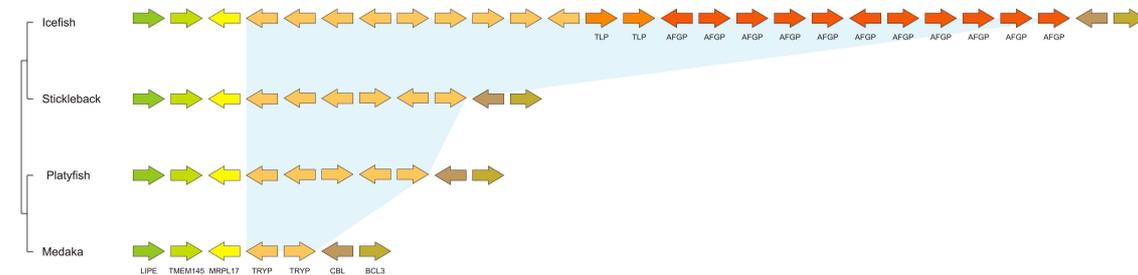


그림 10. 아열대성 어류 내 소실된 AFP와 대량 확장증폭된 남극빙어 AFP 비교 분석 결과

3. Transient receptor potential(TRP) 분석

◦ Transient receptor potential(TRP) 유전자는 superfamily 형태로 염색체 내에 보존되어 있으며, 특히 환경 스트레스에 민감하게 반응하여 생물체의 항상성 유지에 도움을 주는 것으로 알려져 있음. 남극어류 내 TRP 유전자가 특히 발달하여 있을 것이라 기대하였으나, TRPM8이 온도 변화에 관련된 유전자이기 때문에 드래곤피쉬 위주로 분석을 진행하였으나 synteny 상 큰 특이적 유전진화 양상은 발견되지 않았음(그림 11). 그러나 특이하게도 어류의 조상격으로 알려진 실러캔스는 TRPM8 유전자를 보유하고 있으나, 드래곤피쉬를 비롯해 모든 경골어류에서 이 유전자가 관찰이 되지 않음을 확인하고 현재 다른 TRP 유전자 클러스터들로 연구를 확장하고 있음

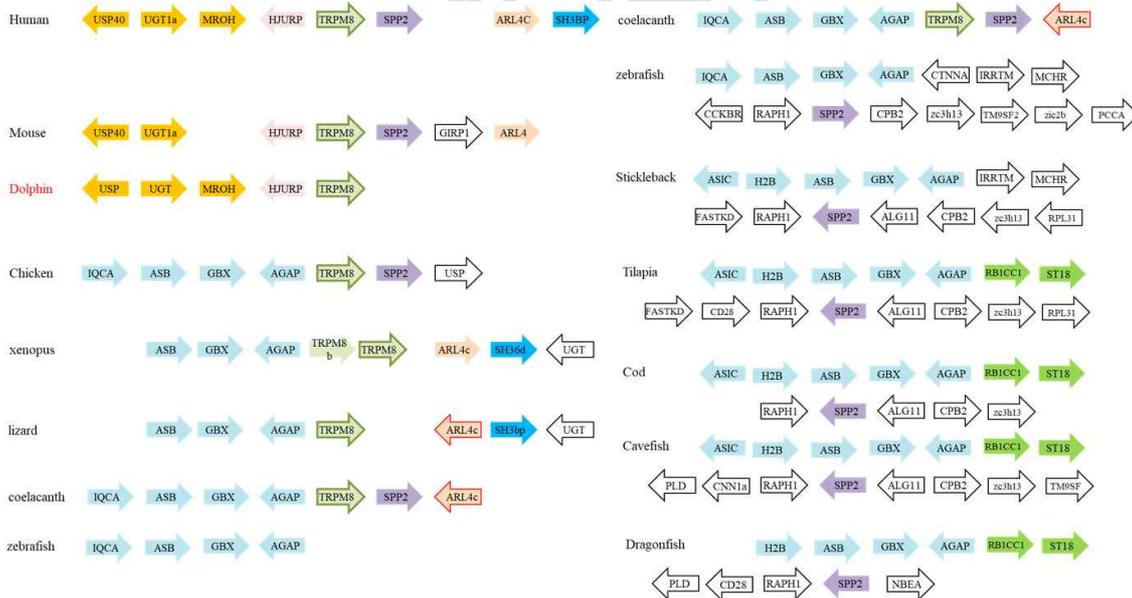


그림 11. 남극어류 포함한 경골어류 내 transient receptor potential gene superfamily 비교 분석 결과

4. Olfactory receptor superfamily 분석

◦ Olfactory receptor는 고등생물 계통 내 매우 복잡하게 다수의 유전자가 분포하고 있으며, 각기 환경 조건의 변화 또는 그에 대한 반응에 깊게 관여하고 있는 것으로 알려져 있음. 특이하게도 아이스피쉬 계통 내에서 일반적인 경골어류 대비 2~5배에 달하는 olfactory 유전자 superfamily가 증폭된 것을 확인하였음(그림 12좌). 이들 아이스피쉬 olfactory 유전자 superfamily의 경우 classification이 성공적으로 완성되었으나(그림 12우) 어류 내 유전자 수가 워낙 방대하여 전체 어류를 이용한 전체 classification 작업을 수행하고 있으며, 최종 논문으로 제출할 계획임

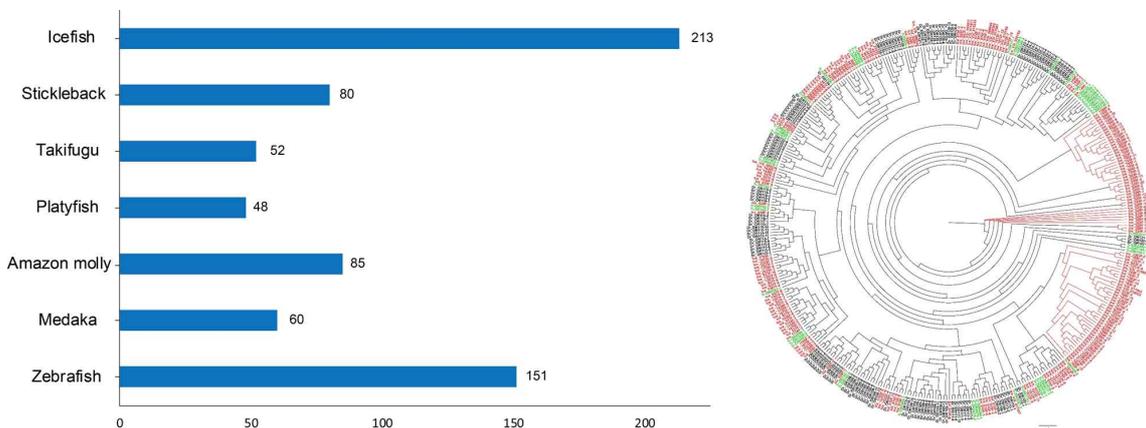


그림 12. 경골어류와 아이스피쉬 계통 내 olfactory gene superfamily 분포도(좌) 및 아이스피쉬 olfactory gene superfamily classification 결과(우)

5. Heat shock protein(Hsp) superfamily 분석

◦ Heat shock protein(Hsp)는 온도 변화에 민감하게 반응해 적응 또는 생리활성 조절에 관여하는 것으로 잘 알려진 유전자임. 극지어류를 포함하여 경골어류 계통 내 전체 Hsp superfamily 개수와 분포도를 확인한 결과 극지어류와 아열대 서식 어류 간 큰 차이점은 없는 것으로 확인되었음(그림 13)

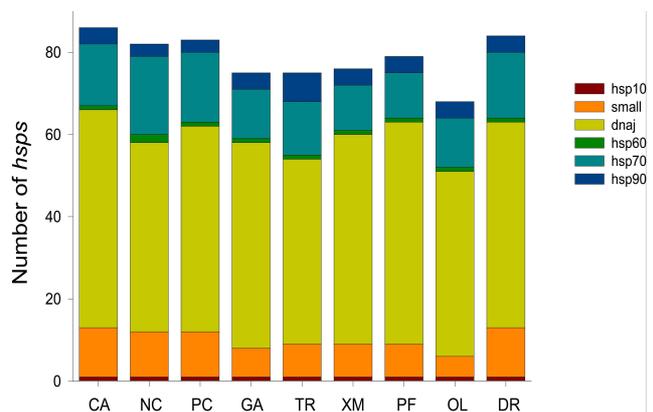


그림 13. 경골어류 계통 내 Hsp gene superfamily 비교 분석 결과

6. Zona pellucida gene family 분석

◦ Zona pellucida(ZP) gene은 어류의 난막 형성에 중요한 유전자이며, 특히 남극어류와 같이 극저온에서 알을 생성하는 메커니즘의 이해는 매우 중요함. ZP gene은 family 형태로 염색체에 분포하고 있으나 유전자 상동성이 높고 반복서열의 연속된 존재로 고품질 염색체 정보가 아닐시 분석할 수 없음. 본 연구를 통해 고품질 염색체 정보를 assembly할 수 있었고, 이를 통해 정확한 ZP gene family를 동정하였음. 특히 ZP gene family의 경우 ZPC5 유전자가 아이스피쉬 내 다량으로 증폭된 것으로 확인하였음(그림 14), 이를 통해 아이스피쉬가 극저온에서 알을 보호할 수 있는 메커니즘을 확인하였음

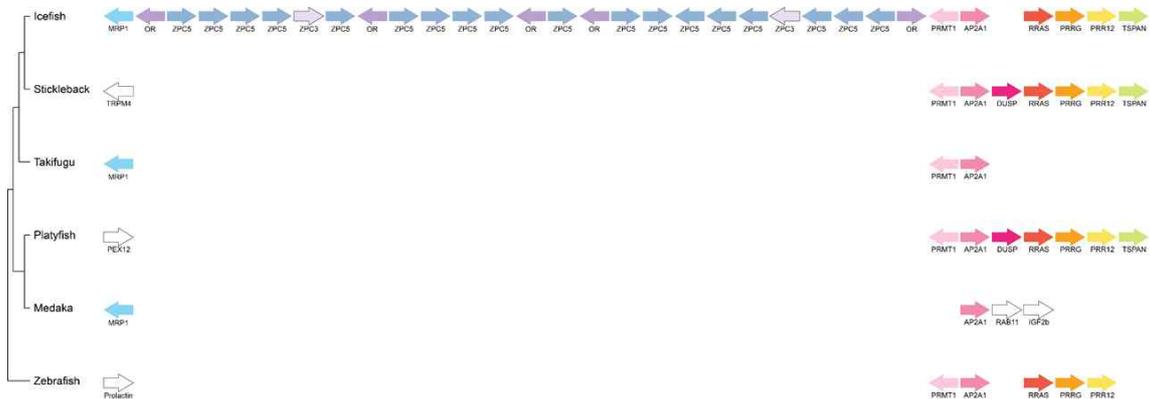


그림 14. 아이스피쉬와 경골어류 게놈 내 ZPC5 유전자 분포도 비교 분석 결과

7. Opsin gene family 분석

◦ 남극의 경우 빛의 양이 시간이나 일, 월 단위로 일정치 않기에 green, blue, red 계열의 opsin 유전자들을 분석하였으나 경골어류 대비 남극어류만의 큰 차이점을 발견하지는 못했음 (그림 15)

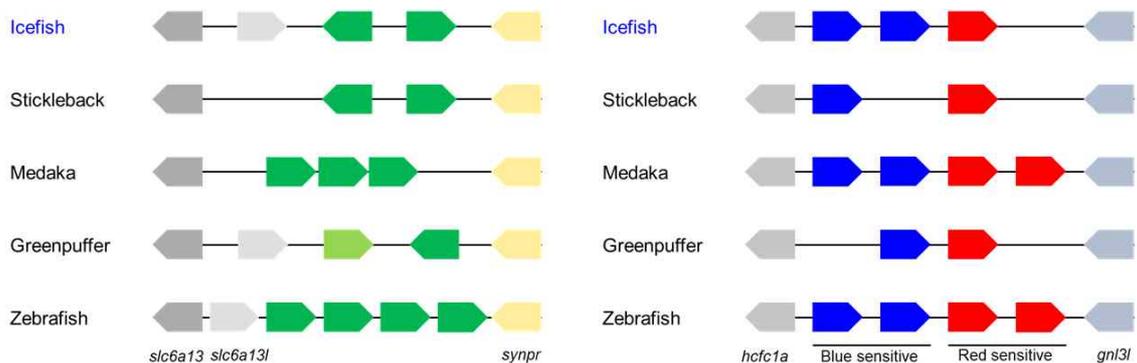


그림 15. 아이스피쉬와 경골어류 게놈 내 opsin 유전자 분포도 비교 분석 결과

8. Hemoglobin 유전자 분석

- 남극어류 중 아이스피쉬 등 몇 어류의 혈액은 헤모글로빈이 없기 때문에 혈액이 하얀색으로 나타남이 밝혀짐. 이에 대한 계통 상 정확한 검증이 현재까지 이루어지지 않았기에 본 연구를 통해 이를 경골어류 계통 내에서 정확하게 검증하였음. 아이스피쉬의 경우 LA 클러스터 내 a-globin cluster가 있는 것으로 나타났으나 partial 형태로 그 기능이 소실된 것으로 판단되며, MN 클러스터 내에서는 a-globin cluster와 a/b globin cluster가 모두 소실된 것으로 확인되었으며, 이로 인해 아이스피쉬 혈액은 하얀색을 유지하는 것이 최종 확인됨(그림 15)

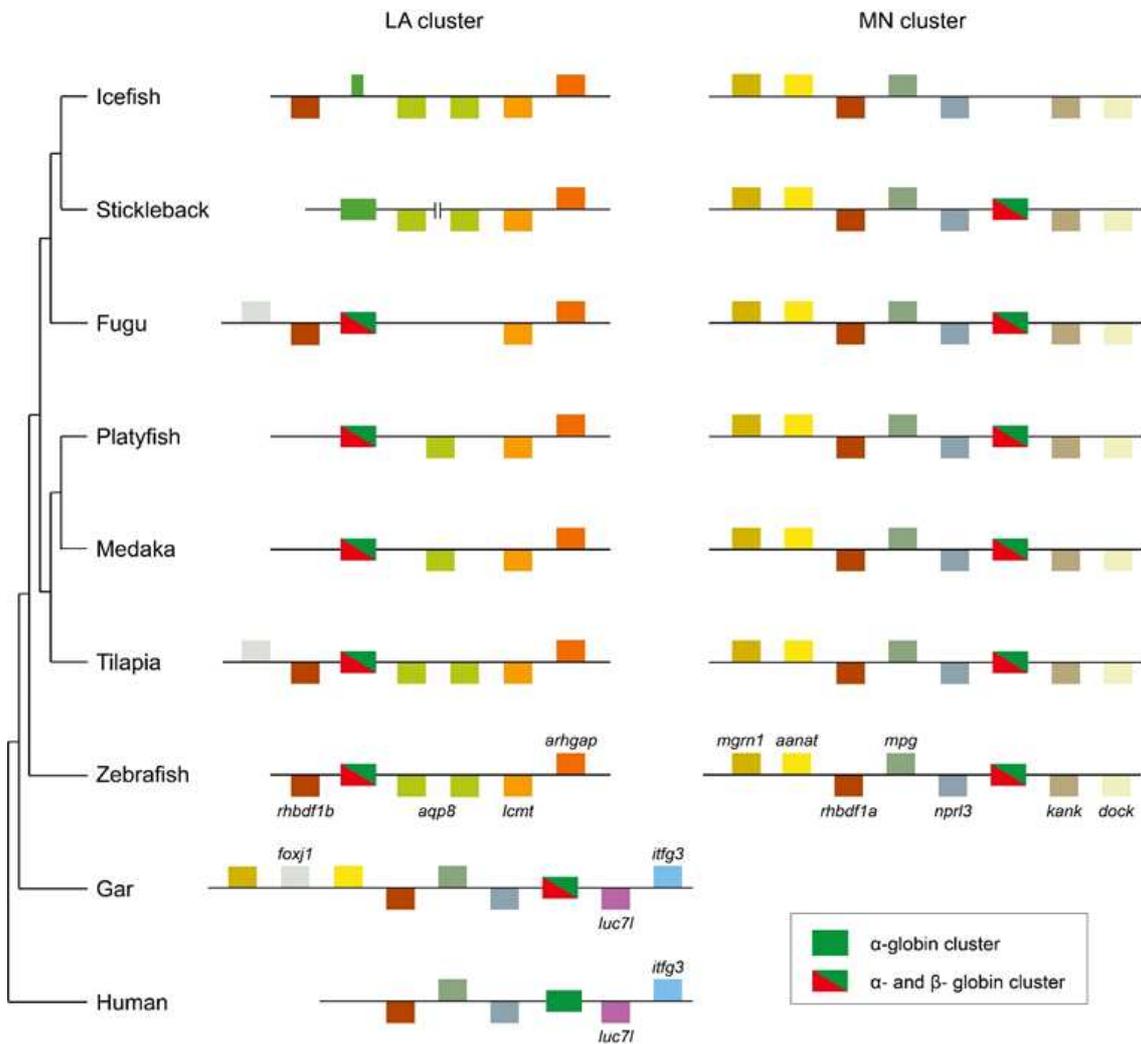


그림 16 아이스피쉬와 경골어류 계통 내 globin 유전자 분포도 비교 분석 결과

제 4 장 연구개발목표 달성도 및 대외기여도

제 1 절 연구개발목표의 달성도

성과목표	세부목표		평가지표	달성도 (%)
전체 어류 계통 내 주요 유전자군 비교분석을 통한 남극어류 특이적 계통 진화 분석	1-1	남극어류 계통 내 경골어류 진화를 이끌어온 주요 유전자군 정보 확보	<ul style="list-style-type: none"> - 경골어류 계통 진화에 관여하는 주요 유전자군 정보 분석 - 남극어류 계통 정보와의 비교 분석 	100
	1-2	남극 어류 계통 내 주요 유전자군 정보 확립	<ul style="list-style-type: none"> - 상기 남극어류 내 유전자군 분류 - 대상유전자군의 진화속도 비교분석 	
	1-3	남극어류 유전자군의 어류 내 계통학적 보존유전 또는 독립유전정보 분석	<ul style="list-style-type: none"> - 남극어류 유전자군의 계통학적 비교 분석 - 주요 유전자군의 보존 또는 독립진화여부 분석 	100
남극 환경 적응에 있어 남극어류 특이적 계통 진화의 역할 규명	2-1	남극어류 계통 특이적 진화양상 제시	<ul style="list-style-type: none"> - 주요 유전자군의 계통 내 synteny 비교 분석 - 남극 어류 특이적 유전자군의 insertion 및 deletion 여부 분석 	100
	2-2	남극어류 특이 계통 진화가 환경적응에 미치는 영향 분석	<ul style="list-style-type: none"> - 계통 내 특정 유전자 또는 유전자군의 변화가 환경적응성에 미치는 영향 예측 - 유전자군 기능예측을 통한 환경적응 능력과의 상관관계 분석 	100
	2-3	남극어류 유전자군의 어류 내 계통학적 보존유전 또는 독립유전정보 분석	<ul style="list-style-type: none"> - 현재까지 계통 정보가 밝혀진 어류들을 대상으로 남극어류 특이 유전자군 진화 분석 - 남극어류 특이 유전자군의 진화속도 분석 	100

제 2 절 관련분야의 기술발전예의 기여도

- 어류는 유전자의 유사도가 인간과 높은 가장 하등한 척추동물이기에 실험동물로의 개발, 유전자 기능 분석, 돌연변이 실험 등 다양한 목적을 위해 꾸준히 게놈 정보가 밝혀지고 있음. 그러나 약 3만여 종의 어류 중 게놈 정보가 밝혀진 경골어류 내 극지어류의 게놈 진화와 관련된 논문은 큰 임팩트를 만들어낼 수 있으나 극히 적은 수의 논문이 발표되었기에 극지어류 유전체 정보 활용도를 학술적으로 극대화할 수 있음. 실제 최근 발표된 아이스피쉬의 경우 Nature Ecology and Evolution에 발표되었으며(Kim et al., 2019), 이와 같이 최정상급 논문에 꾸준히 극지어류들의 유전체 정보가 발표될 시 국제적인 양 극지 생물들의 유전체 발굴에 우위를 점할 수 있으며, 선도적으로 연구진을 구축해 나갈 수 있음
- 특히 어류 게놈 진화 또는 특정 gene family 발굴 및 비교분석은 genome duplication이나 expansion/contraction 등 척추동물 진화 및 유전 관련 매우 중요한 분야임에도 우리나라에서 연구를 진행하고 있는 연구진이 거의 전무하기 때문에 국가 내 연구분야 불균형을 해소할 수 있음. 특히 국내에는 인간 유전체의 분석 능력이나 유전학적 통계는 수준급의 연구진이 있으나, 비모델생물이나 reference genome이 없는 생물들을 다루는 전문적인 연구진이 부족한 실정이기 때문에 전문 인력 양성에도 중요한 구심점으로 작용할 수 있음
- 극지어류 게놈 정보를 주도적으로 추진하고 있는 극지연구소의 유전체 연구 투자 및 추진력과 기발표된 게놈 정보를 활용한 후속연구 연계성을 하나의 platform으로 구축하여 극지어류 유전체 분야 내 독보적인 연구진 양성 가능하며, earth genome project 등 국제적인 유전체 분석기구에 투입되어 유전체 분석 능력을 크게 키울 수 있음
- 경골어류 게놈과 유전체 연구를 하는 세계적인 연구네트워크는 독자적인 연구네트워크를 구축하여 주기적으로 최정상급 학술지에 논문을 출판하고 있음. 이에 따라 상기 platform이 갖추어질 시 경골어류 게놈 내 극지어류 게놈에 관련된 정보를 바탕으로 국제적인 네트워크와 컨소시엄에 자연스레 합류할 수 있으며, 이를 통해 공동연구를 통한 최정상급 학술지에 논문을 출판하여 과급효과를 극대화할 수 있음.

제 5 장 연구개발결과의 활용계획

제 1 절 연구개발결과 활용계획

- 나고야의정서에 발맞추어 세계 각국은 자국 내 생물종들의 유전자원을 우선 등록하는 데 혈안이 되어 있음. 유전 정보로부터 당장 경제적 이득을 볼 수는 없을지라도 유전 정보 내 특이단백질을 생합성하는 유전자를 추후 발굴시 예측할 수 없는 경제적 효과를 불러올 수 있기 때문에 생물종 유지와 유전정보 확보는 각국의 이익에 크게 귀결되며 이를 통해 역으로 유전정보를 확보한 국가나 단체가 보호받을 수 있음. 이에 따라 독립적으로 극지어류의 유전정보를 확보하여 활용할 수 있는 극지유전체사업단의 데이터베이스는 장기적으로 보았을 때 국익 신장 및 예상할 수 없는 큰 경제적 이득을 가져올 수 있을 것이라 사료됨
- 극지어류로부터 추출되어 상용화된 대표적인 결빙방지단백질(antifreezing protein) 등의 예를 통해 확인할 수 있듯이 낮은 온도에 대한 적응 메커니즘의 주요 인자들을 발굴하면 큰 상업적인 효과를 불러올 수 있음. 위의 결빙방지단백질 역시 유전정보가 있어야만 추가 modification이나 editing 등을 통해 대량 생산이 가능하므로, 유전정보를 바탕으로 낮은 온도 적응 인자 개발 등을 통해 경제적인 이득을 확보할 수 있음. 본 연구를 통해 구축한 남극 어류를 포함한 경골어류 계통 정보는 관심 있는 국내 연구진에 우선적으로 오픈할 필요가 있으며, 이를 통해 상업화 또는 논문 작성에 필요한 유전자원을 공유할 시 국익과 연구력 향상에 큰 도움을 줄 수 있을 것으로 사료됨
- 극지생물 유전체 연구의 선점을 확보할 수 있고, 국제적인 입지를 강화시켜 대규모 연구비 소스를 만들어 내 이를 주도적으로 활용할 수 있음. 특히 과거의 1000K genome project나 현재 추진되는 Earth Genome Project 등 genome 분석과 관련된 사업은 그 연구적 임팩트와 경제산업적으로 중요한 부가 물질 개발과 직접적으로 연동되므로 국내 연구진의 실력을 강화하여 이러한 대규모 프로젝트에서 중추적 역할을 함과 동시에 대규모 연구비 소스를 모색할 필요가 있음. 부가적으로 유전체 분석 사업을 통해 전문 bioinformatics 인력을 양성할 수 있으며, 세계기구의 주도적 전략에 유연하게 대처할 수 있는 전문 체계를 구축할 수 있음

제 2 절 추가연구의 필요성

- 연구적 측면에서 membrane potential과 lipid metabolism에 관여하는 gene family 발굴 및 비교분석이 중요하나 lipid 생합성에 관여하는 유전자만 한정하여도 대규모 분석이 실시되어야 함. 극한 온도에 대한 분자적응능력 중 중요한 요소인 membrane potential 유지 및 막과 세포 내 지질 함유량 조절에 대한 연구는 현재까지 주로 미생물을 대상으로 연구가 이루어짐. 상대적으로 작은 게놈 크기와 적은 수의 유전자를 지닌 미생물을 대상으로 상기 유전요인들을 조사하는 것은 그리 어렵지 않으나, 게놈 크기가 크며 유전자의 개수와 isoform과 같은 gene variant가 높은 어류에서는 이러한 연구가 거의 이루어지지 않음. 이러한 니체를 잘 활용하여 막 구성에 주요한 역할을 하는 gene family와 지질 대사에 관여하는 대규모 gene family를 분석하여 일반적인 어류에서도 잘 연구되지 않은 미개척 분야에 도전할 계획임. 특히 어류는 포유류와 지방 대사 메커니즘이 완전히 다름에도 제대로 된 검증이 이루어지지 않고 있음. 특히 갈색지방과 백색지방 메커니즘은 어류에서 많이 소실된 것으로 알려져 있으나, 극지어류의 경우 이미 극저온에 적응하였기에 독립적인 지방 메커니즘을 지닐 것으로 예상되며, 이를 위해 유전체 및 유전자 기능 연구가 추진되어야 할 것으로 판단됨
- 유전체 분석은 꾸준히 개발되는 알고리즘과 게놈 정보를 직접 다루며 오랜 시간동안 집중해야 전문 실력이 쌓일 수 있음. 이에 따라 후학 양성과 유전체 분석 전문 인력 양성을 위해 계속해서 어류를 비롯한 극지생물의 유전체를 다룰 수 있는 인력 양성 사업을 지속적으로 추진해야 할 것으로 사료됨

제 6 장 연구개발과정에서 수집한 해외과학기술정보

연구 내용	연구 결과
남극대구, 에메랄드피시, 아이스피시 등 주요 남극어류 유전체 정보 확보	현재까지 공개적으로 발표된 남극에 서식하는 대표 어류 남극대구, 에메랄드피시, 아이스피시에 대한 유전체 정보를 NCBI 등 genome server에서 확보하였음
남극대구, 에메랄드피시, 아이스피시 등 주요 남극어류 내 유전자군 분류	현재까지 공개적으로 발표된 남극에 서식하는 대표 어류 남극대구, 에메랄드피시, 아이스피시에 대한 유전체 정보로부터 어류 발생에 중요한 hox gene cluster, 성결정 및 성분화에 중요한 sex gene family, transposable element gene family 등 주요 유전자군을 분류하였음
현재까지 게놈 정보가 밝혀진 어류들을 대상으로 남극어류 특이 유전자군 진화 분석	현재까지 게놈 정보가 공개된 아열대성 어류들의 유전체 정보를 Ensemble과 NCBI 및 자체 수립 서버로부터 확보하였으며 ¹⁾ , 전체 유전체 정보로부터 유전자군의 비교분석을 위해 분석대상 유전자들을 분류하였음 ²⁾

1) Fish genome assemblies and/or gene models used in this study.

Common name	Species	Assembly/Gene model	#Genes	Reference
Tiger puffer	<i>Takifugu rubripes</i>	FUGU 5.0	22,760	Aparicio et al., 2002
Green spotted puffer	<i>Tetraodon nigroviridis</i>	TETRAODON 8.0	23,118	Jaillon et al., 2004
Stickleback	<i>Gasterosteus aculeatus</i>	BROAD S1	27,576	Jones et al., 2012
European seabass	<i>Dicentrarchus labrax</i>	dicLab v1.0c	26,719	Tine et al., 2014
Medaka	<i>Oryzias latipes</i>	HdrR	24,674	Kasahara et al., 2007
Clown fish	<i>Amphiprion ocellaris</i>	dx.doi.org/10.5524/100397	27,240	Tan et al., 2018
Nile tilapia	<i>Oreochromis niloticus</i>	Orenill.0	26,763	Conte et al., 2017
Platyfish	<i>Xiphophorus maculatus</i>	Xipmac4.4.2	20,454	Schartl et al., 2013
Amazon molly	<i>Poecilia formosa</i>	Poecilia_formosa-5.1.2	30,898	Warren et al., 2018

Atlantic cod	<i>Gadus morhua</i>	gadMor1	22,100	Star et al., 2011
Zebrafish	<i>Danio rerio</i>	GRCz10	25,403	Howe et al., 2013
Cavefish	<i>Astyanax mexicanus</i>	AstMex102	23,698	McGaugh et al., 2014

2) Summary of orthologous gene clusters analyzed in 13 teleost species.

Species name	No. of coding genes	No. of expanded families*	No. of contracted families*	No. of genes lost
<i>Takifugu rubripes</i>	22,760	3,139 (136)	2,913 (71)	3,026
<i>Tetraodon nigroviridis</i>	23,118	1,788 (38)	2,910 (2)	2,927
<i>Gasterosteus aculeatus</i>	27,576	3,899 (179)	3,337 (8)	3,480
<i>Dicentrarchus labrax</i>	26,719	623 (127)	1,893 (2)	1,985
<i>Oryzias latipes</i>	24,674	3,812 (282)	3,141 (7)	3,194
<i>Amphiprion ocellaris</i>	27,240	871 (86)	1,889 (20)	2,080
<i>Oreochromis niloticus</i>	26,763	2,676 (208)	4,695 (9)	4,876
<i>Xiphophorus maculatus</i>	20,454	127 (4)	2,480 (145)	2,797
<i>Poecilia formosa</i>	30,898	4,847 (1,070)	630 (4)	634
<i>Gadus morhua</i>	22,100	801 (132)	7,775 (8)	8,047
<i>Danio rerio</i>	25,403	760 (58)	8,123 (0)	8,359
<i>Astyanax mexicanus</i>	23,698	975 (11)	8,251 (3)	8,481

*Numbers in parentheses are the number of significantly changed genes.

제 7 장 참고문헌

- Amemiya CT, Alföldi J, Lee AP, Fan S, Philippe H, Maccallum I, Braasch I, Manousaki T, Schneider I, Rohner N, Organ C, Chalopin D, Smith JJ, Robinson M, Dorrington RA, Gerdol M, Aken B, Biscotti MA, Barucca M, Baurain D, Berlin AM, Blatch GL, Buonocore F, Burmester T, Campbell MS, Canapa A, Cannon JP, Christoffels A, De Moro G, Edkins AL, Fan L, Fausto AM, Feiner N, Forconi M, Gamielien J, Gnerre S, Gnirke A, Goldstone JV, Haerty W, Hahn ME, Hesse U, Hoffmann S, Johnson J, Karchner SI, Kuraku S, Lara M, Levin JZ, Litman GW, Mauceli E, Miyake T, Mueller MG, Nelson DR, Nitsche A, Olmo E, Ota T, Pallavicini A, Panji S, Picone B, Ponting CP, Prohaska SJ, Przybylski D, Saha NR, Ravi V, Ribeiro FJ, Sauka-Spengler T, Scapigliati G, Searle SM, Sharpe T, Simakov O, Stadler PF, Stegeman JJ, Sumiyama K, Tabbaa D, Tafer H, Turner-Maier J, van Heusden P, White S, Williams L, Yandell M, Brinkmann H, Volff JN, Tabin CJ, Shubin N, Schartl M, Jaffe DB, Postlethwait JH, Venkatesh B, Di Palma F, Lander ES, Meyer A, Lindblad-Toh K. The African coelacanth genome provides insights into tetrapod evolution. *Nature*. 2013 Apr 18;496(7445):311-6.
- Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels A, Rash S, Hoon S, Smit A, Gelpke MD, Roach J, Oh T, Ho IY, Wong M, Detter C, Verhoef F, Predki P, Tay A, Lucas S, Richardson P, Smith SF, Clark MS, Edwards YJ, Doggett N, Zharkikh A, Tavtigian SV, Pruss D, Barnstead M, Evans C, Baden H, Powell J, Glusman G, Rowen L, Hood L, Tan YH, Elgar G, Hawkins T, Venkatesh B, Rokhsar D, Brenner S. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science*. 2002 Aug 23;297(5585):1301-10.
- Austin CM, Tan MH, Croft LJ, Hammer MP, Gan HM. Whole Genome Sequencing of the Asian Arowana (*Scleropages formosus*) Provides Insights into the Evolution of Ray-Finned Fishes. *Genome Biol Evol*. 2015 Oct 6;7(10):2885-95.
- Bao, Z., & Eddy, S. R. (2002). Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Research*, 12, 1269 - 1276.
- Benson, G. (1999). Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Research*, 27, 573 - 580.
- Berthelot C, Brunet F, Chalopin D, Juanchich A, Bernard M, Noël B, Bento P, Da Silva C, Labadie K, Alberti A, Aury JM, Louis A, Dehais P, Bardou P, Montfort J, Klopp C, Cabau C, Gaspin C, Thorgaard GH, Boussaha M, Quillet E, Guyomard R, Galiana D, Bobe J, Volff JN, Genêt C, Wincker P, Jaillon O, Roest Crollius H, Guiguen Y. The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nat Commun*. 2014 Apr 22;5:3657.
- Braasch I, Gehrke AR, Smith JJ, Kawasaki K, Manousaki T, Pasquier J, Amores A, Desvignes T, Batzel P, Catchen J, Berlin AM, Campbell MS, Barrell D,

- Martin KJ, Mulley JF, Ravi V, Lee AP, Nakamura T, Chalopin D, Fan S, Weisel D, Cañestro C, Sydes J, Beaudry FE, Sun Y, Hertel J, Beam MJ, Fasold M, Ishiyama M, Johnson J, Kehr S, Lara M, Letaw JH, Litman GW, Litman RT, Mikami M, Ota T, Saha NR, Williams L, Stadler PF, Wang H, Taylor JS, Fontenot Q, Ferrara A, Searle SM, Aken B, Yandell M, Schneider I, Yoder JA, Volff JN, Meyer A, Amemiya CT, Venkatesh B, Holland PW, Guiguen Y, Bobe J, Shubin NH, Di Palma F, Alfo Ldi J, Lindblad-Toh K, Postlethwait JH. The spotted gar genome illuminates vertebrate evolution and facilitates human–teleost comparisons. *Nat Genet.* 2016 Apr;48(4):427–37.
- Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, 17, 540 - 552.
- Chalopin, D., Naville, M., Plard, F., Galiana, D., Volff, J.N., 2015. Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. *Genome Biol. Evol.* 7, 567–580.
- Chin CS, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, Dunn C, O'Malley R, Figueroa-Balderas R, Morales-Cruz A, Cramer GR, Delledonne M, Luo C, Ecker JR, Cantu D, Rank DR, Schatz MC. (2016). Phased diploid genome assembly with single molecule real-time sequencing. *Nature Methods*, 13, 1050 - 1054.
- Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., & Robles, M. (2005). blast2go: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 21, 3674 - 3676.
- Conte, M.A., Gammerdinger, W.J., Bartie, K.L., Penman, D.J., Kocher, T.D., 2017. A high quality assembly of the Nile Tilapia (*Oreochromis niloticus*) genome reveals the structure of two sex determination regions. *BMC Genomics* 18, 341.
- Fischer, S., Brunk, B. P., Chen, F., Gao, X., Harb, O. S., Iodice, J. B., Stoeckert Jr, C. J. (2011). Using orthomcl to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. *Current Protocols in Bioinformatics*, Chapter 6, Unit 6.12, 1 - 19.
- Gardner PP, Daub J, Tate J, Moore BL, Osuch IH, Griffiths-Jones S, Finn RD, Nawrocki EP, Kolbe DL, Eddy SR, Bateman A. (2010). rfam: Wikipedia, clans and the “decimal” release. *Nucleic Acids Research*, 39, D141 - D145.
- Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A. (2008). High-throughput functional annotation and data mining with the blast2go suite. *Nucleic Acids Research*, 36, 3420 - 3435.
- Harris, R. S. (2007). Improved pairwise alignment of genomic DNA. PhD thesis. State College, PA: Pennsylvania State University.
- Holt, C., & Yandell, M. (2011). maker2: An annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*, 12, 491.
- Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, Collins JE, Humphray S, McLaren K, Matthews L, McLaren S, Sealy I, Caccamo M, Churcher C, Scott C, Barrett JC, Koch R, Rauch GJ, White S, Chow W,

- Kilian B, Quintais LT, Guerra-Assunção JA, Zhou Y, Gu Y, Yen J, Vogel JH, Eyre T, Redmond S, Banerjee R, Chi J, Fu B, Langley E, Maguire SF, Laird GK, Lloyd D, Kenyon E, Donaldson S, Sehra H, Almeida-King J, Loveland J, Trevanion S, Jones M, Quail M, Willey D, Hunt A, Burton J, Sims S, McLay K, Plumb B, Davis J, Clee C, Oliver K, Clark R, Riddle C, Elliot D, Threadgold G, Harden G, Ware D, Begum S, Mortimore B, Kerry G, Heath P, Phillimore B, Tracey A, Corby N, Dunn M, Johnson C, Wood J, Clark S, Pelan S, Griffiths G, Smith M, Glithero R, Howden P, Barker N, Lloyd C, Stevens C, Harley J, Holt K, Panagiotidis G, Lovell J, Beasley H, Henderson C, Gordon D, Auger K, Wright D, Collins J, Raisen C, Dyer L, Leung K, Robertson L, Ambridge K, Leongamornlert D, McGuire S, Gilderthorp R, Griffiths C, Manthravadi D, Nichol S, Barker G, Whitehead S, Kay M, Brown J, Murnane C, Gray E, Humphries M, Sycamore N, Barker D, Saunders D, Wallis J, Babbage A, Hammond S, Mashreghi-Mohammadi M, Barr L, Martin S, Wray P, Ellington A, Matthews N, Ellwood M, Woodmansey R, Clark G, Cooper J, Tromans A, Grafham D, Skuce C, Pandian R, Andrews R, Harrison E, Kimberley A, Garnett J, Fosker N, Hall R, Garner P, Kelly D, Bird C, Palmer S, Gehring I, Berger A, Dooley CM, Ersan-Ürün Z, Eser C, Geiger H, Geisler M, Karotki L, Kirn A, Konantz J, Konantz M, Oberländer M, Rudolph-Geiger S, Teucke M, Lanz C, Raddatz G, Osoegawa K, Zhu B, Rapp A, Widaa S, Langford C, Yang F, Schuster SC, Carter NP, Harrow J, Ning Z, Herrero J, Searle SM, Enright A, Geisler R, Plasterk RH, Lee C, Westerfield M, de Jong PJ, Zon LI, Postlethwait JH, Nüsslein-Volhard C, Hubbard TJ, Roest Crollius H, Rogers J, Stemple DL. The zebrafish reference genome sequence and its relationship to the human genome. *Nature*. 2013 Apr 25;496(7446):498–503.
- Jaillon, O., Aury, J. M., Brunet, F., Petit, J. L., Stange-Thomann, N., Mauceli, E., ... Roest Crollius, H. (2004) Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature*, 431, 946 - 957.
- Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC, White S, Birney E, Searle S, Schmutz J, Grimwood J, Dickson MC, Myers RM, Miller CT, Summers BR, Knecht AK, Brady SD, Zhang H, Pollen AA, Howes T, Amemiya C; Broad Institute Genome Sequencing Platform & Whole Genome Assembly Team, Baldwin J, Bloom T, Jaffe DB, Nicol R, Wilkinson J, Lander ES, Di Palma F, Lindblad-Toh K, Kingsley DM. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*. 2012 Apr 4;484(7392):55–61.
- Kim BM, Amores A, Kang S, Ahn DH, Kim JH, Kim IC, Lee JH, Lee SG, Lee H, Lee J, Kim HW, Desvignes T, Batzel P, Sydes J, Titus T, Wilson CA, Catchen JM, Warren WC, Schartl M, Detrich HW 3rd, Postlethwait JH, Park H. Antarctic blackfin icefish genome reveals adaptations to extreme environments. *Nat Ecol Evol*. 2019 Mar;3(3):469–478.
- Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, 16, 111 - 120.

- Li, L., Stoeckert, C. J. Jr, & Roos, D. S. (2003). orthomcl: Identification of ortholog groups for eukaryotic genomes. *Genome Research*, 13, 2178 - 2189.
- Lien S, Koop BF, Sandve SR, Miller JR, Kent MP, Nome T, Hvidsten TR, Leong JS, Minkley DR, Zimin A, Grammes F, Grove H, Gjuvsland A, Walenz B, Hermansen RA, von Schalburg K, Rondeau EB, Di Genova A, Samy JK, Olav Vik J, Vigeland MD, Caler L, Grimholt U, Jentoft S, Våge DI, de Jong P, Moen T, Baranski M, Palti Y, Smith DR, Yorke JA, Nederbragt AJ, Tooming-Klunderud A, Jakobsen KS, Jiang X, Fan D, Hu Y, Liberles DA, Vidal R, Iturra P, Jones SJ, Jonassen I, Maass A, Omholt SW, Davidson WS. The Atlantic salmon genome provides insights into rediploidization. *Nature*. 2016 May 12;533(7602):200–5.
- Lin Q, Fan S, Zhang Y, Xu M, Zhang H, Yang Y, Lee AP, Woltering JM, Ravi V, Gunter HM, Luo W, Gao Z, Lim ZW, Qin G, Schneider RF, Wang X, Xiong P, Li G, Wang K, Min J, Zhang C, Qiu Y, Bai J, He W, Bian C, Zhang X, Shan D, Qu H, Sun Y, Gao Q, Huang L, Shi Q, Meyer A, Venkatesh B. The seahorse genome and the evolution of its specialized morphology. *Nature*. 2016 Dec 14;540(7633):395–399.
- Lowe, T. M., & Eddy, S. R. (1997). trnscan-se: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, 25, 955 - 964.
- Löytynoja, A., & Goldman, N. (2005). An algorithm for progressive multiple alignment of sequences with insertions. *Proceedings of the National Academy of Sciences USA*, 102, 10557 - 10562.
- McGaugh SE, Gross JB, Aken B, Blin M, Borowsky R, Chalopin D, Hinaux H, Jeffery WR, Keene A, Ma L, Minx P, Murphy D, O'Quin KE, Rétaux S, Rohner N, Searle SM, Stahl BA, Tabin C, Volff JN, Yoshizawa M, Warren WC. The cavefish genome reveals candidate genes for eye loss. *Nat Commun*. 2014 Oct 20;5:5307.
- Nawrocki, E. P., Kolbe, D. L., & Eddy, S. R. (2009). infernal 1.0: Inference of RNA alignments. *Bioinformatics*, 25, 1335 - 1337.
- Price, A. L., Jones, N. C., & Pevzner, P. A. (2005). De novo identification of repeat families in large genomes. *Bioinformatics*, 21, i351 - i358.
- Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, Calef R, Troll CJ, Fields A, Hartley PD, Sugnet CW, Haussler D, Rokhsar DS, Green RE. (2016). Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Research*, 26, 342 - 350.
- Schartl M, Walter RB, Shen Y, Garcia T, Catchen J, Amores A, Braasch I, Chalopin D, Volff JN, Lesch KP, Bisazza A, Minx P, Hillier L, Wilson RK, Fuerstenberg S, Boore J, Searle S, Postlethwait JH, Warren WC. The genome of the platyfish, *Xiphophorus maculatus*, provides insights into evolutionary adaptation and several complex traits. *Nat Genet*. 2013 May;45(5):567–72.
- Shao C, Bao B, Xie Z, Chen X, Li B, Jia X, Yao Q, Ortí G, Li W, Li X, Hamre K, Xu J, Wang L, Chen F, Tian Y, Schreiber AM, Wang N, Wei F, Zhang J, Dong Z, Gao L, Gai J, Sakamoto T, Mo S, Chen W, Shi Q, Li H, Xiu Y, Li Y, Xu W, Shi Z, Zhang G, Power DM, Wang Q, Schartl M, Chen S. The

- genome and transcriptome of Japanese flounder provide insights into flatfish asymmetry. *Nat Genet.* 2017 Jan;49(1):119–124.
- Shin SC, Ahn DH, Kim SJ, Pyo CW, Lee H, Kim MK, Lee J, Lee JE, Detrich HW, Postlethwait JH, Edwards D, Lee SG, Lee JH, Park H. The genome sequence of the Antarctic bullhead notothen reveals evolutionary adaptations to a cold environment. *Genome Biol.* 2014 Sep 25;15(9):468.
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31, 3210 - 3212.
- Smith JJ, Timoshevskaya N, Ye C, Holt C, Keinath MC, Parker HJ, Cook ME, Hess JE, Narum SR, Lamanna F, Kaessmann H, Timoshevskiy VA, Waterbury CKM, Saraceno C, Wiedemann LM, Robb SMC, Baker C, Eichler EE, Hockman D, Sauka-Spengler T, Yandell M, Krumlauf R, Elgar G, Amemiya CT. The sea lamprey germline genome provides insights into programmed genome rearrangement and vertebrate evolution. *Nat Genet.* 2018 Feb;50(2):270–277.
- Star B, Nederbragt AJ, Jentoft S, Grimholt U, Malmstrøm M, Gregers TF, Rounge TB, Paulsen J, Solbakken MH, Sharma A, Wetten OF, Lanzén A, Winer R, Knight J, Vogel JH, Aken B, Andersen O, Lagesen K, Tooming-Klunderud A, Edvardsen RB, Tina KG, Espelund M, Nepal C, Previti C, Karlsen BO, Moum T, Skage M, Berg PR, Gjøen T, Kuhl H, Thorsen J, Malde K, Reinhardt R, Du L, Johansen SD, Searle S, Lien S, Nilsen F, Jonassen I, Omholt SW, Stenseth NC, Jakobsen KS. The genome sequence of Atlantic cod reveals a unique immune system. *Nature.* 2011 Aug 10;477(7363):207–10.
- Stamatakis, A. (2014). raxml version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30, 1312 - 1313.
- Tan, M.H., Austin, C.M., Hammer, M.P., Lee, Y.P., Croft, L.J., Gan, H.M., 2018. Finding Nemo: hybrid assembly with Oxford Nanopore and Illumina reads greatly improves the clownfish (*Amphiprion ocellaris*) genome assembly. *Gigascience* 7, 1–6.
- Tamura, K., Stecher, G., Peterson, D., Filipowski, A., & Kumar, S. (2013). mega6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, 30, 2725 - 2729.
- Tine M, Kuhl H, Gagnaire PA, Louro B, Desmarais E, Martins RS, Hecht J, Knaust F, Belkhir K, Klages S, Dieterich R, Stueber K, Piferrer F, Guinand B, Bierne N, Volckaert FA, Bargelloni L, Power DM, Bonhomme F, Canario AV, Reinhardt R. European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nat Commun.* 2014 Dec 23;5:5770.
- Valenzano DR, Benayoun BA, Singh PP, Zhang E, Etter PD, Hu CK, Clément-Ziza M, Willemsen D, Cui R, Harel I, Machado BE, Yee MC, Sharp SC, Bustamante CD, Beyer A, Johnson EA, Brunet A. The African Turquoise Killifish Genome Provides Insights into Evolution and Genetic Architecture of Lifespan. *Cell.* 2015 Dec 3;163(6):1539–54.
- Wang Y, Lu Y, Zhang Y, Ning Z, Li Y, Zhao Q, Lu H, Huang R, Xia X, Feng

- Q, Liang X, Liu K, Zhang L, Lu T, Huang T, Fan D, Weng Q, Zhu C, Lu Y, Li W, Wen Z, Zhou C, Tian Q, Kang X, Shi M, Zhang W, Jang S, Du F, He S, Liao L, Li Y, Gui B, He H, Ning Z, Yang C, He L, Luo L, Yang R, Luo Q, Liu X, Li S, Huang W, Xiao L, Lin H, Han B, Zhu Z. The draft genome of the grass carp (*Ctenopharyngodon idellus*) provides insights into its evolution and vegetarian adaptation. *Nat Genet.* 2015 Jun;47(6):625-31.
- Warren WC, García-Pérez R, Xu S, Lampert KP, Chalopin D, Stöck M, Loewe L, Lu Y, Kuderna L, Minx P, Montague MJ, Tomlinson C, Hillier LW, Murphy DN, Wang J, Wang Z, Garcia CM, Thomas GCW, Volff JN, Farias F, Aken B, Walter RB, Pruitt KD, Marques-Bonet T, Hahn MW, Kneitz S, Lynch M, Schartl M., 2018. Clonal polymorphism and high heterozygosity in the celibate genome of the Amazon molly. *Nat Ecol Evol.* 2, 669-679.
- Xu P, Zhang X, Wang X, Li J, Liu G, Kuang Y, Xu J, Zheng X, Ren L, Wang G, Zhang Y, Huo L, Zhao Z, Cao D, Lu C, Li C, Zhou Y, Liu Z, Fan Z, Shan G, Li X, Wu S, Song L, Hou G, Jiang Y, Jeney Z, Yu D, Wang L, Shao C, Song L, Sun J, Ji P, Wang J, Li Q, Xu L, Sun F, Feng J, Wang C, Wang S, Wang B, Li Y, Zhu Y, Xue W, Zhao L, Wang J, Gu Y, Lv W, Wu K, Xiao J, Wu J, Zhang Z, Yu J, Sun X. Genome sequence and genetic diversity of the common carp, *Cyprinus carpio*. *Nat Genet.* 2014 Nov;46(11):1212-9.
- Yang, Z. (2007). paml 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, 24, 1586 - 1591.
- You, X., Bian, C., Zan, Q., Xu, X., Liu, X., Chen, J., Wang, J., Qiu, Y., Li, W., Zhang, X., Sun, Y., Chen, S., Hong, W., Li, Y., Cheng, S., Fan, G., Shi, C., Liang, J., Tom Tang, Y., Yang, C., Ruan, Z., Bai, J., Peng, C., Mu, Q., Lu, J., Fan, M., Yang, S., Huang, Z., Jiang, X., Fang, X., Zhang, G., Zhang, Y., Polgar, G., Yu, H., Li, J., Liu, Z., Zhang, G., Ravi, V., Coon, S.L., Wang, J., Yang, H., Venkatesh, B., Wang, J., Shi, Q. Mudskipper genomes provide insights into the terrestrial adaptation of amphibious fishes. *Nat Commun.* 2014 Dec 2;5:5594.



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.