BSPE9921C-10706-3

볼락류(*Sebastes spp.*)의 체색관련 기능유전체 연구

Functional genomic analysis of pigment patterns in rockfishes (*Sebastes* spp.)

2015. 3

한국해양과학기술원

제출문

한국해양과학기술원장 귀하

본 보고서를 " 볼락류(Sebastes spp.)의 체색관련 기능유전체 연구"과제의 최종보고서로 제출합니다.

2015. 3.

총괄연구책임자 : 장요순

참여연구원 : 이은경, 황선완,
유재명, 명정구
오승용, 김종관
이정아, 장영석
권순열, 윤진영
오지나, 최해영
조혜경, 박서윤

보고서 초록

과제고유 번호	PE9921C	해당단계 연구기간	2014.09.01.～ 2014.12.31.	단계 구분	
연구사업명	중사업명	미래 해양생물자원 개발			
	세부사업명	볼락류(*Sebastes spp.*)의 체색관련 기능유전체 연구			
연구과제명	대과제명				
	세부과제명				
연구책임자	장요순	해당단계 참여연구원수	총 : 15 명 내부: 7 명 외부: 8 명	해당단계 연구비	정부: 229,659,000원 기업: 원 계 : 229,659,000원
		총연구기간 참여연구원수	총 : 15 명 내부: 7 명 외부: 8 명	총 연구비	정부: 229,659,000원 기업: 원 계 : 229,659,000원
연구기관명 및 소속부서명	동해연구소		참여기업명		
국제공동연구					
위탁연구					
요약(연구결과를 중심으로 개조식 500자 이내)				보고서 면수	83페이지

- 동해서식 볼락류 4종(우럭볼락, 흰꼬리볼락, 황볼락, 노랑볼락)의 유전자 정보를 분석하였고, 미토게놈 염기서열 정보를 GenBank에 등록하였음
- RNA-seq *de novo* assembly 방법으로 개볼락 근육조직 전사체 정보를 체색타입별로 조사하여 101,112개의 unigene을 확보하였고, 염기서열 및 아미노산 서열 정보를 뱅크화 하였음
- 개볼락 체색타입별 근육조직 전사체의 발현량을 분석하여 164개의 차등발현유전자를 확인하였고, 개볼락 color type 근육조직에서만 특이적으로 발현되는 51개의 유전자와 wild type 근육조직에서만 특이적으로 발현되는 68개 유전자를 확보하였음
- 개볼락 근육조직의 체색타입에 따른 차등발현유전자의 기능정보를 조사한 결과, 2개는 molecular function, 46개는 biological process, 6개는 cellular component 기능그룹에 해당하였음

색인어 (각 5개 이상)	한 글	볼락류, 미토게놈, 기능유전체, 전사유전체, 차등발현유전자
	영 어	*Sebastes*, Mitogenome, Transcriptome, RNA-seq analysis, Differentially expressed gene

요 약 문

Ⅰ. 제 목

볼락류(Sebastes spp.)의 체색관련 기능유전체 연구

Ⅱ. 연구개발의 필요성 및 목적

다양한 해양생명자원의 유전체 정보 해독은 신품종 개발과 식의약 소재 개발에 도움이 되고 식량안보 및 기후변화 등의 위기 대응에 도움이 되는 것으로 인식되고 있다. 해양생명자원의 유전체 정보를 분석하여 확보한 유용한 물질은 신약, 신소재, 바이오에너지 생산을 위한 자원으로 개발할 가치가 매우 높다.

방대한 양의 염기서열을 신속하게 분석할 수 있는 NGS(next generation technology) 기술 개발로 기존 유전체 연구의 패러다임이 기술이나 시간, 비용 측면에서 급격하게 변하였다. NGS 분석 방법으로 생산된 대량의 염기서열 데이터는 유전체 재분석 연구, 유전자 발현 연구, allele variation 및 association mapping 등에 적용할 수 있는 새로운 연구 플랫폼을 제공하고 있다. 유전체 분석의 가속화로 다수의 생물 유전체가 해독됨에 따라 많은 양의 새로운 유전체 정보, 이른바 빅데이터 처리를 위한 기술 및 유전체 정보 실용화 기술 개발이 필요하다.

본 연구를 통하여 해양어류 자원의 기능유전체 정보뱅크를 구축하고 미이용 어류자원의 유용자원화 및 고부가가치 산업화 연구에 활용하고자 한다. 또한, 양식이 가능한 수산생물의 생산성 향상 관련 분자유전학적 도구를 개발하고 생명자원의 유전체 정보 활용연구 기반을 마련하고자 한다.

Ⅲ. 연구개발의 내용 및 범위

1. 동해서식 볼락류의 유전자 정보 탐색
   - 미토게놈 정보 분석

2. 동해서식 볼락류의 체색관련 기능유전체 탐색
   - 개볼락 체색타입별 근육조직 전사체 정보 분석
   - 개볼락 체색타입별 차등발현유전자 분석

Ⅳ. 연구개발결과

1. 동해서식 볼락류의 미토게놈 정보 분석 및 Genbank 등록

- 동해서식 볼락류 4종(우럭볼락, 흰꼬리볼락, 황볼락, 노랑볼락)의 미토게놈을 분석하고 염기서열 정보를 NCBI (National Center for Biotechnology Information)의 GenBank에 등록하였다. 우럭볼락 미토게놈의 전체 크기는 16,453 bp (GenBank Accession No. KJ 525745), 흰꼬리볼락 미토게놈의 전체 크기는 16,445 bp (GenBank Accession No. KJ 834061), 황볼락 미토게놈의 전체 크기는 16,465 bp (GenBank Accession No. KJ 834063), 노랑볼락 미토게놈의 전체 크기는 16,450 bp (GenBank Accession No. KJ 834060)이었다.

2. 동해서식 볼락류의 체색관련 기능유전체 탐색 및 차등발현유전자 기능 분석

- 개볼락의 체색 타입별 근육조직 전사체를 RNA-seq 방법으로 분석하여 101,112개의 unigene을 확보하였다. BLASTx 검색과 InterProScan(v5) 검색으로 unigene의 기능을 예측하였다. 기능이 밝혀진 35,482개 유전자와 기능이 밝혀지지 않은 유전자 65,630개의 염기서열 정보 및 아미노산 서열을 확보하였다. 개볼락의 체색타입별 근육조직 전사유전체의 발현량을 측정하여 두 개의 근육조직 시료에서 발현량 차이를 나타내는 164개의

차등발현 유전자(DEG, differentially expressed gene)를 확보하였다. 개볼락 근육조직의 체색 타입에 따른 차등발현 유전자의 기능 정보를 파악하기 위하여 Gene Ontology (GO) enrichment 분석한 결과, 164개의 DEG가 1,613개의 GO ID에 해당함을 확인하였으며, DEG 2개는 molecular function, 46개는 biological process, 6개는 cellular component 기능그룹에 해당하였다.

V. 연구개발결과의 활용계획

본 연구를 통하여 확보한 볼락류 미토게놈 정보 및 개볼락의 전사유전자 정보는 유전자 기능분석을 통하여 어류의 체색관련 생명현상 규명에 활용할 계획이다. 또한, 어류의 기능유전체 정보를 이용하여 분자유전학적 도구를 개발하고 더 나아가 생물의 유전체 정보 자원을 산업화 할 수 있는 기술 개발을 위한 참고자료로 사용하고자 한다.

SUMMARY & KEYWORDS

A development of genome sequence Big-data application technology is needed for bio-industrial utilization of marine fish functional genome sequence. Rockfishes are members of the genus *Sebastes*. Currently, we know of about 110 rockfishes species worlwide. We analysed the complete mitochondrial genome of 4 rockfish species were determined using next-generation sequencing, including Sanger sequencing.

Hybrids of *Sebastes hubbsi*, considered a valid species, have been found in the ocean. The complete mitogenome of *S. hubbsis* was 16,453 bp (GenBank Accession No. KJ 525745) in length and comprised 13 protein-coding genes (PCGs), 2 rRNAs, 22 tRNAs and 1 control region. The complete mitochondrial DNA had 27.9% A, 26.7% T, 17.2% G and 28.3% C, showing a slight AT bias (54.5%). The start codon of all PCGs was ATG, except COX1 (GTG). The stop codons in the PCGs were mainly TAA, except ND1 and ND3 (TAG), ND4 (AGA) and Cytb (incomplete termination codon, T). The tRNAs had a typical cloverleaf structure, except tRNASer (AGY) which lacked the DHU arm.

Sebastes longispinis is considered to be a valid species despite its incomplete mitochondrial lineage. The complete sequence *S. longispinis* is 16,445 bp (GenBank Accession No. KJ 834061) in length and comprised of 13 protein-coding genes (PCGs), 2 rRNAs, 22 tRNAs and a control region. The complete mitogenome contains 28.0% A, 26.7% T, 17.1% G and 28.3% C nucleotides, with a slight AT bias (54.6%). The start codon of each PCG is ATG, except for COX1 (GTG). The stop codons found in the PCGs are mainly TAA, except for ND1 (TAG), ND3 (TAG), ND4 (AGA), and Cytb (incomplete termination codon, T). The tRNAs have a typical cloverleaf form,

except for tRNASer (AGY), which lacks the DHU arm.

Sebastes owstoni is popular fishery species in the northern western Pacific Ocean. The assembled mtDNA was 16,465 bp (GenBank Accession No. KJ 834063) in length, comprising 13 protein-coding genes (PCGs), 2 rRNAs, 22 tRNAs and 1 control region. The complete mitogenome comprised 27.7% A, 26.6% T, 17.3% G and 28.4% C, indicating a slight AT bias (54.3%). The start codon of all PCGs was ATG, with the exception of COX1 (GTG). The stop codons in the PCGs were mainly TAA, with the exception of ND1 and ND3 (TAG), ND4 (AGA) and Cytb (incomplete termination codon, T). All tRNAs had a typical cloverleaf shape, with the exception of tRNASer (AGY), which lacked the DHU arm.

The mitogenome of *S. steindachneri* was 16,450 bp (GenBank Accession No. KJ 834060) in length, and comprised of 13 protein-coding genes (PCGs), 2 rRNAs, 22 tRNAs, and a control region. The mitogenome is 27.4% A, 21.1% T, 17.5% G, and 28.0% C, showing a slight AT bias (54.5%). ATG was the start codon in all of the PCGs, except COX1 (GTG). The termination codon in the PCGs was TAA, except in ND1 and ND3 (TAG), ND4 (AGA), and Cytb (incomplete termination codon, T). All of the tRNAs had a typical cloverleaf shape, except tRNASer(AGY), which lacked the DHU arm. The sequenced mitogenome of *Sebastes* can be used in genetics-based management initiatives or for constructing the mitochondrial lineages within Scorpaenidae.

We constructed cDNAs from total RNA isolated from the skeletal muscles of two blass bloched rockfish (*S. pachycephalus*) genotypes : wild type and color type. We use next-generation sequencing (Illumina) RNAseq analyses to compare skeletal muscle transcriptome-wide expression levels in two distinct color pattern in blass bloched rockfish. Sequencing generated 156,145,718 raw reads and this number was reduced to 146,847,238 sequences after the cleaning pipeline was implemented. The latter sequences were used to build the de novo assembly. Trinityt produced a high number of transcripts that were

clustered by TGICL into 101,112 unigenes. Using a combination of differential expression analyses (DE) we identified 164 candidate genes that showed DE between the color morphs. We assigned Gene Ontology (GO) terms to the 1,613 annotated loci. Fourty-six 'Biological Process' categories were assigned among 1,191 blass bloched rockfish loci; Six 'Cellular Component' categories were assigned among 192 blass bloched rockfish loci; and Two 'Molecular Function' categories were assigned among 230 blass bloched rockfish loci. Our study provides valuable transcriptome sequence data for *S. pachycephalus.*

Our transcriptomic sequences will enhance the quality of gene annotation and functional analysis of the blass bloched rockfish genome and serve as a material basis for future genomic research on marine fishes. Using transcriptomic analyses we successfully identified key expression differences between different color morphs of blass bloched rockfish. These differentially expressed genes have important implications for our understanding of the molecular mechanisms underlying pigmentaion. The sequencing data from this study can be used to develop molecular markers and to identify the extreme differences among the phenotypes of different species in the genus *Sebastes*.

KEYWORDS : 볼락류, 미토게놈, 기능유전체, 전사유전체, 차등발현유전자
Rockfish, Mitogenome, Functional genome, Transcriptome,
Differentially expressed gene

목 차

표목차 ······ 9

그림목차 ······ 10

제 1장 서 론 ······ 12

제 1절 연구 필요성 ······ 13

제 2절 연구 목표 ······ 14

제 3절 연구내용 및 범위 ······ 15

제 2장 국내외 기술개발 현황 ······ 17

제 1절 국내 기술개발 현황 ······ 18

제 2절 국외 기술개발 현황 ······ 18

제 3장 연구개발 수행내용 및 결과 ······ 20

제 1절 유용 해양어류 유전자원 정보 탐색 ······ 21

제 2절 해양어류자원의 기능유전체 정보 이용기술 개발 탐색 ······ 51

제 4장 연구개발 목표 달성도 및 대외기여도 ········· 75

제 1절 연구개발 목표 달성도 ········· 76

제 2절 대외기여도 ········· 76

제 5장 연구개발결과의 활용계획 ········· 79

제 6장 참고문헌 ········· 80

List of Tables

Table 1 Summary of the mitochondrial genome of *S. hubbsi*

Table 2. Summary of the mitochondrial genome of *S. longispinis*

Table 3 Summary of the mitochondrial genome of *S. owstoni*

Table 4. Summary of the mitochondrial genome of *S. steindachneri*

Table 6. Unigenes transcripts detected in this study

Table 7. List of color pattern type-specific differentially expressed genes

Table 8. Known genes of color-type skeletal muscle specific expressed genes

Table 9. Known genes of wild color-type skeletal muscle specific expressed genes

Table 10. Molecular function group of the differentially expressed genes

List of Figures

Figure 1. Sequences of the mitogenome of *S. hubbsi* (GenBank Accession No. KJ525745)

Figure 2. A gene arrangement in the complete mitogenome *Sebastes hubbsi*.

Figure 3. Sequences of the mitogenome of *S. longispins* (GenBank Accession No. KJ834061)

Figure 4. A gene arrangement in the complete mitogenome *Sebastes longispinis*.

Figure 5. Sequences of the mitogenome of *S. owstoni* (GenBank Accession No. KJ834063)

Figure 6. A gene arrangement in the complete mitogenome *Sebastes owstoni*.

Figure 7. Sequences of the mitogenome of S. *steindachneri* (GenBank Accession No. KJ834060)

Figure 8. A gene arrangement in the complete mitogenome *Sebastes steindachneri*.

Figure 9. Workflow of amplicon library construction and Hiseq2500 sequencing.

Figure 10. The skeletal muscles of *S. pachycephalus* for RNA-seq analysis.

Figure 11. The result of quality assessment of total RNA.

Figure 11. The result of quality assessment of total RNA.

Figure 12. Results of the blass bloched rockfish (*S. pachycephalus*) de novo assemly

Figure 13. Frequancy of the blass bloched rockfish unigenes BLAST annotation

Figure 14. Predict of CDS of unigenes by InterProScan

Figure 15. Annotation consensus betwwen Blastx and InterProScan of unigenes

Figure 16. DNA sequences of unigene transcripts

Figure 17. Amino acid sequences of unigene transcripts

Figure 18. Expressiom levels of unigene transcripts

Figure 19. Differentially expressed genes (DEG) analysis of unigene transcripts

Figure 20. Histogram of the Gene Ontology classification

제 1장 서 론

제1절 연구 필요성

생명공학 관련 기술의 발달로 생물의 유전체 정보를 신속하게 대량으로 생산할 수 있게 됨에 따라 유전자의 기능과 생명현상 규명 등 유전체 정보 활용 연구가 활발히 진행되고 있다. 유전체 데이터 분석 및 이용기술 개발 연구의 필요성 증가에 따라 다양한 생물의 유전체 정보를 활용하기 위한 맞춤형 데이터베이스가 구축되고 있으며 생명체가 가진 유전체 정보는 매우 유용한 연구자산으로 개별 유전자의 기능 및 유전자의 상호 네트워크를 이해하기 위한 유전체 정보 활용기술 개발이 요구되고 있다. 방대한 양의 염기서열을 신속하게 분석할 수 있는 NGS(next generation technology) 기술 개발로 기존 유전체 연구의 패러다임이 기술이나 시간, 비용 측면에서 급격하게 변하였고, NGS 분석 방법으로 생산된 대량의 염기서열 데이터는 유전체 재분석 연구, 유전자 발현 연구, allele variation 및 association mapping 등에 적용할 수 있는 새로운 연구 플랫폼을 제공하고 있다. 유전체 분석의 가속화로 다수의 생물 유전체가 해독 되었으며, 방대한 양의 새로운 유전체 정보, 이른바 빅데이터를 처리하기 위한 기술 및 유전체 정보 실용화 기술 개발이 시급하다.

해양생명자원의 유전체 정보 분석 및 이용기술 개발을 통하여 유용한 물질 및 기능을 활용한 신약, 신소재, 바이오에너지 등을 생산하는 해양바이오산업 육성을 위한 기반이 마련되어야 한다. 맥킨지(2013년 보고서)는 유전체학을 인류의 삶을 근본적으로 바꿀 미래기술로 분류하였으며, 2025년까지 글로벌 경제 파급효과를 1.6조 달러로 추정하였다. 해외 선진국들은 유전체 정보 활용기술의 막대한 경제효과를 인식하고 유전체 산업 자체가 미래 성장동력이 될 것으로 보고 치열한 경쟁을 벌이고 있다. 다양한 해양 생명자원의 유전체 정보 해독은 신품종 개발과 식의약 소재 개발에 도움이 되고 식량안보 및 기후변화 등의 위기 대응에 도움이 되는 것으로 인식되고 있다.

NGS 유전체 해독 기술 등 생명공학기술의 발전으로 다양한 생물의 유전체 정보 및 오믹스 정보가 축적되었고, 이들 정보를 활용하기 위한 데이터베이스 및 프로그램이 구축되고 있다. 『포스트 게놈 신산업 육성을 위한 다부처 유전체 연구사업』(보건복지부, 농림축산식품부, 해양수산부, 미래창조과학부, 산업통상자원부, 농촌진흥청 /

2014년~2021년)이 ‘유전체 연구를 통한 국민건강 증진과 산업화 연계시스템 구현’이라는 비젼으로 동·식물, 해양생물등 유전정보를 활용한 고부가가치 생명자원을 개발하기 위한 프로젝트가 시작되었다. NGS 기술 개발로 생물의 유전체 정보가 매우 빠른 속도로 축적되고 있으나 유용한 연구자산인 유전체 정보를 효율적으로 활용할 수 있는 준비가 부족하다. 매우 빠른 속도로 발전하고 있는 생명공학은 다양한 분야와 융복합하여 많은 새로운 첨단기술이 개발되고 있으나 생명공학 연구의 활용은 매우 미흡한 실정이다.

제2절 연구 목표

해양어류 자원을 효율적으로 활용하기 위하여 볼락류(*Sebastes spp.*)를 중심으로 기능유전체 정보뱅크를 구축한 후, 유전체 빅데이터 정보 데이터베이스를 융복합한 기능유전체 이용기술을 개발하고자 한다. 볼락류는 전 세계에 걸쳐 약 110종이 분포하고 있으며, 우리나라에는 19종이 출현하고 있다. 체색이나 생태 및 행동이 매우 다양하게 나타나는 특성을 가진 볼락류는 특정 해역에서만 출현하는 지리학적인 분포 특성을 가진 종(species)도 있다. 본 연구사업에서는 우리나라에 서식하는 볼락류를 중심으로 유전특성을 조사하고 기능유전체를 분석하여 어류자원을 효율적으로 활용하기 위한 기능유전체 정보뱅크를 구축할 계획이다. 또한, 볼락류 기능유전체 연구 결과는 미이용 어류자원의 유용자원화 및 고부가가치 산업화 연구에 활용하고자 하며, 양식이 가능한 수산생물의 생산성 향상 관련 분자유전학적 도구를 개발하고 생명자원의 유전체 정보 활용연구 기반을 마련하고자 한다.

제3절 연구내용 및 범위

구 분	목 표	내 용 및 범 위
1차년도 (2014)	○ 유용 해양어류 유전자원 정보 탐색	
	● 유용 어류자원 탐색 및 볼락류 유전자 정보 탐색	● 동해서식 유용 어류자원 조사 - 동해서식 볼락류의 자원특성 조사 ● 동해서식 볼락류의 유전특성 조사 - 동해서식 볼락류의 유전정보 확보 - 개볼락 시료의 유전특성 조사
	○ 해양어류자원 기능유전체 이용기술 탐색	
	● 볼락류 기능유전체 이용기술 탐색	● 동해서식 볼락류의 기능유전체 정보 탐색 ● 개볼락의 체색관련 유전특성 분석

구 분	목 표	내 용 및 범 위
2차년도 (2015)	○ 유용 해양어류 유전자원 확보	
	● 유용 어류자원 탐색 및 볼락류 유전자 정보 탐색	● 동해서식 볼락류 시료 확보 - 동해 고유서식 볼락류 확보 - 체색 및 spot 패턴이 다양한 개볼락 확보 ● 동해 고유서식 볼락류의 유전특성 조사 - 개체 또는 집단수준에서 유전자 변이 분석 - 동해 고유서식 볼락류 미토게놈 분석 ● 체색 및 spot 패턴이 다양한 개볼락의 유전특성 조사 - 개볼락의 미토게놈 분석
	○ 해양어류자원 기능유전체 이용기술 탐색	
	● 볼락류 기능유전체 이용기술 탐색	● 동해서식 볼락류 기능유전체 정보뱅크 구축 ● 개볼락의 체색 관련 기능유전자 확보 및 정보 분석 ● 개볼락의 다양한 체색과 기능유전자 발현과의 상관성 조사 ● 개볼락 기능유전체 활용방안 탐색

구 분	목 표	내 용 및 범 위
3차년도 (2016)	○ 유용 해양어류 유전자원의 활용 연구	
	• 볼락류 유전자 정보 활용 연구	• 볼락류의 유전특성 비교분석 • 양식이 가능한 수산생물의 생산성 향상 관련 유전자마커 개발
	○ 해양어류자원 기능유전체 이용기술 개발	
	• 볼락류 기능유전체 이용기술 개발	• 개볼락 기능유전체를 활용한 유전자 마커 개발 • 볼락류 기능유전체 이용기술 개발 - 분자유전학적 도구 개발 / 적용기술 개발 - 바이오산업소재화 가능성 탐색

제 2 장 국내외 기술개발 현황

제1절 국내 기술개발 현황

국내 해양생물 유전체 연구는 해양생명공학사업으로「해양극한생물분자유전체 연구단」에서 유전체 연구를 추진(2004년～2013년)하였으며, 해양극한생물을 대상으로 유전자 및 단백질 수준에서 생명기능과 현상을 해석하고 유용유전자 및 단백질을 발굴하여 활용하기 위한 기술 개발 연구가 진행된 바 있다. 이 연구사업을 통하여 밍크고래 전장유전체가 해독되었고, 해양 동식물 12종과 해양미생물 43건의 유전체 정보가 확보되었다. 국립수산과학원에서는「양식생물 및 수산생물의 육종(분자육종 포함) 연구」가 넙치, 전복, 돌돔, 멍게를 대상으로 2004년터 2013년까지 10년간 진행되었다. 이 연구사업을 통하여 넙치 유전체를 완전해독하고 유전자 지도를 작성하였으나, 유전자 정보 활용 연구 결과는 없는 상태이다.

국내에서 진행되고 있는 어류의 기능유전체 연구는 제주대학교에서 까막전복과 돌돔을 대상으로 수행하였으며, 한국해양과학기술원에서는 강도다리와 볼락을 대상으로 근육성장과 관련된 기능유전체를 분석하고 기능유전자를 이용한 어류의 생산성 향상 관련 유전자마커 개발 연구를 수행하였다.

제2절 국외 기술개발 현황

해양생명자원의 유전체 연구는 인간유전체를 비롯하여 육상생명자원을 대상으로 다른 바이오 분야에 비하여 늦게 시작되었으나, 최근 미국 등 선진국을 중심으로 대규모 연구가 진행되고 있다.

미국은「21세기 해양청사진」과 같은 해양에 대한 지속적인 정책을 선언하고 세계 해양바이오에 대한 연구를 주도하며 다양한 해양시료 수집 활동 및 해양생물종 유전체 연구를 활발하게 진행하고 있다.

일본에서는 산학연 협력 연구활성화를 통하여 해양유전체 연구를 추진하고 있으며, 문부과학성, 국토 교통성 등 부처 공동으로 해양유전체 연구 분야 지원을 확대하였고, 해양과학기술연구센터(JAMSTEC)와 해양생명공학연구소(MBI), 수산종합연구센터(FRA)를 중심으로 유전체 연구에 많은 예산을 투자하고 있다.

중국에서는 해양환경변화와 해양생물자원의 지속가능한 이용 및 대양 심해 연구에 대한 국가 지원을 강화하고 중국과학기술부를 중심으로 서해, 동중국해, 남중국해에 서식하는 어류, 원핵생물, 식물, 부유동물, 무척추동물 등 주요 해양생물 2천종의 유전자 정보를 담은 DNA 바코드를 데이터베이스화하는 사업인 '중국 근해 해양생물 DNA 바코드 DB구축 프로젝트'를 착수하였다.

프랑스에서는 로스콥해양연구소와 프랑스 국립시컨싱센터(Genoscope)가 중심이 되어 해조류 유전체 분석연구를 진행하였고, 캐나다에서는 수산해양부(DFP, Department of Fisheries and Oceans)를 중심으로 해양생명공학 분야에서 유전체학 및 생명공학 발전계획을 수립하였으며, 대서양연어 유전자지도 작성 등 연어 유전체 연구를 수행하였다.

제 3 장 연구개발 수행내용 및 결과

제1절 유용 해양어류 유전자원 정보 탐색

1. 연구 재료 및 방법

가. 어류시료 확보

동해에 서식하는 볼락류 시료를 확보하였으며, 임원항과 죽변항 일대에서 개볼락, 우럭볼락, 흰꼬리볼락, 황볼락, 노랑볼락을 구입하였다. 유전자 분석시료는 근육조직 및 지느러미 조직을 이용하였으며 절단하여 냉동 보관하였다.

나. 볼락류 어류의 genomic DNA 분리 및 정제

우럭볼락, 흰꼬리볼락, 황볼락, 노랑볼락의 꼬리지느러미를 0.5㎝ X 0.5㎝(가로X세로) 크기로 잘라서 lysis buffer[10mM Tris-HCl pH7.5, 125mM NaCl, 10mM EDTA, 0.5% SDS, 5M Urea, 0.1㎎/㎖ proteinase K]를 첨가하여 충분히 용해한 후, binding buffer와 isopropanol을 사용하여 genomic DNA를 분리하였고, *Accuprep*® Genomic DNA Extraction Kit(Bioneer Co., Korea)의 column을 이용하여 정제하였다. 분리・정제한 genomic DNA는 전기영동으로 확인한 후, NanoDrop® ND-1000 Spectrophotometer(NanoDrop Technologies, USA)를 사용하여 농도를 측정하고 -80℃에 보관하였다.

다. 미토게놈 정보 분석 및 GenBank 등록

분리・정제한 genomic DNA의 순도를 확인한 후 HiSeq2000(Illumina Inc.)을 이용하여 염기서열을 분석하였다. 볼락류의 미토게놈 정보를 SOAPdenovo2 software로 어셈블리 한 후, 빠진 영역의 염기서열은 PCR 증폭산물을 Sanger sequencing하여 확보하였다. 어셈블리가 끝난 미토게놈 정보는 DOGMA와 tRNA-SE를 활용하여 annotation 하여 GenBank에 등록하였다.

2. 연구결과

가. 동해서식 볼락류의 자원특성 조사 및 시료 확보

볼락류(*Sebastes* spp.)는 전 세계에 걸쳐 약 110종이 분포하고 있으며, 우리나라에는 19종이 출현하고 있다. 체색이나 생태 및 행동이 매우 다양하게 나타나는 특성을 가진 볼락류는 특정 해역에서만 출현하는 지리학적인 분포특성을 가진 종(species)도 있다. 볼락류는 대부분 난태생으로 연안의 암초가 많은 수심 20～170m 이내 지역에 서식하며, 돌삼뱅이(*Sebastes baramenuke*), 황볼락(*Sebastes owstoni*), 노랑볼락(*Sebastes steindachneri*) 등 일부 종은 100～400m 이내의 다소 깊은 바다에 서식한다. 동해안에는 17종의 볼락류가 출현하는 것으로 알려져 있으며, 돌삼뱅이, 황볼락, 노랑볼락, 탁자볼락(*S. taczanowskii*), 말락볼락(*S. wakiyai*), 좀볼락(*S. minor*)은 동해에서만 서식하는 것으로 보고되었다. 체색과 관련된 기능유전체를 분석하기 위하여 다양한 체색 패턴을 나타내는 개볼락(*S. pachycephalus*), 우럭볼락(*S. hubbs*i) 및 흰꼬리볼락(*S. longispinis*)을 확보하였다. 또한, 동해 중부 이북에 서식하는 황볼락(*S. pachycephalus*), 노랑볼락(*S. pachycephalus*), 탁자볼락(*S. taczanowskii*), 세줄볼락(*S. trivittatus*), 좀볼락을 확보하여 유전정보 분석을 위한 시료로 활용하였다.

나. 동해서식 볼락류의 미토게놈 정보 분석 및 GenBank 등록

우럭볼락 미토콘드리아 DNA의 전체 염기서열(미토게놈) 정보를 분석하여 세계유전자은행인 NCBI의 GenBank (http://www.ncbi.nlm.nih.gov/genbank/)에 등록(Accession number KJ525745)하였다(Figure 1). 우럭볼락의 미토게놈은 13개의 단백질 암호화 영역, 2개의 리보좀, 22개의 tRNA 및 1개의 조절영역으로 이루어져 있었고, 전체 크기는 16,453 bp 이었다. 우럭볼락의 미토콘드리아 DNA의 염기 조성비는 27.9% (A), 26.7% (T), 17.2% (G) 및 28.3% (C)이었고, AT 비율이 약간 높았으며(54.6%), COX1 (GTG)을 제외한 모든 단백질 암호영역의 개시코돈은 ATG이었고, 종결코돈은 TAA가 주를 이루었으나 ND1과 ND3는 TAG, ND4는 AGA였으며 Cytb는 종결코돈이 불완전(T)하였다(Table 1). tRNA형태는 tRNA$^{Ser(AGY)}$(lacked DHU-arm 구조)을 제외하고 모두 전형적인 cloverleaf 구조로 되어 있었다(Figure 2).

```
   1 GCTAGCGTAG CTTAATTAAA GCATAACACT GAAGATGTTA AGATGGACCC TAGAAAGTCC
  61 CGCCCGCACA AAGGCTTGGT CCTGACTTTA CTATCAACTT TAGCCAAATT TACACATGCA
 121 AGTATCCGCC CCCCTGTGAG AATGCCCTAC AGCTCCCTGC CCGGGAGCAA GGAGCTGGTA
 181 TCAGGCACAC ATCTGTAAGC CCATGACACC TTGCTTAGCC ACACCCTCAA GGGAACTCAG
 241 CAGTGATAAA CATTAAGCCA TAAGTGAAAA CTTGACTTAG TTAAAGCTAA GAGGGCCGGT
 301 AAACCTCGTG CCAGCCACCG CGGCTATACG AGAGACCCAA GTTGATACCA TTCGGCGTAA
 361 AGAGTGGTTA TGGAAAATAA AGACTAAAGC CGCACACCTT CAAAGCTGTT ATACGCATCC
 421 GAAGGCTAGA AGATCAACCA CGAAGGTAGC TTTACAACCC CTGACCCCAC GAAAGCTCTG
 481 GCACAAACTG GGATTAGATA CCCCACTATG CCTAGCCCTA AACCTTGGTA ATATATCACA
 541 TACCCTGCCC GCCTGGGAAC TACGAGCACC AGCTTAAAAC CCAAAGGACT TGGCGGTGCT
 601 TTAGACCCCC CTAGAGGAGC CTGTTCTAGA ACCGATAACC CCCGTTCAAC CTCACCCTTC
 661 CTTGTTTATC CCGCCTATAT ACCGCCGTCG TCAGCTTACC CTGTGAAGGA CTAAAAGTAA
 721 GCACAACTGG CACAACCCAA AACGTCAGGT CGAGGTGTAG CGCATGGAAG GGGAAGAAAT
 781 GGGCTACATT CCCTACACTA GGGAACACGA ACGGCGCACT GAAACACGCG CCTGAAGGAG
 841 GATTTAGTAG TAAGCGGAAA ATAGCGTGTT CCGCTGAAAT CGGCCCTGAA GCGCGCACAC
 901 ACCGCCCGTC ACTCTCCCCA AGCCTACCAC TTTAAGTAAT TAAAAACCCA AAAATCGCAG
 961 AGGGGAGGCA AGTCGTAACA TGGTAAGGGT ACCGGAAGGT GCACTTGGTA ATATCAGAGT
1021 GTAGTTAAAA TAGAATAACA CTTCCCTTAC ACTGAAGAGA CATCCGTGCA AATCGGATCA
1081 CCCTGATGCC CAACAGCTAG CCCACAAACA CAACAACAAC CAACCATTAT TTATAACCCC
1141 AAATGCACGA ATGTTTCAAT TAAACAAACC ATTTTTCCCC TTTAGTATGG GCGACAGAAA
1201 AAGGACTTAG GAGCAATAGA GAAAGTACCG CAGGGGATCG CTGAAAGAGA AATGAAACAA
1261 CCCAGTGAAG CTAAGTAAAG CAGAGATTTA TTCTCGTACC TTTTGCATCA TGATTTAGCC
1321 AGCGTGACCC AAGCAAAGAG TGCTTTAGTT TGACACCCCG AAACTAGGGG AGCTACTCCA
1381 AGACAGCCTA TTTATAGGGC GAACCCGTCT CTGTGGCAAA AGAGTGGAAT GAGCTTTGAG
1441 TAGAGGTGAT AAACCTACCG AACCTAGTTA TAGCTGGTTG CCCGAGAAAT GGATAGAAGT
1501 TCAGCCTCTC AGATTCTTTA TTCACCTCAG TATTACCCCA CCTGATACCA CAAGAAACTG
1561 TGAGAGTTAT TCAAAGGGGG TACAGCCCCT TTGAAACAAG ATACAACTTT TCCGGGAGGA
1621 AAAAGATCAT AATTAAATAA AGGTAAGTAT TTGGGTGGGC CTAAAAGCAG CCACCCCAGT
1681 AGAAAGCGTT ATAGCTCAAA TACATCACTA CCCCTCTCTA TCCTGATCAT TAATTCTTAC
1741 TCCCCCCTTC CCTACCGGGC CATCCCATGC ACACATGGGA GAGACCCTGC TAATATGAGT
1801 AATAAGAGAG CCAAGCCTCT CTCCTCGCAC ACATGTAATT CGGAACGAAC CCGCACCGAG
1861 CATTAACGGC CCCAAACGAA GAGGGACCTG AACAACAACC CAAACAACCA GAAAAAAATT
1921 CAAACATAAA CCGTTAACCC TACACAGGTG TGCATTTAAG GAAAGACTAA GAGAAAGAGA
1981 AGGAACTCGG CAAACAAATC AAGCCTCGCC TGTTTACCAA AAACATCGCC TCTTGCAAAG
2041 CTAAAGAATA AGAGGTCCCG CCTGCCCTGT GACTATTAGT TTAACGGCCG CGGTATTTTG
2101 ACCGTGCAAA GGTAGCGCAA TCACTTGTCT TTTAAATGAA GACCTGTATG AATGGCACAA
2161 CGAGGGCTTA ACTGTCTCCT CTTTCAAGTC AATGAAATTG ATCTCCCCGT GCAGAAGCGG
2221 GGATATAAAC ATAAGACGAG AAGACCCTAT GGAGCTTTAG ACACCAAAGA AGATCCTGTC
2281 AAGTAACCCT CTATAAAGGC CTGAACTAAT GGAACCCTTC CCTAATGTCT TTGGTTGGGG
2341 CGACCGCGGG GAAACAAAAA ACCCCCACGT GGAAAGGGAG CACCCCCTCC TACAACTAAG
2401 AGCCGCAGCT CTAATTAACA GAATATCTGA CCAATAAGAT CCGGCAATGC CGATCAACGT
2461 ACCGAGTTAC CCTAGGGATA ACAGCGCAAT CCCCTTTTAG AGCCCATATC GACAAGGGGG
2521 TTTACGACCT CGATGTTGGA TCAGGACATC CTAATGGTGC AGCCGCTATT AAGGGTCCGT
2581 TTGTTCAACG GTTAAAGTCC TACGTGATCT GAGTTCAGAC CGGAGTAATC CAGGTCAGTT
2641 TCTATCTATG GTGTGCTCTT TTTTAGTACG AAAGGACCGT AAAGAAGAGG CCCCTGCTTT
2701 AAGCAAGCCT CACCCCCACC TAGTGAAAAC AACTAAAGTA GGCAAGAGGG CATACCCCCA
2761 ATGCCTGAGA GAACGGCATG TTGGGGTGGC AGAGCCCGGT GAATGCAAAA GACCTAAGCC
2821 CTTTTTACAG AGGTTCAAGT CCTCTCCTTA ACTATGATTT CAGTCCTTAT TACCCATATT
2881 CTCAATCCCT TGGCCTTCAT TGTCCCCATC CTCTTAGCCG TCGCCTTCCT CACACTTTTA
2941 GAACGTAAGG TACTAGGGTA TATACAACTA CGAAAAGGCC CAAATATTGT AGGGCCTTAC
3001 GGACTATTAC AGCCTATCGC TGACGGTGTG AAGCTCTTTA TTAAAGAGCC CGTTCGCCCC
3061 TCCACTTCTT CCCCAGTACT ATTCCTCCTC GCCCCACTAC TCGCACTCAC ACTTGCCTTA
3121 ACCCTTTGAG CCCCTATACC TCTCCCATAC CCAGTAATTG ACTTAAACCT TGGAATCCTA
3181 TTTATTTTGG CCCTATCAAG CCTCGCTGTC TACTCCATTC TAGGCTCAGG TTGAGCATCA
3241 AATTCAAAAT ACGCTCTCAT CGGGGCCCTT CGGGCTGTAG CCCAAACCAT TTCATATGAA
3301 GTTAGTCTAG GCTTAATCCT ATTAAGTACC ATTATCTTTA CAGGAGGTTT CACACTACAA
3361 ACCTTCAACA TTGCCCAAGA AAGCGTCTGA ATACTACTCC CAGCCTGACC ACTAGCCGCA
3421 ATGTGATATA TTTCAACCCT TGCAGAGACA AACCGTGCAC CCTTTGACCT TACTGAAGGC
3481 GAATCCGAAC TAGTCTCTGG CTTCAATGTC GAGTATGCAG GTGGCCCATT CGCCCTATTT
```

Figure 1. Sequences of the mitogenome of *S. hubbsi* (GenBank Accession No. KJ525745).

```
3541 TTCCTGGCCG AATATGCTAA CATTCTGCTT ATAAACACGC TTTCCGCTAC CCTCTTTTTA
3601 GGGGCCTCCC ATTTTCCAAT GCTACCTGAA CTCACCGCAG TGAACCTAAT AACCAAAGCA
3661 GCCCTTCTGT CCGTCTTATT TCTATGAGTC CGAGCCTCTT ACCCACGATT CCGGTACGAC
3721 CAACTTATAC ATCTAATTTG AAAAAACTTC CTCCCACTTA CACTAGCCCT GGTTATCTGA
3781 CACCTAGCCC TCCCAATTGC ATTTGCTGGC CTGCCACCCC AGCTATAGAT AAGAAGCCGT
3841 GCCTGAAGTA AAGGGCCACT TTGATAGAGT GACTTATGGG GGTTCAAATC CCCCCCGCTT
3901 CTTAAAAAAG GGGGACTCGA ACCCCGCCTA AGGAGAGCAA AACTCCTGGT GCTCCCACTA
3961 CACTATTTCC TAGTAAAGTC AGCTAATTCT AAGCTCTTGG TCCCATACCC CAAACACGAA
4021 GGTTAAAATC CCTCCTTTAC TAATGAACCC TTACATCTTA ACCGCCCTGC TATTTGGTAT
4081 TGGTTTAGGC ACTACTACCA CCTTCGCAAG CTCCCACTGA TTACTAGCCT GAATAGGCCT
4141 GGAAATAAAT ACTCTTGCCA TCATTCCCCT AATAGCTCAA CACCATCACC CCCGAGCAGT
4201 TGAAGCAGCC ACAAAATATT TCTTAATTCA AGCTGCAGGA GCAGCCATAC TACTCTTTGC
4261 CAGCACCACC AACGCTTGAT TAACTGGACA ATGGGACCTT TTACAAATTG CCCACCCTTT
4321 CCCAACTGCC CTTGTCACTT TGGCCCTCGC ACTAAAAGTG GGACTTGCAC CTGTGCACTC
4381 ATGACTGCCT GAAGTACTTC AGGGCCTAGA CCTAACCACA GGACTTATTT TGTCTACCTG
4441 GCAAAAACTT GCCCCATTTG CCTTATTAGT CCAAACCCCC TGTGCCAACA CCACCCTTCT
4501 AATCATTCTA GGACTCACCT CAACCATTGT AGGAGGCTGA GGAGGCCTTA ACCAAACCCA
4561 ACTTCGCAAA ATTCTTGCTT ACTCCTCCAT CGCACACCTA GGCTGAATAG TAATTGTGCT
4621 ACAATTCTCC CCCTCCTTGA CTATTTTAAC ATTATTCACA TACTTCATTA TAACATTCTC
4681 AGCATTTCTT ATGTTTAAAC TTAATAAAGC AACCAACATT AATACTCTAG CAACCTCATG
4741 AGCAAAAACC CCCGCCCTAA CGGCCCTTGC ACCCCTTTTA TTATTATCCT TAGGGGGACT
4801 CCCACCACTT ACAGGCTTTA TGCCAAAATG ACTTATCCTT CAAGAACTTA CTAAGCAAGA
4861 CCTCGCCCCA GCTGCAACAC TAGCAGCAAT AACCGCCCTC CTCAGCCTAT ATTTTTACCT
4921 ACGACTATCA TACGCAATGG CATTAACTAT TTCACCCAAC AACCTCACCG CAATTTCCCC
4981 ATGACGCCTC CCCTCCTTAC AACTAACACT GCCACTTGCT ACCTCAGCCA TAGCTACGCT
5041 ACTGCTTCTA CCCCTAACAC CCGCCGCAAT AGCACTAATA ACCCTTTAAG GGACTTAGGT
5101 TAAAACAAGA CCAAGGGCCT TCAAAGCCCT AAGTGAGGGT GGAAGTCCCC CAGTCCCTGA
5161 TAAGGCTTGC GGGACACTAC CCCACATCTC CTGTATGCAA AACAGGTACT TTAATTAAGC
5221 TAAAGCCTTC CTAGAAGGGC AGGCCTCGAT CCTGCAAGAT CTTAGTTAAC AGCTAAGCGC
5281 TCAAACCAGC GAGCATCCAT CTACCTTTCC CCCGCCTAAA AAGCGGGCTG AAGGCGGGGG
5341 AAAGTCCCGG CAGACGACTA ACCTGCATCT TCAGATTTGC AATCTGATAT GTATAACACC
5401 TCAAGACTTC TGGTAAGAAG AGGATTCAAA CCTCTGTTTG TGGGGCTACA ATCCATCGCT
5461 TAGAAACTCA GCCATCCTAC CTGTGGCCAT CACACGTTGA TTTTTCTCCA CTAATCACAA
5521 AGACATCGGC ACCCTTTATC TAGTATTTGG TGCCTGAGCC GGTATAGTAG GCACAGCCCT
5581 CAGCCTACTC ATTCGAGCAG AACTAAGCCA ACCGGGCGCT CTCCTTGGAG ACGACCAAAT
5641 TTATAATGTA ATCGTTACGG CACATGCCTT CGTAATGATT TTCTTTATAG TAATGCCAAT
5701 TATAATCGGA GGTTTTGGAA ACTGATTAAT TCCCCTAATG ATTGGAGCCC CAGATATAGC
5761 ATTTCCTCGT ATAAATAATA TAAGTTTCTG ACTTCTTCCC CCTTCTTTTC TGCTACTACT
5821 TGCCTCCTCT GGAGTAGAGG CAGGCGCCGG GACCGGCTGA ACAGTGTACC CGCCCCTGGC
5881 TGGCAACTTA GCCCACGCAG GAGCATCAGT CGACCTAACA ATCTTTTCAC TCCACTTAGC
5941 AGGTATCTCC TCAATTCTTG GGGCAATCAA TTTTATTACC ACAATTATTA ATATGAAGCC
6001 TCCGGCCATC TCCCAATACC AGACGCCCCT ATTTGTGTGA GCCGTCTTAA TTACCGCTGT
6061 TCTTCTCCTT CTCTCTCTAC CAGTTCTCGC TGCAGGTATT ACAATGCTCC TTACGGACCG
6121 AAACCTTAAC ACCACCTTCT TCGACCCGGC CGGAGGGGGA GATCCAATCC TTTACCAACA
6181 CCTATTCTGG TTTTTTGGGC ACCCGGAGGT ATATATTCTT ATTCTGCCTG GCTTTGGTAT
6241 GATTTCACAT ATTGTCGCCT ATTACTCTGG CAAAAAAGAA CCCTTTGGAT ATATAGGCAT
6301 GGTCTGAGCA ATAATGGCTA TTGGCCTCCT AGGATTTATT GTATGGGCTC ATCACATATT
6361 CACAGTTGGT ATGGACGTAG ACACACGTGC TTACTTTACG TCTGCCACAA TAATCATCGC
6421 AATTCCCACC GGTGTTAAAG TATTTAGCTG ACTTGCAACT CTACATGGGG GCTCTATTAA
6481 ATGAGAGACA CCCCTTTTAT GGGCCCTTGG CTTTATCTTC CTATTTACAG TAGGCGGGCT
6541 TACAGGTATT GTTCTGGCCA ATTCATCTCT AGATATTGTA CTACACGATA CATACTATGT
6601 AGTAGCCCAC TTCCACTACG TACTATCTAT GGGGGCCGTA TTTGCCATTG TCGCCGCCTT
6661 CGTGCACTGA TTCCCGCTAT TCTCAGGATA TACACTCCAC AGCACTTGGA CAAAAATCCA
6721 TTTCGGCATT ATGTTCTTGG GGGTAAACTT AACCTTCTTC CCACAACACT TCCTCGGATT
6781 AGCCGGAATG CCCCGACGAT ACTCCGACTA CCCTGACGCC TATACCCTAT GAAATACAGT
6841 CTCCTCAATC GGATCACTTA TCTCGTTAGT GGCTGTTATT ATGTTCTTAT TTATTATTTG
6901 AGAGGCATTC GCCGCCAAAC GTGAAGTTCT AGCAACAGAT TTAACAACAA CCAATGTAGA
6961 ATGACTGCAT GGCTGCCCTC CCCCCTACCA CACATTCGAG GAGCCTGCCT TTGTACAAGT
7021 ACAAGCAGAC TAATGAGAAA GGGAGGAGTC GAACCCCATT AGGTCGGTTT CAAGCCGACC
7081 ACATAACCGC TCTGCCACTT TCTTTATAAG ACACTAGTAA AAGAGAACAT TACACCGCCT
```

(Figure 1. Continued)

```
 7141 TGTCAAGGCG GAAGTGTGGG TTAGACCCCC GCGTGTCTTG CTTTTAATGG CCCATCCATC
 7201 ACAGCTTGGA TTTCAAGATG CAGCTTCACC TGTTATAAAA GAACTTCTTC ATTTTCACGA
 7261 CCATGCTTTA ATAATCGTCT TCCTGATTAG CACACTTGTG CTTTACATTA TTCTTGCTAT
 7321 AGTTACCACT AAATTAACGA ACAAATATAT TTTAGATTCA CAAGAAATTG AAATTATCTG
 7381 AACAATTCTC CCAGCTATTA TTTTAATTCT AATTGCACTC CCCTCCCTCC GCATCCTCTA
 7441 CCTTATAGAT GAAATTAACA ACCCCCTATT AACAATTAAA GCCGTTGGTC ACCAATGATA
 7501 CTGAAGCTAT GAATATACTG ACTACGAAGA CCTTGGCTTT GATTCATACA TAATCCCCAC
 7561 CCAAGACCTA ACCCCTGGAC AATTCCGCCT ATTAGAAGCT GACCATCGCA TGGTTATCCC
 7621 AGTTGAATCC CCGATCCGAG TTTTAGTATC TGCAGATGAT GTACTCCACT CATGGGCAGT
 7681 CCCAGCCCTA GGGGTAAAAA TGGACGCAGT ACCAGGACGC CTTAATCAAA CAGCCTTCAT
 7741 CGCATCCCGA CCAGGCGTAT TCTACGGACA ATGCTCTGAA ATCTGCGGAG CAAATCACAG
 7801 CTTTATGCCT ATTGTAGTGG AAGCAGTTCC CCTAGAACAC TTTGAAAACT GATCATCTCG
 7861 AATACTTGAA GACGCCTCGC TAGGAAGCTA AATAGGGTAT AGCGTTAGCC TTTTAAGCTA
 7921 AAGATTGGTG GCCCCCAACC ACCCCTAACG ACATGCCCCA ACTCAACCCC GCACCTTGAT
 7981 TTGCTATTTT AGTCTTCTCG TGAATGGTCT TCCTGGCCGT TATTCTCGCT AAAGTTACAG
 8041 CCCACACCTT CCCAAACACT CCTACTCTGC AAAGCGCAGA AAAACCTAAA ACAGACCCCT
 8101 GAACTTGACC ATGACACTAA GCTTTTTTGA CCAGTTTATA AGCCCCACCT ATCTTGGAAT
 8161 CCCATTAATA GCCCTTGCCC TTACCCTACC CTGACTCCTT TACCCCACAC CTACAACTCG
 8221 ATGATTAAAT AACCGATTCC TCTCGCTTCA AGGTTGATTT ATTAACCGTT TTACTCAACA
 8281 GCTTCTCCTC CCCCTAAATA TGGGAGGTCA CAAGTGAGCC GCCCTCCTAA CTTCATTAAT
 8341 GATCTTTTTA ATTTCCCTAA ATATATTAGG ACTTCTTCCC TATACTTTTA CCCACACCAC
 8401 CCAATTATCA CTTAATTTAG GGCTCGCGGT ACCTCTCTGA TTAGCAACTG TTATTATTGG
 8461 CATGCGAAAC CAACCAACCC ACGCCCTAGG ACACCTCCTA CCAGAAGGCA CACCCGGCCC
 8521 CCTTATTCCA GTGCTTATTA TTATCGAAAC AATTAGTCTC TTTATTCGCC CCCTTGCCCT
 8581 AGGGGTACGG CTAACAGCCA ATTTAACAGC CGGTCACCTA TTAATTCAAT TAATTGCTAC
 8641 AGGTGCGTTC GTACTTCTTC CCCTAATACC AACCGTCGCA TTCATCACAA CAACAGTATT
 8701 GGTTCTCCTT ACCCTGTTAG AAGTTGCCGT AGCAATAATT CAAGCCTACG TCTTCGTTCT
 8761 CCTACTAACA CTGTACCTAC AAGAAAACGT CTAATGGCCC ATCAAGCACA CCCTTACCAC
 8821 ATAGTTGACC CCAGCCCTTG ACCCCTAACA GGGGCAATTG CTGCCCTCCT GATAACATCA
 8881 GGCCTCGCGA CCTGATTTCA TTTTCGCTCA ACAACCTTAA TAACCTTAGG AACAGCTCTA
 8941 CTGCTTCTTA CAATATATCA ATGATGACGA GATATCGTAC GAGAAGGCAC ATTCCAAGGA
 9001 CATCACACGC CCCCCGTACA AAAAGGTCTT CGATACGGAA TAATTCTTTT CATTACCTCC
 9061 GAAGTATTCT TTTTCCTAGG GTTCTTCTGA GCCTTTTACC ACGCAAGCCT CGCTCCCACT
 9121 CCTGAGTTAG GGGGCTGCTG GCCCCCCACG GGCATTACAA CCCTTGACCC ATTTGAAGTC
 9181 CCCCTCCTTA ATACAGCTGT CCTGCTTGCC TCCGGGGTAA CAGTTACCTG GGCCCACCAC
 9241 AGCATTATGG AAGGTGAACG AAAACAGACC ATTCAATCAC TAGCCTTAAC TATTCTTCTA
 9301 GGCTTTTACT TTACATTTCT TCAAGCCTTG GAATACTATG AAGCCCCCTT TACAATTGCA
 9361 GACGGCGTAT ACGGCTCTAC ATTTTTCGTA GCCACCGGAT TCCACGGACT ACACGTTATT
 9421 ATTGGCTCCA CATTTTTAGC CGTTTGCCTC CTACGGCAAA TCCAATACCA TTTTACATCC
 9481 GAGCACCATT TCGGGTTCGA AGCAGCTGCC TGATACTGAC ATTTCGTAGA CGTTGTCTGA
 9541 TTATTCCTAT ATATCTCTAT CTACTGATGA GGCTCTTAAT CTTTCTAGTA TTAAAACTAG
 9601 TATAAGTGAC TTCCAATCAC CCGGTCTTGG TTAAAATCCA AGGAAAGATA ATGAACGTAG
 9661 CAATAGCTGT AATTACCATC ACTATTTTGC TTTCCGTAGT CCTGGCCATT GTATCCTTCT
 9721 GACTCCCCCA AATGACCCCC GACCACGAAA AGCTCTCCCC ATATGAATGT GGTTTCGACC
 9781 CTTTAGGATC AGCCCGCCTA CCATTTTCCC TCCGCTTCTT CCTAGTCGCC ATTCTCTTCC
 9841 TACTTTTCGA TTTAGAAATT GCCCTTCTCC TCCCCCTCCC ATGAGGAGAC CAATTAACCT
 9901 CCCCTTTATT GACACTCTTC TGAGCCGTAG CCGTGCTTAT TCTTCTCACC CTTGGCTTAG
 9961 TCTACGAGTG AATTCAAGGA GGGTTAGAAT GAGCCGAATA GCCAATTAGT TTAAAAAAAA
10021 TATTTGATTT CGGCTCAAAA GCTTATGGTT AAAGTCCATA ATTGTCTAAT GACTCCCGCT
10081 CACTTCGCTT TCTCATCGGC CTTTACTCTA GGATTGACAG GCCTAGCATT CCATCGAACC
10141 CACCTCCTCT CCGCTCTTTT ATGCTTAGAA GGGATGATGC TCTCTTTATT TATTGGACTT
10201 TCAATTTGAA CCCTTCAACT GGGCTCCACA AGTTTCTCTG CGGCTCCTAT GCTTCTGTTA
10261 GCTTTTTCAG CTTGTGAAGC AAGCGCGGGG CTTGCCCTGC TGGTAGCCAC AGCTCGCACA
10321 CATGGCTCGG ACCGCCTTCA GACCTTAAAC CTCTTACAAT GCTAAAAATC CTAATTCCCA
10381 CTCTAATGCT TCTCCCCACA GCCTGGCTTG CCCCTGCCAA ATGATTGTGA CCTACTACTC
10441 TCTCCCACAG CCTAGTCATT GCATTAGCCA GCCTCACCTG ACTAAAAAAT ACATCCGAAA
10501 CAGGCTGGTC TTGCCTCACA CCCTTCATAG CCACAGACCC CCTCTCAACC CCCCTTCTTG
10561 TCCTCACCTG CTGACTACTC CCCCTTATGA TTTTGGCAAG CCAAAACCAC ACAGCACTCG
10621 AACCTATTAA CCGCCAACGA ACCTACATTA GCCTATTGAC ATCACTGCAA GTGTTCCTTA
```

(Figure 1. Continued)

```
10681 TTATAGCATT TGGTGCCACT GAACTACTTA TATTTTATGT CATATTTGAA GCTACTCTTA
10741 TCCCCACGCT AATTATTATT ACTCGCTGAG GTAACCAAGC AGAACGTCTT AATGCAGGAG
10801 TATATTTTTT GTTTTATACC CTGGCAGGCT CTCTCCCGTT ACTAGTCGCC CTATTACTTC
10861 TTCAGAAAGA TACAGGCTCC CTCTCCCTCT TAACCATCCA ATATGCCAGC TCTACCCCCC
10921 TTTCATCTTA TGCTGATAAG ATTTGATGGG CAGGCTGCCT GATTGCATTT TTAGTAAAAA
10981 TACCCCTATA CGGAGCACAT CTATGGCTGC CAAAAGCACA TGTAGAAGCC CCAGTTGCAG
11041 GCTCAATAGT CCTGGCTGCA GTTCTTCTAA AACTAGGGGG CTACGGCATA ATCCGAATAA
11101 TGACCATATT GGAACCTCTC ACTAAAGAAT TAAGTTATCC ATTTATTATC CTCGCCCTCT
11161 GAGGTGTAAT TATAACTGGC TCCACCTGCC TTCGCCAAAC AGATCTTAAA TCCCTCATCG
11221 CTTACTCATC CGTAAGCCAC ATGGGCCTAG TCGTTGGAGG TATTCTTATC CAAACACCAT
11281 GGGGCCTTGC CGGCGCCGTA ATTCTCATAA TTGCACACGG CCTAACATCC TCAGCCCTTT
11341 TCTGCTTGGC CAACACAAAT TACGAACGCC TCCATAGCCG GACAATACTA TTAGCTCGAG
11401 GGTTACAAAT AGTACTTCCA CTCATAGCAA CATGATGATT TATTGCCAGC CTCGCAAACT
11461 TAGCCCTTCC CCCCCTGCCC AACCTCATAG GGGAACTTTT AATTATTACC TCATTATTTG
11521 GTTGATCGTG ATGAACCCTC GTACTCACAG GGGCAGGAAC CCTCATTACC GCAAGCTATT
11581 CACTCTATAT ATTCCTCATG ACCCAGCGGG GCCCACTCCC AGCACATATT ATTAGCCTAA
11641 ACCCTTCCTA CACGCGGGAG CACCTAGTCA TAGCCCTTCA CCTCCTCCCC CTGCTTCTAC
11701 TTGTCTTAAA GCCCGAACTA GTATGAGGCT GAACCACCTG TAGATATAGT TTAACAAAAA
11761 TATTAGATTG TGATTCTAAA GACAGAGGTT AAAATCCCCT TATCCACCGA GAGAGGCTCG
11821 CCAGCAACGA AGACTGCTAA TCTCCGTGAC CTTGGTTGGA CCCCAGGGCT CACTCGGCCT
11881 GCTCCTAAAG GATAACAGCT CATCCATTGG TCTTAGGAAC CAAAAACTCT TGGTGCAAAT
11941 CCAAGTAGCA GCTATGCACT CCTCATCACT TATTATATCA TCCAGCTTAG TCATTATCTT
12001 TTTACTGTTA GCATATCCTA TTTTTACGAC TCTAGAGCCT CGCCCTCGAA ACCCCGACTG
12061 GGCCGTCTCA CATGTTAAGA CAGCAGTCGC CCTAGCTTTC TTCATCAGCC TAATTCCCCT
12121 ATTCCTGTTT CTTAACGAAG GAGCAGAAAC AATCATCACC TCATGAAATT GAATAAACAC
12181 ACTGACCTTC GACGTGAACA TTAGTTTCAA ATTTGACCAC TACTCAGTTA TCTTTGTACC
12241 TATTGCCCTC TATGTCACCT GATCCATTCT AGAATTTGCA TCATGATATA TACACGCAGA
12301 CCCATACATA AATCGATTTT TTAAATATCT CCTAATTTTC CTTATTGCCA TAATTATTCT
12361 TGTTACAGCA AATAATCTAT TCCAGCTTTT CATTGGTTGG GAAGGAGTAG GCATTATGTC
12421 ATTTCTTCTC ATTGGCTGAT GGTACGGACG AGCAGATGCC AACACAGCGG CCCTCCAGGC
12481 CGTTGTGTAT AACCGGGTCG GAGACATTGG ACTGCTATTC ACAATAGCAT GAATAGCAAC
12541 CAACGCTAAC TCCTGAGAGC TACAACAAAT TTTTGTGGCA ACAAAAGACC TAGATCTGAC
12601 CCTACCGCTA CTAGGCCTGA TTATCGCCGC TACAGGTAAA TCTGCCCAAT TTGGTCTCCA
12661 CCCTTGACTC CCCTCTGCTA TAGAGGGTCC TACACCGGTC TCTGCCCTAC TGCATTCAAG
12721 CACAATAGTC GTCGCCGGTA TTTTTCTTTT AGTGCGAACA AGCCCACTCC TAGAAAATAA
12781 TCAAACTGCC CTCACCACCT GCCTATGCCT AGGAGCCCTA ACAACACTAT TCACAGCCAC
12841 CTGTGCCCTG ACCCAAAATG ATATCAAGAA AATTGTAGCA TTCTCCACAT CAAGTCAACT
12901 TGGCCTAATA ATAGTCACTA TTGGACTAAA TCAGCCCCAA TTAGCCTTCC TACACATTTG
12961 CACCCACGCC TTCTTCAAGG CAATATTATT CCTCTGTTCT GGCTCAATTA TTCACAGTCT
13021 CAACGACGAA CAAGATATCC GAAAAATAGG AGGCATACAT CACCTTACCC CCTTTACATC
13081 CTCCTGCCTC ACTATTGGTA GTTTAGCCCT CACAGGCACC CCCTTCCTGG CAGGGTTCTT
13141 CTCCAAAGAT GCCATCATTG AAGCACTAAA CACATCCCAC CTAAACGCCT GAGCCCTAGT
13201 CCTAACCCTT CTAGCCACCT CATTCACGGC CATCTACAGT CTCCGCGTAG TATTCTTTGT
13261 CTCTATGGGC TACCCACGAT TTAACCCTAT TTCTCCCATC AATGAGAACA ACCCAGCAGT
13321 TATTAACCCC TTAAAACGAC TTGCATGAGG AAGCATTGTC GCCGGCCTCC TAATCATTTC
13381 AAGCATTACC CCCCTCAAGA CCCCTGTGAT ATCCATACCC CCCTTACTCA AATTAGCTGC
13441 CCTCGGAGTT ACAATTACAG GATTATTCAT TGCCCTCGAG CTGGCAACAT TGACCAATAA
13501 ACAGTACAAA ATCACCCCTA ATCTAGTTAC TCACCACTTC TCCAACATAC TAGGCTTTTT
13561 CCCCTCGATC ATTCACCGAT TTATACCCAA ACTAAATCTA GTTTTAGGAC AGGCACTTGC
13621 CAGCCAACTA ATTGACCAAA CTTGACTAGA GAAAGTTGGC CCCAAAGCAA TCTCTTCATC
13681 AAATATTCCC TTGATTACAA CAACAAGTAA TACCCAACAA GGAATAATTA AAACATATCT
13741 CACCCTATTC CTTCTCACCT TAACCCTCGC TGCCTTATTA TTTACCCGTT AAACTGCCCG
13801 AAGAGCCCCC CGACTTAGTC CACGAGTTAA CTCCAACACA ACAAATAAGG TAAGGAGCAA
13861 AACCCACGCA CTAAGTACCA GTATACCTCC CCCTAATGAA TACATTAACG CAACCCCTCC
13921 AATATCGCCT CGCAATACAG AGAGCTCACT AAGCTCATCA GCCGGCACCC ATGAGGACTC
13981 ATATCACCCC CCTCAAAATA CACTAGAAGC CACCCCCACC CCTACTAGGT ACATCAACAT
14041 ATCACCTACA ACAGGACCAC TTACCCAACT CTCCGGGTAC GGCTCAGCGG CGAGCGCCGC
14101 CGAATATGCA AACACAACTA ATATACCACC CAGATAAATC AAAAACAACA CCAATGATAG
14161 AAAAGGTCCC CCATGACCTA CCAATACTCC GCACCCCATG CCCGCCACAA CTACTAACCC
```

(Figure 1. Continued)

```
14221 CAAGGCAGCA AAGTAGAGAG AAGGGTTGCA GGCCTTGGCA CCCCCCCCCA GAACTAATCC
14281 AATTAAAAAT AAAGACATAA TGTAAATCAT AATTCCTGCC GGGACTTTAA CCAAAACCAA
14341 TGGCGTGAAA AACCCCCGTA GTTATTCAAC GACAACAATC CTGTAATGGC AAGTCTACGA
14401 AAGACACACC CTCTCCTCAA AATCGCAAAC AATGCCCTAG TTGACCTACC CGCCCCCTCA
14461 GATATTTCAG TGTGATGAAA CTTCGGATCT CTCTTAGGAC TCTGCTTAAT TATTCAAATC
14521 CTCACAGGAC TATTTCTAGC CATACACTAC ACCTCTGATA TTGCGACAGC TTTTTCTTCC
14581 GTCGCCCATA TTTGCCGAGA CGTAAATTAC GGCTGACTCA TCCGAAACCT TCACGCCAAC
14641 GGCGCATCTT TCTTCTTCGT ATGTATTTAT GCCCACATTG GCCGCGGACT TTACTACGGC
14701 TCATACCTCT ATAAAGAGAC ATGAAACATC GGAGTAGTCC TGCTACTTCT AGTTATAATA
14761 ACTGCTTTCG TCGGCTATGT TCTACCCTGA GGCCAAATGT CCTTTTGAGG TGCAACCGTT
14821 ATTACCAACC TACTCTCTGC AGTGCCCTAC GTAGGTGGCT CTTTAGTTCA ATGAATTTGA
14881 GGCGGATTCT CAGTAGACAA TGCAACCCTT ACCCGATTCT TTGCTTTCCA CTTCCTATTC
14941 CCCTTTGTAA TTGCAGGCGC AACCATAGTC CACCTTCTCT TCCTTCATCA AACAGGGTCA
15001 AATAATCCCC TCGGCCTAAA TTCAGACGCA GATAAAATAA GCTTTCACCC CTACTTCTCA
15061 TACAAAGACT TATTAGGGTT TGCAGTACTT GTTATCGCCC TCACATCTTT AGCTTTATTC
15121 TCACCCAACC TGCTAGGAGA CCCAGACAAC TTCACCCCAG CCAATCCGCT AGTTACTCCT
15181 CCCCACATCA AACCAGAGTG ATACTTCCTG TTCGCATATG CAATTCTGCG CTCCATCCCC
15241 AACAAACTAG GAGGAGTATT GGCCCTACTG GCCTCAATCC TCATTCTGAT GCTCGTACCA
15301 TTTCTACACA CGTCTAAACA ACGAAGTCTT ACCTTCCGGC CGCTTACACA ATTCTTGTTT
15361 TGAACCCTAA TTGCAGATGT TATTATTCTC ACCTGAATTG GAGGTATGCC TGTATCACAC
15421 CCGTTTGTTA TTATTGGACA AATTGCGTCC TTTTTATACT TTTTCCTCTT CCTAGTCCTT
15481 ACACCATTAG CAGGTTATGC AGAGGATAAA GCACTTGAAT GAACTTGCAT TAGTAGCTCA
15541 GCGTCAGAGC CCTGGTCTTG TAAACCAGAT GTCGGGGGTT AAAATCCTCC CTACTGCTCA
15601 AAGAAAGGAG ATTTTAACTC CCACCCCTGG CTCCCAAAGC CAGGATTCTT AATTAAACTA
15661 TTCTTTGTAG TGTATGTACA ATAATTTTAA ATACATATAT GTATTATCAA CATTAATTTA
15721 TATTAACCAT ATCATAGGGC ATTCAAGTAC ATATATGTTT TATCACCATA TCTAGGGTTA
15781 CACCATTCAG GAATTACATG ATACGAAAAT TTTACATAAA GCATGATAAT AATAATAAAC
15841 AAGTACTTAT AAACACCAGG CGAAATTTAA GACCTAACAC AAAAATCCAT AGGTTAAGTT
15901 ATACCTTTAC TCAAAATCCC GCCAAACTCA AATATTTAAT GTAGTAAGAG CCGACCAACA
15961 AGTCCATTTC TTAATGCCAA AGTTTATTGA AGGTGAGGGA CAAAAATTGT GGGGGTTTCA
16021 CAGAGTGATT TATTCCTGGC ATTTGGTTCC TATTTCAGGG CCACAAATTG TAAACATCCC
16081 CATAAATTAT TTCACCGGGC ATAGGTTAAT GGTGGAATAC AATAGGGGGA GGGGCCCCCA
16141 TGCCGAGCCT TCTTTTTATA GGGCAATGGG GTTTTTTTTT TTCTTTTTCT TTTCAATAGA
16201 CATTTCACAG TGCACGAAAT CTAGTTAACA AGGTGGGAAT CATCCTAGGA AGCATGGAAA
16261 TAGTATGAGT GATGAAAAGT CTTTACTAAA GAATTACATA TAAAAATCTC AAGGACATAA
16321 TATAGTGAAA TTTAGTCGGA AGATATCTAT ATGACCCCCT TTTGGCTTTT TCGCGTTAAA
16381 CCCCCCTACC CCCCTAAACT CCTGAGATAG CTAACGCTCC TGTAAACCCC CCGGAAACAG
16441 GAAAACCTCG AGT
```

(Figure 1. Continued)

Table 1. Summary of the mitochondrial genome of *S. hubbsi*

*H and L denote heavy and light strands, respectively.

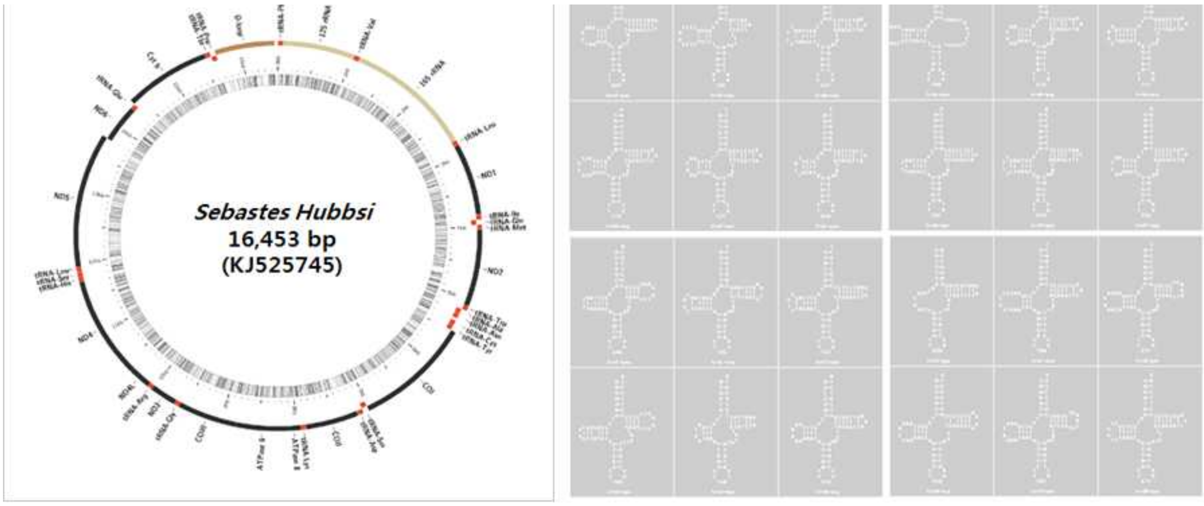


Figure 2. A gene arrangement in the complete mitogenome *Sebastes hubbsi*.

흰꼬리볼락의 미토게놈은 13개의 단백질 암호화 영역, 2개의 리보좀, 22개의 tRNA 및 1개의 조절영역으로 이루어져 있었고, 전체 크기는 16,445 bp 이었다(GenBank Accession number KJ834061)(Figure 3). 흰꼬리볼락의 미토콘드리아 DNA의 염기 조성비는 28.0% (A), 26.7% (T), 17.1% (G) 및 28.3% (C)이었고, AT 비율이 약간 높았으며(54.6%), COX1 (GTG)을 제외한 모든 단백질 암호영역의 개시코돈은 ATG 이었고, 종결코돈은 TAA가 주를 이루었으나 ND1과 ND3는 TAG, ND4는 AGA였으며 Cytb는 종결코돈이 불완전(T)하였다(Table 2). tRNA형태는 tRNA$^{Ser(AGY)}$(lacked DHU-arm 구조)을 제외하고 모두 전형적인 cloverleaf 구조로 되어 있었다(Figure 4).

```
   1 GCTAGCGTAG CTTAATTAAA GCATAACACT GAAGATGTTA AGATGGGCCC TAGAAAGCCC
  61 CGCCCGCACA AAGGCTTGGT CCTGACTTTA CTATCAACTT TAGCCAAATT TACACATGCA
 121 AGTATCCGCC CCCCTGTGAG AATGCCCTAC AGCTCCCTGC CCGGGAGCAA GGAGCTGGTA
 181 TCAGGCACAC ATCTGTAAGC CCATGACACC TTGCTTAGCC ACACCCTCAA GGGAACTCAG
 241 CAGTGATAAA CATTAAGCCA TAAGTGAAAA CTTGACTTAG TTAAAGCTAA GAGGGCCGGT
 301 AAACCTCGTG CCAGCCACCG CGGCTATACG AGAGACCCAA GTTGATACCA TTCGGCGTAA
 361 AGAGTGGTTA TGGAAAATAA AGACTAAAGC CGCACACCTT CAAAGCTGTT ATACGCATCC
 421 GAAGGCTAGA AGATCAACCA CGAAGGTAGC TTTACAACCC CTGACCCCAC GAAAGCTCTG
 481 GCACAAACTG GGATTAGATA CCCCACTATG CCTAGCCCTA AACCTTGGTA ATATATCACA
 541 TACCCTGCCC GCCTGGGAAC TACGAGCACC AGCTTAAAAC CCAAAGGACT TGGCGGTGCT
 601 TTAGACCCCC CTAGAGGAGC CTGTTCTAGA ACCGATAACC CCCGTTCAAC CTCACCCTTC
 661 CTTGTTTATC CCGCCTATAT ACCGCCGTCG TCAGCTTACC CTGTGAAGGA CTAAAAGTAA
 721 GCACAACTGG CACAACCCAA AACGTCAGGT CGAGGTGTAG CGCATGGAAG GGGAAGAAAT
 781 GGGCTACATT CCCTACACTA GGGAACACGA ACGGCGCACT GAAACACGCG CCTGAAGGAG
 841 GATTTAGTAG TAAGCGGAAA ATAGCGTGTT CCGCTGAAAT CGGCCCTGAA GCGCGCACAC
 901 ACCGCCCGTC ACTCTCCCCA AGCCTACCAC TTTAAGTAAT TAAAAACCCA AAAATCGCAG
 961 AGGGGAGGCA AGTCGTAACA TGGTAAGGGT ACCGGAAGGT GCACTTGGTA ATATCAGAGT
1021 GTAGTTAAAA TAGAATAATA CTTCCCTTAC ACTGAAGAGA CATCCGTGCA AATCGGATCA
1081 CCCTGATGCC CAACAGCTAG CCCACAAACA CAACAACAAC CAACCATTAT TTATAACCCC
1141 AAATGCACGA ATGTTTCAAT TAAACAAACC ATTTTTCCCC TTTAGTATGG GCGACAGAAA
1201 AAGGACTTAG GAGCAATAGA GAAAGTACCG CAAGGGATCG CTGAAAGAGA AATGAAACAA
1261 CCCAGTGAAG CTAAGTAAAG CAGAGATTTA TTCTCGTACC TTTTGCATCA TGATTTAGCC
1321 AGCGTGACCC AAGCAAAGAG TGCTTTAGTT TGACACCCCG AAACTAGGGG AGCTACTCCA
1381 AGACAGCCTA TTTATAGGGC GAACCCGTCT CTGTGGCAAA AGAGTGGAAT GAGCTTTGAG
1441 TAGAGGTGAT AAACCTACCG AACCTAGTTA TAGCTGGTTG CCCGAGAAAT GGATAGAAGT
1501 TCAGCCTCTC AGATTCTTTA TTCACCTCAG TATTACCCCA CCTGATACCA CAAGAAACTG
1561 TGAGAGTTAT TCAAAGGGGG TACAGCCCCT TTGAAACAAG ATACAACTTT TCCGGGAGGA
1621 AAAAGATCAT AATTAAATAA AGGTAAGTAT TTGGGTGGGC CTAAAAGCAG CCACCCCAGT
1681 AGAAAGCGTT ATAGCTCAAA TACATCACTA CCCCTCTCTA TCCTGATCAT TAATTCTTAC
1741 TCCCCCCTTC CCTACCGGGC CATCCCATGC ACACATGGGA GAGACCCTGC TAATATGAGT
1801 AATAAGAGAG CCAAGCCTCT CTCCTCGCAC ACATGTAATT CGGAACGAAC CCGCACCGAG
1861 CATTAACGGC CCCAAACGAA GAGGGACCTG AACAACAACC CAAACAACCA GAAAAAAATT
1921 CAAACATAAA CCGTTAACCC TACACAGGTG TGCATTTAAG GAAAGACTAA GAGAAAGAGA
1981 AGGAACTCGG CAAACAAATC AAGCCTCGCC TGTTTACCAA AAACATCGCC TCTTGCAAAG
2041 CTAAAGAATA AGAGGTCCCG CCTGCCCTGT GACTATTAGT TTAACGGCCG CGGTATTTTG
2101 ACCGTGCAAA GGTAGCGCAA TCACTTGTCT TTTAAATGAA GACCTGTATG AATGGCACAA
2161 CGAGGGCTTA ACTGTCTCCT CTTTCAAGTC AATGAAATTG ATCTCCCCGT GCAGAAGCGG
2221 GGATATAAAC ATAAGACGAG AAGACCCTAT GGAGCTTTAG ACACCAAAGA AGATCCTGTC
2281 AAGTAACCCT CTATAAAGGC CTGAACTAAT GGAACCCTTC CCTAATGTCT TTGGTTGGGG
2341 CGACCGCGGG GAAACAAAAA ACCCCCACGT GGAAAGGGAG CACCCCCTCC TACAACTAAG
2401 AGCCGCAGCT CTAATTAACA GAATATCTGA CCAATAAGAT CCGGCAATGC CGATCAACGG
2461 ACCGAGTTAC CCTAGGGATA ACAGCGCAAT CCCCTTTTAG AGCCCATATC GACAAGGGGG
2521 TTTACGACCT CGATGTTGGA TCAGGACATC CTAATGGTGC AGCCGCTATT AAGGGTCCGT
2581 TTGTTCAACG GTTAAAGTCC TACGTGATCT GAGTTCAGAC CGGAGTAATC CAGGTCAGTT
2641 TCTATCTATG GTGTGCTCTT TTCCAGTACG AAAGGACCGA AAAGAAGAGG CCCCTGCTCT
2701 AAGCAAGCCT CACCCCCACC TAGTGAAAAC AACTAAAGTA GGCAAGAGGG CATACCCCCA
2761 GTGCCTGAGA GAACGGCATG TTGGGGTGGC AGAGCCCGGT GAATGCAAAA GACCTAAGCC
2821 CTTTTTACAG AGGTTCAAGT CCTCTCCTTA ACTATGATTT CAGTCCTTAT TACCCATATT
2881 CTCAATCCCT TGGCCTTCAT TGTCCCCATC CTCTTAGCCG TCGCCTTCCT CACACTTTTA
2941 GAACGTAAGG TACTAGGGTA TATACAACTA CGAAAAGGCC CAAATATTGT AGGGCCTTAC
3001 GGACTATTAC AGCCTATCGC TGACGGTGTG AAGCTCTTTA TTAAAGAGCC CGTTCGCCCC
3061 TCCACTTCTT CCCCAGTACT ATTCCTCCTC GCCCCACTAC TCGCACTCAC ACTTGCCTTA
3121 ACCCTTTGAG CCCCTATACC TCTCCCCTAC CCAGTAATTG ACTTAAACCT TGGGATCCTA
3181 TTTATTTTGG CCCTATCAAG CCTCGCTGTC TACTCCATTC TAGGCTCAGG TTGAGCATCA
3241 AATTCAAAAT ACGCTCTCAT CGGGGCCCTT CGGGCTGTAG CCCAAACCAT TTCATATGAA
3301 GTTAGTCTAG GCTTAATCCT ATTAAGTACC ATTATCTTTA CAGGAGGTTT CACACTACAA
3361 ACCTTCAACA TTGCCCAAGA AAGCGTCTGA ATACTACTCC CAGCCTGACC ACTAGCCGCA
3421 ATGTGATATA TTTCAACCCT TGCAGAGACA AACCGTGCAC CCTTTGACCT TACTGAAGGC
3481 GAATCCGAAC TAGTCTCTGG CTTCAATGTC GAGTATGCGG GTGGCCCATT CGCCCTATTT
3541 TTCCTGGCCG AATATGCTAA CATTCTGCTT ATAAACACGC TTTCCGCTAC CCTCTTTTTA
```

Figure 3. Sequences of the mitogenome of *S. longispins* (GenBank Accession No. KJ834061).

```
3601 GGGGCCTCCC ATTTTCCAAT GCTACCTGAA CTCACCGCAG TTAACCTAAT AACCAAAGCA
3661 GCCCTTCTGT CCGTCTTATT CCTATGAGTC CGAGCCTCTT ACCCACGATT CCGCTACGAC
3721 CAACTTATAC ATCTAATTTG AAAAAACTTC CTCCCACTTA CACTAGCCCT GGTTATCTGA
3781 CACCTAGCCC TCCCCATTGC ATTTGCTGGC CTGCCACCCC AGCTATAGAT AAGAAGCCGT
3841 GCCTGAAGTA AAGGGCCACT TTGATAGAGT GACTTATGGG GGTTCAAATC CCCCCCGCTT
3901 CTTAGAAAAG GGGGACTCGA ACCCCGCCTA AGGAGAGCAA AACTCCTGGT GCTCCCACTA
3961 CACTATTTCC TAGTAAAGTC AGCTAATTCT AAGCTCTTGG TCCCATACCC CAAACACGAA
4021 GGTTAAAATC CCTCCTTTAC TAATGAACCC TTACATCTTA ACCGCCCTGC TATTTGGTAT
4081 TGGTTTAGGC ACTACTACCA CCTTCGCAAG CTCCCACTGA CTACTAGCCT GAATAGGCCT
4141 GGAAATAAAT ACTCTTGCCA TCATTCCCCT AATAGCTCAA CACCATCACC CCCGAGCAGT
4201 TGAAGCAGCC ACAAAATATT TCTTAATTCA AGCTGCCGGA GCAGCCATAC TACTCTTTGC
4261 CAGCACCACC AACGCTTGAT TAACTGGACA ATGGGACCTT TTACAAATTG CCCACCCTTT
4321 CCCAACTGCC CTTGTCACTT TGGCCCTCGC ACTAAAAGTG GGACTTGCAC CTGTGCACTC
4381 ATGACTGCCT GAAGTACTTC AGGGCCTAGA CCTAACCACA GGACTTATTT TGTCTACCTG
4441 GCAAAAACTT GCCCCATTTG CCTTATTAGT CCAAACCCCC TGTGCCAACA CCACCCTTCT
4501 AATCATTCTA GGACTCACCT CAACCATTGT AGGAGGCTGA GGAGGCCTTA ACCAAACCCA
4561 ACTTCGCAAG ATTCTTGCTT ACTCCTCCAT CGCACACCTA GGCTGAATAG TAATTGTGCT
4621 ACAATTCTCC CCCTCCTTGA CTATTTTAAC ATTATTCACA TACTTCATTA TAACATTCTC
4681 AGCATTTCTT ATGTTTAAAC TTAATAAAGC AACCAACATT AATACTCTAG CAACCTCATG
4741 AGCAAAAACC CCCGCCCTAA CAGCCCTTGC ACCCCTTTTA TTATTATCCT TAGGGGGACT
4801 CCCACCACTT ACAGGCTTTA TGCCAAAATG ACTTATCCTT CAAGAACTTA CTAAACAAGA
4861 CCTCGCCCCA GCTGCAACAC TAGCAGCAAT AACCGCCCTC CTCAGCCTAT ATTTTTACCT
4921 ACGACTATCA TACGCAATGG CATTAACTAT TTCACCCAAC AACCTCACCG CAATTTCCCC
4981 ATGACGCCTC CCCTCCTTAC AACTAACACT GCCACTTGCT ACCTCAGCCA TAGCTACGCT
5041 ACTGCTTCTA CCCCTAACAC CCGCCGCAAT AGCACTAATA ACCCTTTAAG GGACTTAGGT
5101 TAAAACAAGA CCAAGGGCCT TCAAAGCCCT AAGTGAGGGT GGAAGTCCCC CAGTCCCTGA
5161 TAAGGCTTGC GGGACACTAC CCCACATCTC CTGTATGCAA AACAGGTACT TTAATTAAGC
5221 TAAAGCCTTC CTAGAAGGGC AGGCCTCGAT CCTGCAAGAT CTTAGTTAAC AGCTAAGCGC
5281 TCAAACCAGC GAGCATCCAT CTACCTTTCC CCCGCCTAAA AAGCGGGCTG AAGGCGGGGG
5341 AAAGTCCCGG CAAACAACTA ACCTGCATCT TCAAATTTGC AATCTGATAT GTATAACACC
5401 TCAAAACTTC TGGTAAAAAA AGGATTCAAA CCTCTGTTTG TGGGGCTACA ATCCATCGCT
5461 AAAAAACTCA CCCATCTTAC CTGTGGCCAT CACACGTTGA TTTTTCTCCA CTAATCACAA
5521 AAACATCGGC ACCCTTTATC TATTATTTGG TGCCTGAGCC GGTATAGTAG GCACAGCCCT
5581 CAGCCTACTC ATTCGAGCAA AACTAAGCCA CCCGGGCGCT CTCCTTGAAG ACAACCAATT
5641 TTATAATGTA ATCGTTACGG CACATGCCTT CGTAATGATT TTCTTTATAG TAATGCCAAT
5701 TATAATCGGA GGTTTTGGAA ACTGATTAAT TCCCCTAATG ATTGGAGCCC CAGATATAGC
5761 ATTTCCTCGT ATAAATAATA TAAGTTTCTG ACTTCTTCCC CCTTCTTTTC TGCTACTACT
5821 TGCCTCCTCT GGAGTAGAGG CAGGCGCCGG GACCGGCTGA ACAGTGTACC CGCCCCTGGC
5881 TGGCAACTTA GCCCACGCAG GAGCATCAGT CGACCTAACA ATCTTTTCAC TCCACTTAGC
5941 AGGTATCTCC TCAATTCTTG GGGCAATCAA TTTTATTACC ACAATTATTA ATATGAAGCC
6001 TCCGGCCATC TCCCAATACC AGACGCCCCT ATTTGTGTGA GCCGTCTTAA TTACCGCTGT
6061 TCTTCTCCTT CTCTCTCTAC CAGTTCTCGC TGCAGGTATT ACAATGCTCC TTACGGACCG
6121 AAACCTTAAC ACCACCTTCT TCGACCCGGC CGGAGGGGGA GATCCAATCC TTTACCAACA
6181 CCTATTCTGG TTTTTTGGGC ACCCGGAGGT ATATATTCTT ATTCTGCCTG GCTTTGGTAT
6241 GATTTCACAT ATTGTCGCCT ATTACTCTGG CAAAAAAGAA CCCTTTGGAT ATATAGGCAT
6301 GGTCTGAGCA ATAATGGCTA TTGGCCTCCT AGGATTTATT GTATGGGCTC ATCACATATT
6361 CACAGTTGGT ATGGACGTAG ACACACGTGC TTACTTTACG TCTGCCACAA TAATCATCGC
6421 AATTCCCACC GGTGTTAAAG TATTTAGCTG ACTTGCAACT CTACATGGGG GCTCTATTAA
6481 ATGAGAGACA CCCCTTTTAT GGGCCCTTGG CTTTATCTTC CTATTTACAG TAGGCGGGCT
6541 TACAGGTATT GTTCTGGCCA ATTCATCTCT AGATATTGTA CTACACGATA CATACTATGT
6601 AGTAGCCCAC TTCCACTACG TACTATCTAT GGGGGCCGTA TTTGCCATTG TCGCCGCCTT
6661 CGTGCACTGA TTCCCGCTAT TCTCAGGATA TACACTCCAC AGCACTTGGA CAAAAATCCA
6721 TTTCGGCATT ATGTTCTTGG GGGTAAACTT AACCTTCTTC CCACAACACT TCCTCGGATT
6781 AGCCGGAATG CCCCGACGAT ACTCCGACTA CCCTGACGCC TATACCCTAT GAAATACAGT
6841 CTCCTCAATC GGATCACTTA TCTCGTTAGT GGCTGTTATT ATGTTCTTAT TTATTATTTG
6901 AGAGGCATTC GCCGCCAAAC GTGAAGTTCT AGCAACAGAT TTAACAACAA CCAATGTAGA
6961 ATGACTGCAT GGCTGCCCTC CCCCTTACCA CACATTCGAG GAGCCTGCCT TTGTACAAGT
7021 ACAAGCAGAC TAACGAGAAA GGGAGGAGTC GAACCCCCAT AGGTCGGTTT CAAGCCGACC
7081 ACATAACCGC TCTGCCACTT TCTTTATAAG ACACTAGTAA AAAAGTACAT TACACCGCCT
```

(Figure 3 Continued)

```
 7141 TGTCGAGACG GAAGTGTGGG TTAAACCCCC GCGTGTCTTG CTTTTAATGG CCCATCCGTC
 7201 ACAGCTTGGA TTTCAAGATG CAGCTTCACC TGTTATAGAA GAACTTCTTC ATTTTCACGA
 7261 CCATGCTTTA ATAATCGTCT TCCTAATTAG CACACTAGTG CTTTACATTA TTCTTGCTAT
 7321 AGTTACCACT AAATTAACGA ACAAATATAT TTTAGATTCA CAAGAGATTG AAATTATCTG
 7381 AACAATCCTC CCAGCTATCA TTTTAATTCT AATTGCACTA CCCTCCCTTC GAATTCTGTA
 7441 CCTTATAGAT GAAATTAACA ATCCCCTGTT AACAATTAAA GCCGTTGGCC ACCAATGATA
 7501 CTGAAGCTAT GAATACACTG ACTACGAAGA TCTTGGATTT GATTCATACA TAATCCCCAC
 7561 CCAGGACCTA ACCCCTGGAC AATTCCGCCT ATTAGAAGCC GACCATCGCA TGGTTATTCC
 7621 AGTTGAATCC CCCATCCGAG TTTTAGTATC TGCAGACGAT GTACTCCACT CATGGGCAGT
 7681 CCCAGCCCTA GGGGTAAAAA TGGACGCAGT CCCAGGACGC CTTAATCAAA CAGCCTTTAT
 7741 CGCATCCCGA CCAGGCGTAT TCTACGGACA ATGCTCTGAG ATCTGCGGAG CAAATCACAG
 7801 CTTTATACCT ATTGTAGTGG AAGCAGTTCC CCTAGAACAC TTTGAAAATT GATCATCTCG
 7861 AATACTTGAA GACGCCTCGC TAGGAAGCTA AATAGGGTAT AGCGTTAGCC TTTTAAGCTA
 7921 AAGATTGGTG GCCCCCAACC ACCCCTAACG ACATGCCCCA ACTCAACCCC GCACCTTGAT
 7981 TTGCTATTTT AGTCTTCTCG TGAATGGTCT TCCTGGCCGT TATTCCCGCT AAAGTTACAG
 8041 CCCACACCTT CCCAAACACT CCTACTCTGC AAAGCGCAGA AAAACCTAAA ACAGACCCCT
 8101 GAACTTGACC ATGACACTAA GCTTTTTTGA CCAGTTTATA AGCCCCACCT ATCTTGGAAT
 8161 CCCATTAATA GCCCTTGCCC TTACCCTACC CTGACTCCTT TACCCCACAC CTACAACTCG
 8221 ATGATTAAAT AACCGATTCC TCTCGCTTCA AGGTTGATTT ATTAACCGTT TTACTCAACA
 8281 GCTTCTCCTC CCCCTAAATA TTGGGGGTCA CAAGTGAGCC GCCCTCCTAA CTTCATTAAT
 8341 GATCTTTTTA ATTTCCCTAA ATATATTAGG ACTTCTTCCC TATACTTTTA CCCCCACCAC
 8401 CCAACTATCA CTAAATTTAG GGCTCGCGGT ACCTCTCTGA TTAGCAACTG TTATTATTGG
 8461 CATGCGAAAC CAACCAACCC ACGCCCTAGG ACACCTCCTA CCAGAAGGCA CACCCGGCCC
 8521 CCTTATTCCA GTGCTTATTA TTATCGAAAC AATTAGTCTC TTTATTCGCC CCCTTGCCCT
 8581 AGGGGTACGG CTAACAGCCA ATTTAACAGC CGGTCACCTA TTAATTCAAT TAATTGCTAC
 8641 AGGTGCCTTC GTACTTCTTC CCCTAATACC AACCGTCGCA ATTATCACAA CAACAGTATT
 8701 GGTTCTCCTT ACCCTGTTAG AAGTTGCCGT AGCAATAATT CAAGCCTACG TCTTCGTTCT
 8761 CCTGCTAACA CTGTACCTAC AAGAAAACGT CTAATGGCCC ATCAAGCACA CCCTTACCAC
 8821 ATAGTTGACC CCAGCCCTTG ACCCCTAACA GGGGCAATTG CTGCCCTCCT GATAACATCA
 8881 GGCCTCGCGA CCTGATTTCA TTTTCGCTCA ACAACCTTAA TAACCTTAGG AACAGCTCTA
 8941 CTGCTTCTTA CAATATATCA ATGATGACGA GATATCGTAC GAGAAGGTAC ATTCCAAGGA
 9001 CATCACACGC CCCCCGTACA AAAAGGTCTT CGATACGGAA TAATTCTTTT CATTACCTCC
 9061 GAAGTATTCT TTTTCCTAGG GTTCTTCTGA GCCTTTTACC ACGCAAGCCT CGCTCCCACT
 9121 CCTGAGTTAG GGGGCTGCTG GCCCCCCACG GGCATTACAA CCCTTGACCC ATTTGAAGTC
 9181 CCCCTCCTTA ATACAGCTGT CCTGCTTGCC TCCGGGGTAA CAGTTACCTG GGCCCACCAC
 9241 AGCATTATGG AAGGTGAACG AAAACAGACC ATTCAATCAC TAGCCTTAAC TATTCTTCTA
 9301 GGCTTTTACT TTACATTTCT TCAAGCCTTG GAATACTATG AAGCCCCCTT TACAATTGCA
 9361 GACGGCGTAT ACGGCTCTAC ATTTTTCGTA GCCACCGGAT TCCACGGACT ACACGTTATT
 9421 ATTGGCTCCA CATTTTTAGC CGTTTGCCTC CTACGGCAAA TCCAATACCA TTTTACATCC
 9481 GAACACCATT TCGGGTTCGA AGCAGCTGCC TGATACTGAC ATTTCGTAGA CGTTGTCTGA
 9541 TTATTCCTAT ATATCTCTAT CTACTGATGA GGCTCTTAAT CTTTCTAGTA TTAAAACTAG
 9601 TATAAGTGAC TTCCAATCAC CCGGTCTTGG TTAAAATCCA AGGAAAGATA ATGAACGTAG
 9661 CAATAGCTGT AATTACCATC ACTATTTTGC TTTCCGTAGT CCTGGCCATT GTATCCTTCT
 9721 GACTCCCCCA AATGACCCCC GACCACGAAA AGCTCTCCCC ATATGAATGT GGTTTCGACC
 9781 CTTTAGGATC AGCCCGCCTA CCATTTTCCC TCCGCTTCTT CCTAGTCGCC ATTCTCTTCC
 9841 TACTTTTCGA TTTAGAAATT GCCCTTCTCC TCCCCCTCCC ATGAGGAGAC CAATTAACCT
 9901 CCCCTTTATT GACACTCTTC TGAGCCGTAG CCGTGCTTAT TCTTCTCACC CTTGGCTTAG
 9961 TCTACGAGTG AATTCAAGGA GGGTTAGAAT GAGCCGAATA GCCAATTAGT TTAAAAAAAA
10021 TATTTGATTT CGGCTCAAAA GCTTATGGTT AAAGTCCATA ATTGTCTAAT GACTCCCGCT
10081 CACTTCGCTT TCTCATCGGC CTTTACTCTA GGATTGACAG GCCTAGCATT CCATCGAACC
10141 CACCTCCTCT CCGCTCTTTT ATGCTTAGAA GGGATGATGC TCTCTTTATT TATTGGACTT
10201 TCAATTTGAA CCCTCCAACT GGGCTCCACA AGTTTCTCTG CGGCTCCTAT GCTTCTGTTA
10261 GCTTTTTCAG CTTGTGAAGC AAGCGCGGGG CTTGCCCTGC TGGTAGCCAC AGCTCGCACA
10321 CATGGCTCGG ACCGCCTTCA GACCTTAAAC CTCTTACAAT GCTAAAAATC CTAATTCCCA
10381 CTCTAATGCT TCTCCCCACA GCCTGGCTTG CCCCTGCCAA ATGATTGTGA CCTACTACTC
10441 TCTCTCACAG CCTAGTCATT GCATTAGCCA GCCTCACCTG ACTAAAAAAT ACATCCGAAA
10501 CAGGCTGGTC TTGCCTCACA CCCTTCATAG CCACAGACCC CCTCTCAACC CCCCTTCTTG
10561 TCCTCACCTG CTGACTACTC CCCCTTATGA TTTTGGCAAG CCAAAACCAC ACAGCACTCG
10621 AACCTATTAA CCGCCAACGA ACCTACATTA GCCTATTGAC ATCACTGCAA GTGTTCCTTA
```

(Figure 3. Continued)

```
10681 TTATAGCATT TGGTGCCACT GAACTACTTA TATTTTATGT CATATTTGAA GCTACTCTTA
10741 TCCCCACGCT AATTATTATT ACTCGCTGAG GTAACCAAGC AGAACGTCTT AATGCAGGAG
10801 TATATTTTTT GTTTTATACC CTGGCAGGCT CTCTCCCGTT ACTAGTCGCC CTATTACTTC
10861 TTCAGAAAGA TACAGGCTCC CTCTCCCTCT TAACCATCCA ATATGCCAGC TCTACCCCCC
10921 TTTCATCTTA TGCTGATAAG ATTTGATGGG CAGGCTGCCT GATTGCATTT TTAGTAAAAA
10981 TACCCCTATA CGGAGCACAT CTATGGCTGC CAAAAGCACA TGTAGAAGCC CCAGTTGCAG
11041 GCTCAATAGT CCTGGCTGCA GTTCTTCTAA AACTAGGGGG CTACGGCATA ATCCGAATAA
11101 TGACCATATT GGAACCTCTC ACTAAAGAAT TAAGTTATCC ATTTATTATC CTCGCCCTCT
11161 GAGGTGTAAT TATAACTGGC TCCACCTGCC TTCGCCAAAC AGATCTTAAA TCCCTCATCG
11221 CTTACTCATC CGTAAGCCAC ATGGGCCTAG TCGTTGGAGG TATTCTTATC CAAACACCAT
11281 GGGGCCTTGC CGGCGCCGTA ATTCTCATAA TTGCACACGG CCTAACATCC TCAGCCCTTT
11341 TCTGCTTGGC CAACACAAAT TACGAACGCC TCCATAGCCG GACAATACTA TTAGCTCGAG
11401 GGTTACAAAT AGTACTTCCA CTCATAGCAA CATGATGATT TATTGCCAGC CTCGCAAACT
11461 TAGCCCTTCC CCCCCTGCCC AACCTCATAG GGGAACTTTT AATTATTACC TCATTATTTG
11521 GTTGATCGTG ATGAACCCTC GTACTCACAG GGGCAGGAAC CCTCATTACC GCAAGCTATT
11581 CACTCTATAT ATTCCTCATG ACCCAGCGGG GCCCACTCCC AGCACATATT ATTAGCCTAA
11641 ACCCTTCCTA CACGCGGGAG CACCTAGTCA TAGCCCTTCA CCTCCTCCCC CTGCTTCTAA
11701 TTGTCTTAAA GCCCGAACTA GTATGAGGCT GAACCACCTG TAGATATAGT TTAACAAAAA
11761 TATTAGATTG TGATTCTAAA GACAGAGGTT AAAATCCCCT TATCCACCGA GAGAGGCTCG
11821 CCAGCAACGA AGACTGCTAA TCTCCGTGAC CTTGGTTGGA CCCCAGGGCT CACTCGGCCT
11881 GCTCCTAAAG GATAACAGCT CATCCATTGG TCTTAGGAAC CAAAAACTCT TGGTGCAAAT
11941 CCAAGTAGCA GCTATGCACT CCTCATCACT TATTATATCA TCCAGCTTAG TCATTATCTT
12001 TTTACTGTTA GCATATCCTA TTTTTACGAC TCTAGAGCCT CGCCCTCGAA ACCCCGACTG
12061 GGCCGTCTCA CATGTTAAGA CAGCAGTCGC CCTAGCCTTC TTCATCAGCC TAATTCCCCT
12121 ATTCCTGTTT CTTAACGAAG GAGCAGAAAC AATTATCACC TCATGAAATT GAATGAACAC
12181 ACTGACCTTC GACGTGAACA TTAGTTTCAA ATTTGACCAC TACTCAGTTA TCTTTGTACC
12241 CATTGCCCTC TATGTCACCT GATCCATTCT AGAATTTGCA TCATGATATA TACACGCAGA
12301 CCCATACATA AATCGATTTT TTAAATATCT CCTAATTTTC CTTATTGCCA TAATTATTCT
12361 TGTTACAGCA AATAATCTAT TCCAGCTTTT CATTGGTTGG GAAGGAGTAG GCATTATGTC
12421 ATTTCTTCTC ATTGGCTGAT GGTACGGACG AGCAGATGCC AACACAGCGG CCCTCCAGGC
12481 CGTTGTGTAT AACCGGGTCG GAGACATTGG ACTGCTATTC ACAATAGCAT GAATAGCAAC
12541 CAACGCTAAC TCCTGAGAGC TACAACAAAT TTTTGTGGCA ACAAAAGACC TAGATCTGAC
12601 CCTACCGCTA CTAGGCCTGA TTATCGCCGC TACAGGTAAA TCTGCCCAAT TTGGTCTCCA
12661 CCCTTGACTC CCCTCTGCTA TAGAGGGTCC TACACCGGTC TCTGCCCTAC TGCATTCAAG
12721 CACTATAGTC GTCGCCGGTA TTTTTCTTTT AGTACGAACA AGCCCACTCC TAGAAAATAA
12781 TCAAACTGCC CTCACCACCT GCCTATGCCT AGGAGCCCTA ACAACACTAT TCACAGCCAC
12841 CTGTGCCCTG ACCCAAAATG ATATCAAGAA AATTGTAGCA TTCTCCACAT CAAGTCAACT
12901 TGGCCTAATA ATAGTCACTA TTGGACTAAA TCAGCCCCAA TTAGCCTTCC TACACATTTG
12961 CACCCACGCC TTCTTCAAGG CAATATTATT CCTCTGTTCT GGCTCAATTA TTCACAGTCT
13021 CAACGACGAA CAAGATATCC GAAAAATAGG AGGCATACAT CACCTTACCC CCTTTACATC
13081 CTCCTGCCTC ACTATTGGTA GTTTAGCCCT CACAGGCACC CCCTTCCTGG CAGGGTTCTT
13141 CTCCAAAGAT GCCATCATTG AGGCACTAAA CACATCCCAC CTAAACGCCT GAGCCCTAGT
13201 CCTAACCCTT CTAGCCACCT CATTCACGGC CATCTACAGT CTCCGCGTAG TATTCTTTGT
13261 CTCTATGGGC TACCCACGAT TTAACCCTAT TTCTCCCATC AATGAGAACA ACCCAGCAGT
13321 TATTAATCCC TTAAAACGAC TTGCATGAGG AAGCATTGTC GCTGGCCTCC TAATCATTTC
13381 AAGCATTACC CCCCTCAAGA CCCCTGTGAT ATCCATACCC CCCTTACTCA AATTAGCTGC
13441 CCTCGGAGTT ACAATTACAG GATTATTTAT TGCCCTCGAG CTGGCAACAT TGACCAATAA
13501 ACAGTACAAA ATCACCCCTA ATCTAGTTAC TCACCACTTC TCCAACATAC TAGGCTTTTT
13561 CCCCTCGATC ATTCACCGAT TTATACCCAA ACTAAATCTA GTTTTAGGAC AGGCACTTGC
13621 CAGCCAACTA ATTGACCAAA CTTGACTAGA GAAAGTTGGT CCCAAAGCAA TCTCTTCATC
13681 AAATATTCCC TTGATTACAA CAACAAGTAA TACCCAACAA GGAATAATTA AAACATATCT
13741 CACCCTATTC CTTCTCACCT TAACCCTCGC TGCCTTATTA TTTACCCGTT AAACTGCCCG
13801 AAGAGCCCCC CGACTTAGTC CACGAGTTAA CTCCAACACA ACAAATAAGG TAAGGAGCAA
13861 AACCCACGCA CTAAGTACCA GTATACCTCC CCCTAATGAA TACATTAACG CAACCCCTCC
13921 AATATCGCCT CGCAATACAG AGAGCTCACT AAGCTCATCA GCCGGCACCC ATGAGGACTC
13981 ATATCACCCC CCTCAAAATA CACTAGAAGC CACCCCCACC CCTACTAGGT ACATCAACAT
14041 ATCACCTACA ACAGGACCAC TCACCCAACT CTCCGGATAC GGCTCAGCGG CAAGCGCCGC
14101 CGAATATGCA AACACAACTA ATATACCACC CAGATAAATC AAAAACAACA CCAATGATAG
14161 AAAGGGTCCC CCATGACCTA CCAATACTCC GCACCCCATG CCCGCCACAA CTACTAACCC
```

(Figure 3. Continued)

```
14221 CAAGGCAGCA AAGTAAGGAG AAGGGTTTGA AGCAACTGCA ACCAACCCTA GAACTAATCC
14281 AATTAAAAAT AAAGACATAA TATAAGTCAT AATTCCTGCC AGGACTTTAA CCAGAACCAA
14341 TGGCTTGAAA AACCACCGTT GTTATTCAAC TACAAGAACC CGCTAATGGC AAGTCTACGA
14401 AAGACACACC CTCTCCTCAA AATCGCAAAC AATGCCCTAG TTGACCTACC CGCCCCCTCA
14461 AATATTTCAG TGTGATGAAA CTTCGGATCT CTCCTAGGAC TCTGCTTAAT TATCCAGATC
14521 CTCACAGGAC TATTTTTAGC CATACACTAT ACCTCTGATA TTGCTACGGC TTTTTCTTCC
14581 GTTGCCCACA TTTGTCGGGA CGTAAATTAT GGCTGACTCA TCCGAAACCT TCACGCTAAC
14641 GGTGCATCCT TCTTCTTTGT ATGTATTTAT GCCCATATTG GCCGCGGACT TTACTACGGC
14701 TCATACCTCT ATAAAGAAAC ATGAAACATC GGAGTAGTCC TTCTACTTTT AGTTATAATA
14761 ACTGCTTTCG TTGGCTATGT ACTACCCTGA GGCCAAATGT CCTTTTGAGG TGCAACCGTT
14821 ATTACTAACC TACTTTCCGC AGTACCCTAC GTAGGTAGCT CTTTAGTCCA ATGAATCTGA
14881 GGTGGGTTCT CAGTAGACAA TGCAACCCTT ACCCGATTCT TTGCCTTCCA CTTCCTATTC
14941 CCATTCGTAA TTGCAGGAGC AAGCATGGTA CACCTTCTTT TCCTTCATCA AACAGGATCA
15001 AATAATCCCC TCGGCCTAAA TTCAGACGCA GATAAAATAA GCTTCCACCC CTATTTCTCA
15061 TATAAAGACT TACTGGGGTT CGCAGTACTT GTCATTGCCC TCACATGTCT AGCTCTATTT
15121 TCGCCCAACC TACTGGGAGA CCCAGACAAC TTCACCCCCG CCAATCCACT AGTTACCCCT
15181 CCCCACATCA AACCAGAATG ATATTTCCTG TTCGCGTATG CAATTCTACG CTCCATCCCC
15241 AATAAACTAG GAGGAGTATT GGCCCTACTG GCCTCAATCC TCATTCTGAT GCTCGTACCA
15301 TTTCTACACA CGTCTAAACA ACGAAGTCTT ACCTTCCGGC CGCTTACACA ATTCTTGTTT
15361 TGAACCCTAA TTGCAGATGT TATTATTCTC ACCTGAATTG GAGGTATGCC TGTATCACAC
15421 CCGTTTGTTA TTATTGGACA AATTGCGTCC TTTTTATACT TTTTCCTCTT CCTAGTCCTT
15481 ACACCATTAG CAGGTTATGC AGAGGATAAA GCACTTGAAT GAACTTGCAT TAGTAGCTCA
15541 GCGTCAGAGC CCTGGTCTTG TAAACCAGAT GTCGGAGGTT AAAATCCTCC CTACTGCTCA
15601 AAGAAAGGAG ATTTTAACTC CCACCCCTGG CTCCCAAAGC CAGGATTCTT AATTAAACTA
15661 TTCTTTGTAG TGTATGTACA ATAATTTTAA ATACATATAT GTATTATCAA CATTAATTTA
15721 TATTAACCAT ATCATAGGGC ATTCAAGTAC ATATATGTTT TATCACCATA TCTAGGGTTA
15781 CACCATTCAG GAATTACATG ATACGAAAAT TTTACATAAA GCATGATAAT AATAATAAAC
15841 AAGTACTTAT AAACACCAGG CGAAATCTAA GACCTAACAC AAAAACCCAT AGGTTAAGTT
15901 ATACCTTTAC TCAAAATCCC GCCAAACTCA AATATTTAAT GTAGTAAGAG CCGACCAACA
15961 AGTCCATTTC TTAATGCCAA CGTTTATTGA AGGTGAGGGA CAAAAATTGT GGGGGTTTCA
16021 CAGAGTGATT TATTCCCGGC CATTGGGTTC TTTTTCAGGG CCCCAAATTG TAAACCTCCC
16081 CCATAAATTA TTTCCCCGGG CATAGGGTAA TGGGGGGATA CAAAAGCGGG AGCGGCCCCC
16141 CTGCCGGGCC TTTTTTTTAT AGGGCCTTGG GCTTTTTTTT TTTCTTTTTC TTTTCAATAG
16201 ACATTTCACA GTGCACGCAA TCTAGTTAAC AAGGTGGGAG TCATCATAGG AAGCATGGAA
16261 ATAGTATGAG TGATGAAAAG TCTTTACTAA AGAATTACAT ATAAAAATAT CAAGGACATA
16321 ATATAGTGAA ATTTAGTCGG AAGATATCTA TATGACCCCC TTTGGCTTTT TCGCGTAAAC
16381 CCCCCTACCC CCCTAAACTC CTGAGATAAC TAACGCTCCT GTAAACCCCC CGGAAACAGG
16441 AAAAC
```

(Figure 3. Continued)

Table 2. Summary of the mitochondrial genome of *S. longispinis*

*H and L denote heavy and light strands, respectively.

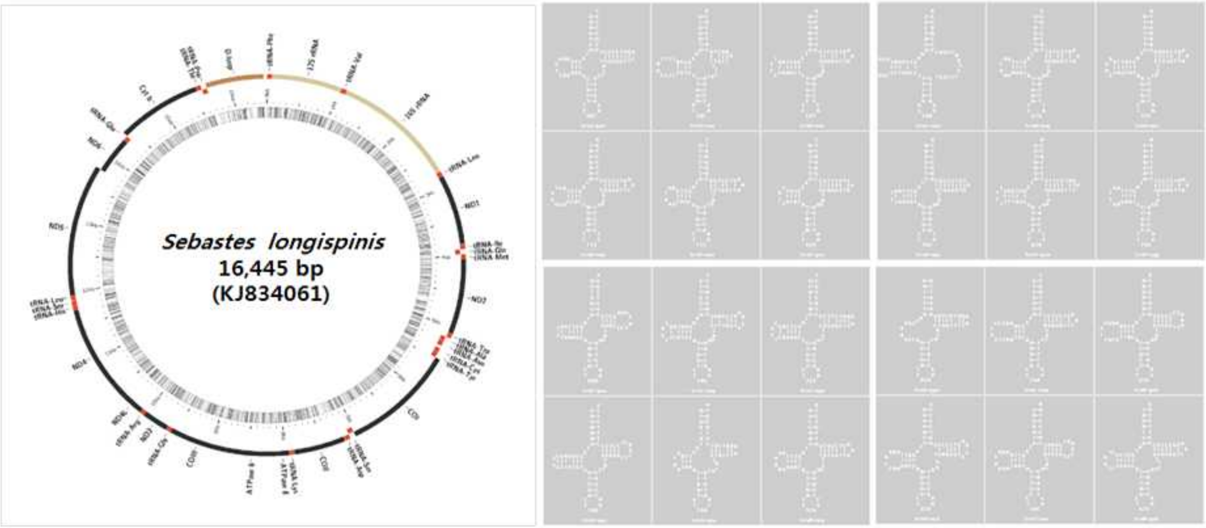


Figure 4. A gene arrangement in the complete mitogenome *Sebastes longispinis*.

황볼락의 미토게놈은 13개의 단백질 암호화 영역, 2개의 리보좀, 22개의 tRNA 및 1개의 조절영역으로 이루어져 있었고, 전체 크기는 16,465 bp 이었다(GenBank Accession number KJ834063)(Figure 5). 황볼락의 미토콘드리아 DNA의 염기 조성비는 27.7% (A), 26.6% (T), 17.3% (G) 및 28.4% (C)이었고, AT 비율이 약간 높았음(54.3%), COX1 (GTG)을 제외한 모든 단백질 암호영역의 개시코돈은 ATG이었고, 종결코돈은 TAA가 주를 이루었으나 ND1과 ND3는 TAG, ND4는 AGA였으며 Cytb는 종결코돈이 불완전(T)하였다(Table 3). tRNA형태는 tRNA$^{Ser(AGY)}$(lacked DHU-arm 구조)을 제외하고 모두 전형적인 cloverleaf 구조로 되어 있었다(Figure 6).

```
   1 GCTAGCGTAG CTTAATTAAA GCATAACACT GAAGATGTTA AGATGGACCC TAGAAAGCCC
  61 CGCCCGCACA AAGGCTTGGT CCTGACTTTA CTATCAACTT TAGCCAAATT TACACATGCA
 121 AGTATCCGCC CCCCTGTGAG AATGCCCTAC AGCTCCCTGC CCGGGAGCAA GGAGCTGGTA
 181 TCAGGCACAC ATCTGTAAGC CCATGACACC TTGCTTAGCC ACACCCTCAA GGGAACTCAG
 241 CAGTGATAAA CATTAAGCCA TAAGTGAAAA CTTGACTTAG TTAAAGCTAA GAGGGCCGGT
 301 AAACCTCGTG CCAGCCACCG CGGCTATACG AGAGACCCAA GTTGATACCA TTCGGCGTAA
 361 AGAGTGGTTA TGGAAAATAA AGACTAAAGC CGCACACCTT CAAAGCTGTT ATACGCATCC
 421 GAAGGCTAGA AGATCAACCA CGAAGGTAGC TTTACAACCC CTGACCCCAC GAAAGCTCTG
 481 GCACAAACTG GGATTAGATA CCCCACTATG CCTAGCCCTA AACCTTGGTA ATATATCACA
 541 CACCCTACCC GCCTGGGAAC TACGAGCACC AGCTTAAAAC CCAAAGGACT TGGCGGTGCT
 601 TTAGACCCCC CTAGAGGAGC CTGTTCTAGA ACCGATAACC CCCGTTCAAC CTCACCCTTC
 661 CTTGTTTATC CCGCCTATAT ACCGCCGTCG TCAGCTTACC CTGTGAAGGC CTAAAAGTAA
 721 GCACAACTGG CACAACCCAA AACGTCAGGT CGAGGTGTAG CGCATGGAAG GGGAAGAAAT
 781 GGGCTACATT CCCTACATTA GGGAACACGA ACGGCGCACT GAAATACGCG CCTGAAGGAG
 841 GATTTAGTAG TAAGCGGAAA ATAGCGTGTT CCGCTGAAAT CGGCCCTGAA GCGCGCACAC
 901 ACCGCCCGTC ACTCTCCCCA AGCCTATCAC TTTAAATAAT TAAAAACCCA AAAATCGCGG
 961 AGGGGAGGCA AGTCGTAACA TGGTAAGGGT ACCGGAAGGT GCACTTGGTA ATATCAGAGT
1021 GTAGTTAAAA TAGAATAACA CTTCCCTTAC ACTGAAGAGA CACCCGTGCA AATCGGATCA
1081 CCCTGATGCC CAACAGCTAG CCCACAAACA CAACAACAAC CAACCATTAT TTATAACCCC
1141 AAACGCACGA GTGTTTTAAT TAAACAAACC ATTTTTCCCC TTTAGTATGG GCGACAGAAA
1201 AAGGACTTAG GAGCAATAGA GAAAGTACCG CAAGGGATCG CTGAAAGAGA AATGAAATAA
1261 CCCAGTGAAG CTAAGTAAAG CAGAGATTTA TTCTCGTACC TTTTGCATCA TGATTTAGCC
1321 AGCGTGACCC AAGCAAAGAG TGCTTTAGTT TGACACCCCG AAACTAGGGG AGCTACTCCA
1381 AGACAGCCTA TTTATAGGGC GAACCCGTCT CTGTGGCAAA AGAGTGGAAT GAGCTTTGAG
1441 TAGAGGTGAT AAACCTACCG AACCTAGTTA TAGCTGGTTG CCCGGGAAAT GGATAGAAGT
1501 TCAGCCTCTC AGATTTTTTA TTCACCTCAG TATTACCCCA CCTGATACCA CAAGAAACTG
1561 TGAGAGTTAT TCAAAGGGGG TACAGCCCCT TTGAAACAAG ATACAACTTT TCCGGGAGGA
1621 AAAAGATCAT AATTAAATAA AGGTAAGTAT TTGGGTGGGC CTAAAAGCAG CCACCCCAGT
1681 AGAAAGCGTT ATAGCTCAAA TACATCACTA CCCCTCTCTA TCCTGATCAT TAATTCTTAC
1741 TCCCCCCTTC CCTACCGGGC CATCCCATGC AAACATGGGA GGGACCCTGC TAATATGAGT
1801 AATAAGAGAG CCAAGCCTCT CTCCTTGCAC ACATGTAATT CGGAACGAAC CCGCACCGAG
1861 CATTAACGGC CCCAAACGAA GAGGGACCTG AACAACAACC CAAACAACCA GAAAAAAACT
1921 CAAACATAAA CCGTTAACCC TACACAGGTG TGCACCTAGG GAAAGACTAA AAGAAAGAGA
1981 AGGAACTCGG CAAACAAATC AAGCCTCGCC TGTTTACCAA AAACATCGCC TCTTGCAAAG
2041 CTAAAGAATA AGAGGTCCCG CCTGCCCTGT GACTATTAGT TTAACGGCCG CGGTATTTTG
2101 ACCGTGCAAA GGTAGCGCAA TCACTTGTCT TTTAAATGAA GACCTGTATG AATGGCACAA
2161 CGAGGGCTTA ACTGTCTCCT CTTTCAAGTC AATGAAATTG ATCTCCCCGT GCAGAAGCGG
2221 GGATATAAAC ATAAGACGAG AAGACCCTAT GGAGCTTTAG ACACCAAAGA AGATCCTGTC
2281 AAGTAACCCC TTATAAGGGC CTGAACTAAT GGAACCCTTC CCTAATGTCT TTGGTTGGGG
2341 CGACCGCGGG GAAACAAAAA ACCCCCACGT GGAAAGGGAG CACCCCCTCC TACAACTAAG
2401 AGCCGCAGCT CTAATTAACA GAATATCTGA CCAATAAGAT CCGGCAACGC CGATCAACGG
2461 ACCGAGTTAC CCTAGGGATA ACAGCGCAAT CCCCTTTTAG AGCCCATATC GACAAGGGGG
2521 TTTACGACCT CGATGTTGGA TCAGGACATC CTAATGGTGC AGCCGCTATT AAGGGTCCGT
2581 TTGTTCAACG GTTAAAGTCC TACGTGATCT GAGTTCAGAC CGGAGTAATC CAGGTCAGTT
2641 TCTATCTATG GTGTGCTCTT TTCTAGTACG AAAGGACCGA AAAGAAGAGG CCCCTGCTCT
2701 AAGCAAGCCT CACCCCCACC TAGTGAAGAC AACTAAAGTA GGCAAGAGGG CATACCCCCA
2761 ATGCCTGAGA GAACGGCATG TTGGGGTGGC AGAGCCCGGT GAATGCAAAA GACCTAAGCC
2821 CTTTTTACAG AGGTTCAAGT CCTCTCCTTA ACTATGATCT CAGTGCTTAT TACCCATATT
2881 CTTAACCCCT TGGCCTTCAT TGTCCCCGTC CTCTTAGCCG TCGCCTTCCT CACACTTCTA
2941 GAACGTAAAG TACTAGGATA CATACAACTA CGAAAGGGTC CAAATATTGT AGGGCCTTAC
3001 GGCCTATTAC AGCCTATTGC TGATGGTGTA AAACTCTTTA TTAAAGAGCC TGTTCGCCCC
3061 TCCACTTCCT CCCCCGTACT ATTTCTCCTC GCCCCACTGC TCGCACTCAC GCTTGCCTTA
3121 ACCCTCTGAG CCCCCATGCC CCTCCCTTAC CCAGTCATCG ACTTAAACCT TGGGATCCTA
3181 TTTATTTTAG CCCTATCAAG CCTCGCTGTC TACTCCATCC TGGGCTCAGG CTGAGCATCA
3241 AATTCAAAAT ATGCTCTCAT CGGGGCCCTT CGGGCTGTAG CCCAAACCAT TTCATATGAA
3301 GTTAGTCTGG GCCTAATCCT ATTAAGTACT ATCATTTTTA CAGGGGGTTT TACACTACAA
3361 ACCTTCAACA TTGCCCAAGA AAGCGTCTGA ATACTTCTCC CAGCCTGACC ATTAGCCGCA
3421 ATATGGTATA TTTCAACCCT TGCAGAGACA AACCGTGCAC CCTTTGACCT TACTGAAGGC
3481 GAATCGGAAC TAGTCTCTGG CTTCAATGTC GAATATGCAG GTGGCCCGTT TGCTCTATTT
3541 TTCTTAGCCG AATATGCTAA TATTCTGCTT ATAAATACAC TTTCCGCCAC CCTCTTCTTA
```

Figure 5. Sequences of the mitogenome of *S. owstoni* (GenBank Accession No. KJ834063)

```
3601 GGGGCCTCTC ACTTTCCAAT ACTACCTGAA CTCACCGCAG TAAACCTGAT AACCAAAGCA
3661 GCCCTTCTAT CTGTCCTATT CCTATGAGTT CGAGCCTCTT ACCCACGATT CCGATACGAC
3721 CAGCTCATGC ATCTAATTTG AAAAAACTTC CTCCCGCTTA CACTAGCCCT AGTTATCTGA
3781 CACCTAGCCC TCCCCATTGC ATTTGCTGGC CTGCCGCCCC AGCTATAGAT AAGAAGCCGT
3841 GCCTGAAGTA AAGGGCCACT TTGATAGAGT GACTTATGGG GGTTCAAATC CCCCCCGCTT
3901 CTTAGAAAAG GGGGACTCGA ACCCCGCCTA AGGAGAGCAA AACTCCTGGT GCTCCCACTA
3961 CACTATTTCC TAGTAAAGTC AGCTAATTCT AAGCTCTTGG TCCCATACCC CAAACACGAA
4021 GGTTAAAATC CCTCCTTTAC TAATGAACCC TTACATCTTA ACCGCCCTGC TATTTGGTAT
4081 TGGTTTAGGC ACTACTACCA CCTTCGCAAG CTCCCACTGA CTACTAGCCT GAATGGGCCT
4141 GGAGATAAAT ACTCTCGCCA TCATTCCCCT AATAGCTCAA CACCATCACC CCCGAGCAGT
4201 TGAAGCAGCC ACTAAATATT TCTTGATTCA AGCTGCCGGG GCAGCTATGC TACTCTTTGC
4261 CAGCACCACC AACGCTTGAT TAACTGGACA ATGAGACCTC TTACAAATTG CCCACCCCTT
4321 CCCAACTGCT CTTGTCACTT TGGCCCTCGC ACTAAAAGTG GGACTTGCAC CTGTACACTC
4381 ATGACTACCT GAAGTACTTC AAGGCCTAGA CCTAACTACA GGACTTATTT TGTCGACCTG
4441 ACAAAAACTT GCCCCATTTG CCTTATTAGT CCAAACCCCC TGTGCCAACA CCACCCTTTT
4501 AGTTATTCTA GGGCTTACCT CAACCATTGT AGGAGGCTGA GGGGGACTCA ACCAAACCCA
4561 ACTTCGCAAA ATCCTTGCCT ACTCCTCCAT CGCACATCTA GGCTGGATAG TAATTGTACT
4621 ACAATTCTCC CCCTCCCTAA CTATTTTAAC ATTATTTACA TACTTCATTA TAACCTTCTC
4681 AGCATTTCTT ATGTTTAAGC TTAATAAAGC AACCAGCATT AATGCTCTAG CAACCTCATG
4741 GGCAAAAACC CCCGCCCTAA CCGCCCTTGC ACCCCTACTA TTATTATCCT TAGGGGGCCT
4801 CCCCCCACTG ACAGGCTTTA TGCCAAAGTG ACTTATTCTT CAAGAACTTA CTAAGCAAGA
4861 CCTTGCCCCC GCTGCAACAC TGGCGGCGAT AACCGCCCTC CTTAGCCTAT ATTTTTATCT
4921 GCGACTATCA TACGCAATAG CACTAACTAT TTCACCCAAC AACCTCACCG CAATTTCCCC
4981 ATGACGCCTC CCCTCCTTAC AACTAACACT ACCACTTGCT ACTTCGGCCA TAGCTACGCT
5041 GCTGCTTCTA CCCCTAACAC CCGCCGCAAT AGCACTAATA ACCCTTTAAG GGACTTAGGT
5101 TAAAACAAGA CCAAGGGCCT TCAAAGCCCT AAGTGAGGGT GGAAGTCCCC CAGTCCCTGA
5161 TAAGGCTTGC GGGACACTAC CCCACATCTC CTGTATGCAA AACAGGTACT TTAATTAAGC
5221 TAAAGCCTTC CTAGAAGGGC AGGCCTCGAT CCTGCAAGAT CTTAGTTAAC AGCTAAGCGC
5281 TCAAACCAGC GAGCATCCAT CTACTTTTCC CCCGCCTGAC GGGCGGGCGG AGGCGGGGGA
5341 AAGTCCCGGC AGACGACTAA CCTGCATCTT CAGATTTGCA ATCTGATATG TATAACACCT
5401 CAAGACTTCT GGTAAGAAGA GGATTCAAAC CTCTGTTTGT GGGGCTACAA TCCATCGCTT
5461 AAAAACTCAG CCATCCTACC TGTGGCCATC ACACGTTGAT TTTTCTCCAC TAATCACAAA
5521 GACATCGGCA CCCTTTATCT AGTATTTGGT GCCTGAGCCG GTATAGTAGG CACAGCCCTC
5581 AGCCTACTCA TTCGAGCAGA ACTAAGCCAA CCGGGCGCTC TCCTTGGAGA CGACCAAATT
5641 TATAATGTAA TCGTTACAGC ACATGCCTTC GTAATGATTT TCTTTATAGT AATGCCAATT
5701 ATGATTGGAG GTTTTGGAAA CTGATTAATC CCCCTAATGA TCGGAGCCCC AGATATAGCA
5761 TTCCCTCGTA TAAATAACAT AAGTTTCTGA CTTCTACCCC CTTCTTTCCT GCTACTACTT
5821 GCCTCCTCTG GAGTAGAAGC AGGTGCCGGA ACCGGGTGAA CAGTGTACCC ACCCCTGGCT
5881 GGTAATTTAG CCCACGCAGG AGCATCAGTC GACCTGACAA TCTTTTCACT TCACCTGGCA
5941 GGTATTTCCT CAATTCTTGG GGCAATCAAT TTTATTACCA CAATTATTAA TATGAAACCT
6001 CCAGCCATCT CTCAATACCA AACACCCCTG TTTGTGTGAG CCGTCCTAAT TACCGCTGTT
6061 CTTCTCCTTC TCTCCTTACC AGTCCTTGCT GCCGGCATCA CAATGCTCCT TACCGACCGA
6121 AACCTTAATA CCACCTTCTT TGACCCGGCC GGAGGAGGAG ATCCAATCCT TTACCAGCAC
6181 TTATTCTGGT TCTTTGGACA CCCGGAAGTA TATATTCTCA TTCTGCCTGG CTTTGGTATG
6241 ATTTCACACA TCGTCGCCTA TTACTCTGGC AAAAAAGAAC CCTTTGGCTA TATGGGTATA
6301 GTGTGAGCAA TAATGGCTAT TGGCCTCTTA GGCTTTATTG TATGAGCTCA CCACATGTTT
6361 ACAGTTGGCA TGGACGTAGA CACACGTGCT TATTTTACGT CTGCCACAAT AATCATCGCA
6421 ATTCCCACCG GTGTTAAAGT ATTTAGCTGA CTTGCAACCC TACATGGGGG CTCTATTAAA
6481 TGAGAAACAC CCCTTCTATG GGCCCTTGGC TTTATTTTCT TATTTACAGT AGGCGGGCTT
6541 ACAGGCATTG TTCTGGCCAA TTCATCTCTA GATATTGTAC TCCACGATAC CTATTATGTA
6601 GTAGCCCACT TCCACTACGT ATTATCTATG GGGGCCGTAT TTGCCATTGT CGCCGCCTTT
6661 GTGCACTGAT TCCCACTATT CTCAGGCTAC ACACTTCACA GTACTTGAAC GAAAATCCAC
6721 TTCGGTATTA TGTTCTTAGG GGTAAACTTA ACCTTCTTCC CACAACACTT CCTCGGATTA
6781 GCCGGAATGC CCCGACGATA CTCTGACTAC CCTGACGCCT ATACCCTATG AAATACAGTC
6841 TCCTCAATCG GATCACTTAT CTCCTTAGTA GCTGTTATTA TGTTCTTATT TATTATTTGA
6901 GAGGCATTCG CCGCCAAACG TGAAGTTCTA GCAACAGATT TAACAACAAC CAATGTAGAA
6961 TGACTACATG GCTGCCCTCC CCCATACCAC ACATTCGAGG AGCCTGCCTT TGTACAAGTA
7021 CAAGCAGACT AACGAGAAAG GGAGGAGTCG AACCCCCATA GGTCGGTTTC AAGCCGACCA
7081 CATAACCGCT CTGCCACTTT CTTTATAAGA CACTAGTAAA AGAGAACATT ACACCGCCTT
```

(Figure 5. Continued)

```
 7141 GTCAAGGCGG AAGTGTGGGT TAGACCCCCG CGTGTCTTGC TTTTAATGGC CCATCCGTCA
 7201 CAGCTTGGAT TTCAAGATGC AGCTTCACCT GTTATAGAAG AACTTCTTCA TTTTCACGAC
 7261 CATGCTTTAA TAATCGTCTT CCTAATTAGC ACACTTGTGC TTTACATTAT TCTTGCTATA
 7321 GTTACCACTA AATTAACGAA CAAATATATT TTAGATTCAC AAGAGATTGA AATTATCTGA
 7381 ACAATTCTCC CAGCTATTAT TCTAATTCTA ATTGCACTCC CCTCCCTCCG CATCCTCTAC
 7441 CTTATAGATG AGATTAATAA CCCCTTATTG ACAATTAAAG CCGTTGGCCA CCAATGATAC
 7501 TGAAGCTATG AGTACACTGA CTACGAAGAC CTTGGCTTTG ACTCATACAT AATCCCCACC
 7561 CAAGACCTAA CCCCTGGACA ATTCCGCCTA TTAGAAGCCG ACCATCGCAT GGTTATTCCA
 7621 GTTGAATCCC CTATCCGAGT CTTAGTATCT GCAGACGATG TACTCCATTC ATGAGCAGTC
 7681 CCAGCCCTGG GGGTAAAAAT GGACGCAGTA CCAGGACGCC TTAATCAAAC AGCCTTCATC
 7741 GCATCCCGAC CAGGCGTATT CTACGGACAA TGCTCTGAAA TCTGCGGAGC AAATCACAGC
 7801 TTTATACCTA TTGTAGTGGA AGCAGTTCCC CTAGAACACT TTGAAAACTG ATCATCTCGA
 7861 ATACTTGAAG ACGCCTCGCT AGGAAGCTAA ATAGGGCATA GCGTTAGCCT TTTAAGCTAA
 7921 AGATTGGTGA CTCCCAACCA CCCCTAACGA CATGCCCCAA CTCAACCCCG CACCTTGATT
 7981 TGCTATTTTA GTCTTCTCGT GAATGGTCTT CCTGGCCGTT ATTCCCGCTA AAGTTACAGC
 8041 CCACACCTTC CCAAATACCC CTAATCTGCA AAGCGCAGAA AAACCCAAAA CAGACCCCTG
 8101 AACCTGACCA TGACACTAAG CTTTTTTGAC CAGTTTATAA GCCCCACCTA TCTTGGAATC
 8161 CCACTAATAG CCCTTGCCCT TACCTTACCC TGACTCCTTT ACCCCACACA TACAACTCGA
 8221 TGATTAAATA ACCGATTCCT CGCGCTTCAG GGTTGATTTA TCAACCGTTT TACTCAACAG
 8281 CTCCTCCTCC CCTTAAATAT TGGAGGCCAT AAGTGAGCCG CCCTCCTGAC CTCATTAATG
 8341 ATCTTTTTAA TTACCCTAAA TATGTTAGGA CTTCTTCCCT ATACTTTTAC CCCCACCACC
 8401 CAACTATCAC TAAATTTAGG GCTTGCGGTA CCTCTCTGAT TAGCAACTGT TATTATTGGC
 8461 ATGCGAAACC AACCAACCCA TGCCCTAGGA CACCTCCTAC CAGAAGGCAC ACCTGGCCCC
 8521 CTCATCCCCG TACTTATCAT TATCGAAACA ATTAGCCTCT TTATTCGCCC CCTTGCCCTA
 8581 GGAGTACGAC TAACGGCCAA TTCAACAGGT GGCCACCTCT TAATTCAAGT AAATGGTACA
 8641 GGTGCATTCG TACTTCTCCC CTTAATACCA ACCGTCGCAA TCATCACAAC AACAGTACTG
 8701 GTTCTCCTCA CCCTATTAGA AGTTGCTGTA GCAATAATTC AAGCCTACGT CTTCGTTCTC
 8761 CTACTAACAC TATACCTACA AGAAAACGTC TAATGGCCCA TCAAGCACAC CCTTACCACA
 8821 TAGTTGACCC CAGCCCTTGA CCCCTAACAG GGGCAATTGC TGCCCTCCTG ATAACATCAG
 8881 GCCTCGCAAC CTGATTTCAT TTTCGCTCAA CAACCTTAAT AACCTTAGGA ACAGCCCTAC
 8941 TGCTTCTTAC AATATATCAG TGATGACGAG ACATCGTACG AGAGGGTACA TTCCAAGGAC
 9001 ACCACACGCC CCCCGTACAA AAAGGTCTTC GATACGGAAT GATTCTTTTC ATTACCTCCG
 9061 AAGTATTCTT TTTCCTAGGA TTCTTCTGAG CCTTTTACCA CGCGAGCCTC GCCCCTACTC
 9121 CTGAGCTAGG GGGCTGCTGA CCTCCCACAG GCATTACAAC TCTTGACCCA TTTGAAGTCC
 9181 CCCTCCTTAA TACAGCTGTC CTGCTTGCTT CTGGAGTAAC GGTCACCTGA GCCCACCACA
 9241 GCATTATGGA AGGTGAACGA AAACAGACCA TCCAATCACT AGCCTTAACC ATTCTTCTAG
 9301 GCTTTTATTT TACATTTCTT CAAGCCCTGG AATACTATGA AGCCCCCTTT ACAATTGCAG
 9361 ATGGCGTATA CGGCTCTACA TTTTTCGTAG CCACTGGATT CCACGGACTA CACGTTATTA
 9421 TTGGCTCCAC ATTTTTAGCT GTCTGCCTCC TACGACAAAT CCAATACCAC TTTACATCCG
 9481 AGCACCACTT CGGGTTCGAA GCAGCTGCCT GATACTGACA TTTCGTAGAC GTTGTCTGAT
 9541 TATTCCTATA TATCTCTATC TACTGATGAG GCTCTTAATC TTTCTAGTAT TAAAACTAGT
 9601 ATAAGTGACT TCCAATCACC CGGTCTTGGT TAAAATCCAA GGAAAGATAA TGAATGTAGC
 9661 TATAGCTGTA ATTACCATCA CTATTTTGCT TTCCGTAGTC CTGGCCATTG TATCCTTCTG
 9721 ACTCCCCCAG ATGACCCCCG ACCATGAAAA GCTCTCCCCA TATGAATGTG GTTTCGACCC
 9781 CTTGGGATCA GCCCGCCTAC CATTTTCCCT CCGCTTCTTC CTAGTCGCCA TTCTCTTCCT
 9841 CCTTTTCGAT TTAGAAATTG CCCTTCTCCT CCCACTCCCC TGAGGGGACC AATTAACCTC
 9901 CCCTTTATTG ACACTCTTCT GAGCCGTGGC CGTGCTTATT CTTCTTACCC TTGGCCTAGT
 9961 TTACGAGTGA ATTCAAGGAG GATTAGAATG AGCCGAATAG CCAATTAGTT TAAGAAAAAT
10021 ATTTGATTTC GGCTCAAAAG CTTATGGTTA AAGTCCATAA TTGTCTAATG ACTCCCGCTC
10081 ACTTCGCTTT CTCATCGGCC TTTACCCTAG GACTGACAGG CCTAGCATTC CATCGAACCC
10141 ACCTCCTCTC TGCTCTTTTA TGCTTAGAAG GGATGATACT CTCTTTATTT ATTGGACTTT
10201 CGATTTGAAC CCTCCAACTA GGGTCCACAA GTTTTTCTGC GGCTCCTATA CTCCTATTGG
10261 CTTTTTCAGC TTGTGAAGCA AGCGCAGGGC TTGCTTTACT GGTAGCCACA GCTCGCACGC
10321 ATGGTTCAGA TCGCCTTCAA ACCTTAAACC TCTTACAATG CTAAAAATTC TAATTCCCAC
10381 CCTAATGCTT CTTCCCACAG CCTGACTTGC CCCTGCCAAA TGATTATGAC CTACTACCTT
10441 ATCCCACAGC CTAGTTATTG CGCTGGCCAG CCTCACTTGA CTAAAGAATA CATCTGAAAC
10501 AGGCTGATCT TGCCTCACGC CCTTCATAGC CACAGATCCC CTCTCAACAC CTCTCCTTGT
10561 TCTTACCTGC TGACTACTTC CTCTTATAAT TTTGGCAAGC CAAAGCCACA CAGCACTAGA
10621 ACCTATTAAC CGCCAACGAA CCTACATTAG CCTGTTAACG TCTCTGCAAG TATTCCTTAT
```

(Figure 5. Continued)

```
10681 TATAGCATTC GGTGCTACTG AACTCCTTAT GTTTTATGTT ATATTTGAAG CTACTCTTAT
10741 CCCCACACTA ATTATTATTA CTCGGTGGGG CAACCAGGCA GAACGCCTTA ATGCAGGAGT
10801 ATATTTTTTG TTTTATACCC TAGCAGGCTC TCTCCCATTA CTAGTTGCCC TCTTGCTTCT
10861 TCAAAAGGAT ACAGGCTCCC TCTCCCTCTT AACCATCCAA TATACTAGCT CTACCCCTCT
10921 TTCATCTTAT GCTGATAAAC TTTGATGAGC AGGTTGCCTA ATTGCATTTT TAGTAAAAAT
10981 ACCCTTATAT GGAGCACATC TCTGGCTACC AAAAGCACAT GTAGAAGCCC CAGTTGCAGG
11041 CTCAATGGTT CTAGCTGCAG TTCTTCTAAA ACTAGGGGGC TACGGTATGA TCCGAATAAT
11101 AGTCATATTA GAACCTCTCA CCAAGGAATT AAGCTATCCC TTTATTGTCC TAGCGCTCTG
11161 AGGTGTAATT ATAACTGGCT CCACCTGCCT TCGCCAAACA GATCTTAAAT CCCTCATCGC
11221 CTATTCATCC GTAAGCCATA TGGGCCTGGT CGTTGGAGGG ATTCTTATCC AGACACCCTG
11281 AGGTCTTGCC GGCGCTGTAA TCCTTATGAT TGCACACGGC CTGACGTCCT CGGCCCTCTT
11341 CTGCTTGGCC AATACAAACT ATGAACGCCT CCATAGCCGA ACAATACTAT TAGCCCGGGG
11401 ATTACAGATA GTGCTTCCAC TCATAACAAC ATGATGATTT ATTGCCAGCC TCGCAAACTT
11461 AGCCCTTCCC CCTCTGCCTA ACCTCATGGG GGAACTTTTA ATTATTACCT CATTATTTGG
11521 CTGATCATGA TGAACTCTCG TACTCACAGG GACAGGGACC CTTATTACCG CGAGCTATTC
11581 ACTTTATATG TTCCTCATGA CCCAACGGGG TCCCCTCCCA GCACATATTA TTAGCCTAAA
11641 CCCCTCCTAT ACCCGGGAAC ACCTAGTTAT AGCCCTTCAC CTCCTCCCGC TACTTCTACT
11701 TGTTTTAAAG CCCGAATTAG TATGAGGCTG AACCACCTGT AGATATAGTT TAACAAAAAT
11761 ATTAGATTGT GATTCTAAAG ACAGAGGTTA AAATCCCCTT ATCCACCGAG AGAGGCTCGC
11821 CAGCAACGAA GACTGCTAAT CTCCGCGACC TTGGTTGGAC CCCAGGGCTC ACTCGGCCTG
11881 CTCCTAAAGG ATAACAGCTC ATCCATTGGT CTTAGGAACC AAAAACTCTT GGTGCAAATC
11941 CAAGTAGCAG CTATGCACTC CTCATCACTT ATTATATCAT CCAGCTTAGT CATTATCTTT
12001 TTACTATTAG CATATCCTAT CTTTACGACT CTGGAGCCTC GCCCCCGAAA CCCTGAATGG
12061 GCCGTTTCAC ATGTTAAGAC AGCGGTAGCC CTGGCCTTCT TCGTCAGCCT AATCCCCCTA
12121 TTTCTCTTTC TTAACGAAGG CGCAGAAGCA ATCATCACCT CATGAAATTG AATGAATACA
12181 CTAACCTTCG ACGTGAACAT TAGTTTTAAA TTTGATCATT ACTCAGTTAT CTTTGTCCCC
12241 ATTGCCCTCT ACGTCACTTG ATCTATTCTA GAGTTTGCAT CATGATATAT ACACGCAGAC
12301 CCATACATAA ACCGATTCTT TAAATATCTC CTAATTTTCC TTATTGCCAT AATCATTCTT
12361 GTCACAGCAA ATAATCTATT CCAGCTTTTC ATTGGTTGGG AAGGAGTAGG CATTATGTCA
12421 TTTCTACTCA TCGGCTGATG ATACGGACGA GCGGATGCCA ACACAGCGGC CCTTCAGGCC
12481 GTTGTGTATA ATCGGGTCGG AGACATTGGA CTGCTATTCA CAATAGCATG AATAGCAACC
12541 AACGCTAACT CCTGAGAGCT ACAACAAATT TTTGTGGCAA CAAAAGACCT GGATCTTACT
12601 TTACCGCTAC TAGGCCTAAT TATTGCCGCT ACAGGCAAGT CGGCCCAATT TGGTCTCCAC
12661 CCTTGACTTC CCTCTGCTAT AGAGGGTCCT ACACCGGTCT CTGCCCTACT ACATTCAAGC
12721 ACCATGGTTG TTGCCGGCAT TTTTCTCTTG GTACGAACAA GCCCCCTCCT AGAGAATAAT
12781 CAAACTGCCC TCACTACCTG CCTATGCCTA GGTGCCCTAA CGACCCTATT TACAGCCACC
12841 TGTGCCTTAA CCCAAAATGA TATCAAAAAA ATCGTAGCAT TCTCCACATC AAGCCAACTT
12901 GGCCTAATAA TAGTTACCAT CGGCTTAAAT CAACCTCAAC TAGCCTTCCT CCACATCTGC
12961 ACCCATCCCT TCTTCAAAGC AATGTTGTTC CTCTGTTCCG GCTCAATTAT CCACAGCCTC
13021 AACGACGAAC AAGATATCCG AAAAATAGGG GGCATACACC ACCTTACCCC CTTTACATCC
13081 TCCTGCCTCA CTATTGGTAG TTTAGCCCTC ACAGGTACTC CCTTCCTAGC AGGATTCTTC
13141 TCCAAAGATG CCATTATTGA GGCACTAAAC ACATCCCACC TAAACGCCTG AGCCCTAGTC
13201 CTAACCCTTC TAGCCACCTC ATTCACAGCC ATCTACAGCC TCCGCGTGGT ATTTTTGTCC
13261 TCAATAGGCC ACCCACGATT TAACGCTATT TCTCCTATCA ATGAGAACAA CCCAGCAGTT
13321 ATTAACCCCT TAAAGCGACT TGCATGAGGA AGCATTGTCG CTGGCCTCCT AATCATTTCA
13381 AGCATTACCC CCCTCAAGAC CCCTGTGATA TCTATACCCC CCTTGCTCAA ACTAGCTGCC
13441 CTCGGAGTTA CAATTACGGG GCTACTCATC GCCCTCGAAC TAGCCACACT AACCAATAAA
13501 CAGTACAAAA TTACCCCCAA CCTAGTTACC CACCACTTCT CCAACATGCT AGGCTTCTTC
13561 CCCTCGATTA TTCACCGATT TACCCCTAAA CTAAATCTAG TCTTAGGGCA GACATTTGCC
13621 AGTCAACTAA TTGACCAAAC TTGACTAGAG AAAGTCGGCC CCAAAGCAAT CTCTTCATCA
13681 AACATCCCCC TAATTACAAC AACAAGCAAC ACACAACAAG GAATAATTAA GACATACCTC
13741 ACCCTATTCC TTCTCACCCT GACCCTTGCT GCCCTATTGT TTACCCGTTA AACTGCCCGA
13801 AGAGCCCCCC GACTTAGTCC TCGAGTTAAC TCCAACACAA CAAATAAAGT GAGAAGCAGA
13861 ACCCACGCAC TAAGTACCAA TATCCCTCCC CCTAGTGAGT ACATTAGCGC AACCCCTCCA
13921 ATATCGCCTC GCAGAACAGA AAGCTCACTA AGCTCATCAG CCGGCACCCA TGAAGACTCA
13981 TATCACCCCC CTCAAGGCAC ATTACAAGCC ACCCCCACCC CCATGAAGTA CATCAACATG
14041 TCACTTACAA CAGGACCAAG CACTCAACTA TCAGGATAAG GATCAGCAGC AAGTGCCGCC
14101 GAGTATGCAA ACACAAATAA TATACCACCC AAGTAAATCA AAAACAACAC TAGTGATAGA
14161 AAAGGTCCCC CATGAGCAAC CAATACTCCA CACCCCATGC CCGCCACGAC TACTAACCCC
14221 AAGGCAGCAA AGTAGGGAGA AGGGTTAAAG GCAACTGCAA CCAACCCTAA AACTAATCCA
```

(Figure 5. Continued)

```
14281 ATTAAAAATA AAGACATAAT GTAAGTCATA ATTCCTGCCA GGACTTTAAC CAGAACCAAT
14341 GGCTTGAAAA ACCACCGTTG TTATTCAACT ACAAGAACCC ACTAATGGCA AGTCTACGAA
14401 AGACACACCC TCTCCTCAAA ATCGCAAACA ATGCCCTAGT TGACCTACCC GCCCCCTCAA
14461 ATATTTCAGT GTGATGAAAC TTCGGATCTC TCTTGGGACT CTGCTTAATT ATTCAAATCC
14521 TCACAGGACT ATTTTTAGCC ATGCACTACA CCTCTGATAT TGCTACGGCC TTTTCTTCCG
14581 TTGCTCATAT CTGCCGGGAC GTAAATTACG GCTGACTCAT CCGAAACCTC CACGCCAACG
14641 GTGCATCCTT CTTCTTTGTA TGCATCTATG CCCACATTGG CCGCGGACTT TACTACGGCT
14701 CATACCTCTA TAAAGAGACA TGAAACATTG GAGTAGTTCT GCTACTTCTA GTTATAATAA
14761 CTGCTTTCGT CGGCTATGTA CTACCCTGAG GCCAAATGTC CTTTTGAGGT GCAACCGTTA
14821 TTACTAACCT ACTCTCTGCA GTACCCTACG TAGGTGGCTC TCTGGTTCAA TGAATTTGAG
14881 GTGGATTCTC AGTAGACAAT GCAACCCTTA CCCGATTCTT TGCCTTCCAC TTCCTATTCC
14941 CCTTTGTAAT TGCAGGCGCA ACCATAGTCC ACCTTCTTTT CCTTCACCAA ACAGGGTCAA
15001 ATAATCCCCT CGGCCTTAAT TCAGACGCAG ATAAAATAAG CTTCCACCCC TATTTTTCAT
15061 ACAAAGACCT ATTAGGGTTT GCAGTACTTG TCATTGCCCT CACATGTTTA GCCTTATTCT
15121 CACCCAACCT GCTAGGAGAC CCAGACAACT TCACCCCCGC CAATCCGCTA GTCACTCCTC
15181 CCCACATTAA ACCAGAATGA TACTTCCTGT TCGCGTACGC AATTCTACGC TCCATCCCCA
15241 ATAAACTGGG GGGAGTCTTA GCCCTCCTGG CCTCAATCCT TATTCTGATA CTCGTACCAT
15301 TTCTACACAC GTCTAAACAA CGAAGCCTCA CTTTCCGACC ACTTACACAA TTCTTGTTTT
15361 GAACCCTAAT CGCAGACGTT ATTATTCTCA CTTGAATTGG AGGAATGCCC GTGTCACACC
15421 CATTCGTCAT TATTGGACAA ATTGCATCCT TTTTATACTT TTTCCTCTTC CTAGTCCTTA
15481 CGCCACTAGC AGGCTATGCA GAAGATAAAG CACTTGAATG AGCTTGCATT AGTAGCTCAG
15541 CGCCAGAGCC CTGGTCTTGT AAACCAGATG TCGGAGGTTA AAATCCTCCC TAATGCTCAA
15601 AGAAAGGAGA TTTTAACTCC CACCCCTGGC TCCCAAAGCC AGGATTCTTA GTTAAACTAT
15661 TCTTTGTATT ATATGTATAA TAATTTTAAA TACATATATG TATTATCAAC ATTAATTTAT
15721 ATTAACCATA TCATATAGCA TTCAAGTACA TACATGTATT ATCACCATAT CTAGGATTTA
15781 ACCATTCAAG AATTACACGA AACGAAAAAT CTTACGCAAA ACATAACAAC AAAAATCAAT
15841 AAACACTTAA AAATACCAGG CGACATTTAA AACCTAACAC AAACCCCCAT GAGTCAAGTT
15901 ATACCTTTAC TCAAATTCCC GTCAAACTCA AATATTTAAT GTAGTAAGAG CCGACCAACA
15961 AGTCCATTTC TTAATGCCAA CGGTTATTGA AGGTGAGGGA CAAAAACTGT GGGGGTTTCA
16021 CACGGTGATC TATTCCTGGC ATTTGGTTCC TATTTCAGGG CCATAAATTG TAAACATCCC
16081 CATAACTTAT TCTAAAAGGC ATAAGTTAAT GGTGGAGAAC AATAGCGGGA GCGGCCACCA
16141 TGCCGAGCGT TCTTTCCATA GGGCATTTAG CTCTTTTTTT TTTTTTTCCT TTTCAATAGA
16201 CATTTCACAG TGCACGCAAT CTAGTTAACA AGGTGGGAAT AATCCTAGGA AGCAAGGAAA
16261 TAGTATGCGT GATGAAAAGT CTTAACTAAA GAATTACATA TAGAACTTTC AAGGACATAA
16321 GATAGTGAAA TTTAGTCGGA AGATATCTAT ATTACCCCCT TTTGGCTTTT TCGCGTTAAA
16381 CCCCCCTACC CCCCTAAACT CCTGAGATAA CTAACGCTCC TGTAAACCCC CCGGAAACAG
16441 GAAAACCTCG AGTCGTTTTT ATGGT
```

(Figure 5. Continued)

Table 3. Summary of the mitochondrial genome of *S. owstoni*

*H and L denote heavy and light strands, respectively.

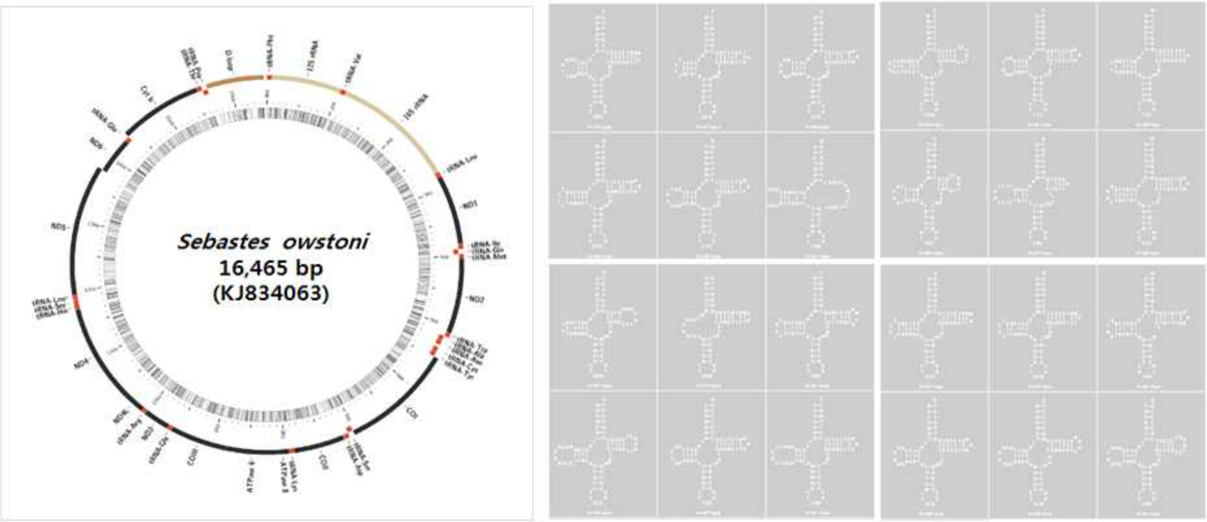


Figure 6. A gene arrangement in the complete mitogenome *Sebastes owstoni*.

노랑볼락의 미토게놈은 13개의 단백질 암호화 영역, 2개의 리보좀, 22개의 tRNA 및 1개의 조절영역으로 이루어져 있었고, 전체 크기는 16,450 bp 이었다(GenBank Accession number KJ834060)(Figure 7). 노랑볼락의 미토콘드리아 DNA의 염기 조성비는 27.7% (A), 26.6% (T), 17.3% (G) 및 28.4% (C)이었고, AT 비율이 약간 높았음(54.3%), COX1 (GTG)을 제외한 모든 단백질 암호영역의 개시코돈은 ATG이었고, 종결코돈은 TAA가 주를 이루었으나 ND1과 ND3는 TAG, ND4는 AGA였으며 Cytb는 종결코돈이 불완전(T)하였다(Table 4). tRNA형태는 tRNA$^{Ser(AGY)}$(lacked DHU-arm 구조)을 제외하고 모두 전형적인 cloverleaf 구조로 되어 있었다(Figure 8).

```
   1 GCTAGCGTAG CTTAATTAAA GCATAACACT GAAGATGTTA AGATGGACCC TAGAAAGTCC
  61 CGCCCGCACA AAGGCTTGGT CCTGACTTTA CTATCAACTT TAGCCAAATT TACACATGCA
 121 AGTATCCGCC CCCCTGTGAG AATGCCCTAC AGCTCCCTGC CCGGGAGCAA GGAGCTGGTA
 181 TCAGGCACAC ATTTGTAAGC CCATGACACC TTGCTTAGCC ACACCCTCAA GGGAACTCAG
 241 CAGTGATAAA CATTAAGCCA TAAGTGAAAA CTTGACTTAG TTAAAGCTAA GAGGGCCGGT
 301 AAAACTCGTG CCAGCCACCG CGGCTATACG AGAGACCCAA GTTGATACCA TTCGGCGTAA
 361 AGAGTGGTTA TGGAAAATAA AGACTAAAGC CGCACACCTT CAAAGCTGTT ATACGCATCC
 421 GAAGGCTAGA AGATCAACCA CGAAGGTAGC TTTACAACCC CTGACCCCAC GAAAGCTCTG
 481 GCACAAACTG GGATTAGATA CCCCACTATG CCTAGCCCTA AACCTTGGTA ATATATCACA
 541 TACCCTACCC GCCTGGGAAC TACGAGCACC AGCTTAAAAC CCAAAGGACT TGGCGGTGCT
 601 TTAGACCCCC CTAGAGGAGC CTGTTCTAGA ACCGATAACC CCCGTTCAAC CTCACCCTTC
 661 CTTGTTTATC CGGCCTATAT ACCGCCGTCG TCAGCTTACC CTGTGAAGGC CTAAAAGTAA
 721 GCACAACTGG CACAACCCAA AACGTCAGGT CGAGGTGTAG CGCATGGAGG GGGAAGAAAT
 781 GGGCTACATT CCCTATATTA GGGAACACGA ACGGCGCACT GAAACACGCG CCTGAAGGAG
 841 GATTTAGTAG TAAGCGGGAA ATAGCGTGTT CCGCTGAAAT CGGCCCTGAA GCGCGCACAC
 901 ACCGCCCGTC ACTCTCCCCA AGCCTATCAC TTTAAATAAT TAAAAACCCA AAAATCGCGG
 961 AGGGGAGGCA AGTCGTAACA TGGTAAGGGT ACCGGAAGGT GCACTTGGTA ATATCAGAGT
1021 GTAGTTAAAA TAGAATAACA CTTCCCTTAC ACTGAAGAGA CACCCGTGCA AATCGGATCA
1081 CCCTGACGCC CAACAGCTAG CCCACAAACA CAACAACAAC CAACCATTAT TTATAACCCC
1141 AAATGCACGA GTGTTTTAAT TAAACAAACC ATTTTTCCCC TTTAGTATGG GCGACAGAAA
1201 AAGGACTTAG GAGCAATAGA GAAAGTACCG CAAGGGATCG CTGAAAGAGA AATGAAACAA
1261 CCCAGTGAAG CTAAGTAAAG CAGAGATTTA TTCTCGTACC TTTTGCATCA TGATTTAGCC
1321 AGCGTGACCC AAGCAAAGAG TGCTTTAGTT TGACACCCCG AAACTAGGGG AGCTACTCCA
1381 AGACAGCCTA TTTATAGGGC GAACCCGTCT CTGTGGCAAA AGAGTGGAAT GAGCTTTGAG
1441 TAGAGGTGAT AAACCTACCG AACCTAGTTA TAGCTGGTTG CCCGGGAAAT GGATAGAAGT
1501 TCAGCCTCTC AGATTCTTTA TTCACCTCAG TATTACCCCA CCTGATACCA CAAGATAACT
1561 GTGAGAGTTA TTCAAAGGGG GTACAGCCCC TTTGAAACAA GATACAACTT TTCCGGGAGG
1621 AAAAAGATCA TAATTAAATA AAGGTAAGTA TTTGGGTGGG CCTAAAAGCA GCCATCCCAG
1681 TAGAAAGCGT TATAGCTCAA ATACATCACT ACCCCTCTCT ATCCTGATCG TTAATTCTTA
1741 CTCCCCCCTT CCCTACCGGG CCATCCCATG CAGACATGGG AGGGACCCTG CTAATATGAG
1801 TAATAAGAGA GCCAAGCCTC TCTCCTTGCA TACGTGTAAT TCGGAACGAA CCCGCACCGA
1861 GCATTAACGA CCCCAAACGA AGAGGGACCT GAACAACAAC CCAAACAACC AGAAAAAAAT
1921 TCAAACATAA ACCGTTAACC CTACACAGGT ATGCACCTCA GGAAAGACTA AAAGAAAGAG
1981 AAGGAACTCG GCAAACAAAT CAAGCCTCGC CTGTTTACCA AAAACATCGC CTCTTGCAAA
2041 GCTAAAGAAT AAGAGGTCCC GCCTGCCCTG TGACTATTAG TTTAACGGCC GCGGTATTTT
2101 GACCGTGCAA AGGTAGCGCA ATCACTTGTC TTTTAAATGA AGACCTGTAT GAATGGCACA
2161 ACGAGGGCTT AACTGTCTCC TCTTTCAAGT CAATGAAATT GATCTCCCCG TGCAGAAGCG
2221 GGGATATAAA CATAAGACGA GAAGACCCTA TGGAGCTTTA GACACCAAAG AAGATCCTGT
2281 CAAGTAACCC CTTATAAGGG CCTGAACTAA TGGAATCCTT CCCTAATGTC TTTGGTTGGG
2341 GCGACCGCGG GGAAACAAAA AACCCCCACG TGGAAAGGGA GCACCCACTC CTACAACTAA
2401 GAGCCGCAGC TCTAATTAAC AGAATATCTG ACCAATAAGA TCCGGCAATG CCGATCAACG
2461 GACCGAGTTA CCCTAGGGAT AACAGCGCAA TCCCCTTTTA GAGCCCATAT CGACAAGGGG
2521 GTTTACGACC TCGATGTTGG ATCAGGACAT CCTAATGGTG CAGCCGCTAT TAAGGGTCCG
2581 TTTGTTCAAC GGTTAAAGTC CTACGTGATC TGAGTTCAGA CCGGAGTAAT CCAGGTCAGT
2641 TTCTATCTAT GGTGTGCTCT TTTCTAGTAC GAAAGGACCG AAAAGAAGAG GCCCCTGCTC
2701 TAAGCAAGCC TCACCCCCAC CTAGTGAAGA CAACTAAAGT AGGCAAGAGG GCATACCCCC
2761 CGTGCCTGAG AGAACGGCAT GTTGGGGTGG CAGAGCCCGG TAAATGCAAA AGACCTAAGC
2821 CCTTTTTACA GAGGTTCAAG TCCTCTCCTT AACTATGATT TCAGTCCTTA TTACCCATAT
2881 TCTTAACCCC TTGGCCTTCA TCGTCCCCGT CCTCTTAGCC GTCGCCTTCC TCACGCTTCT
2941 AGAACGTAAG GTACTAGGGT ATATACAACT ACGAAAGGGT CCAAATATTG TAGGACCTTA
3001 CGGGCTGTTA CAGCCTATCG CCGATGGTGT GAAGCTCTTT ATTAAGGAGC CTGTTCGCCC
3061 CTCCACTTCC TCTCCCGTAC TTTTCCTCCT CGCCCCCCTA CTCGCACTTA CGCTTGCCTT
3121 GACCCTTTGA GCCCCCATGC CTCTCCCATA CCCAGTCATT GACTTGAACC TTGGGATTCT
3181 ATTTATTTTG GCCCTATCAA GCCTCGCTGT CTACTCTATT CTAGGCTCAG GCTGAGCATC
3241 CAATTCAAAA TATGCCCTCA TCGGGGCCCT TCGGGCTGTA GCCCAAACCA TCTCATATGA
3301 GGTTAGTCTA GGCCTAATCC TATTAAGTAC TATTATTTTT ACAGGGGGTT TTACCCTACA
3361 AACCTTCAAC ATTGCTCAAG AGAGCGTCTG AATACTACTC CCAGCTTGAC CACTAGCCGC
3421 AATATGATAT ATTTCAACCC TTGCGGAGAC AAACCGTGCA CCTTTTGACC TTACTGAAGG
3481 CGAATCCGAA CTAGTCTCTG GCTTCAATGT CGAATATGCA GGTGGCCCAT TCGCCCTATT
3541 TTTCTTGGCC GAATATGCTA ATATTCTACT TATAAATACG CTTTCCGCCA CCCTCTTCTT
```

Figure 7. Sequences of the mitogenome of S. *steindachneri* (GenBank Accession No. KJ834060)

```
3601 AGGGGCCTCT CATTTTCCAA CACTACCTGA ACTCACCGCA GTAAACCTAA TAGTCAAAGC
3661 GGCCCTTCTG TCTGTCTTAT TTTTATGAGT TCGAGCCTCT TACCCACGAT TCCGCTACGA
3721 TCAACTCATA CATCTAATTT GAAAAAACTT CCTCCCGCTT ACACTGGCCC TGGTTATTTG
3781 ACACCTAGCC CTCCCCATTG CATTTGCTGG CTTGCCACCC CAGCTATAGA TAAGAAGCCG
3841 TGCCTGAAGT AAAGGGCCAC TTTGATAGAG TGACTTATGG GGGTTCAAAT CCCCCCCGCT
3901 TCTTAGAAAA GGGGGACTCG AACCCCGCCT AAGGAGAGCA AAACTCCTGG TGCTCCCACT
3961 ACACTATTTC CTAGTAAAGT CAGCTAATTC TAAGCTCTTG GTCCCATACC CCAAACACGA
4021 AGGTTAAAAT CCCTCCTTTG CTAATGAACC CTTACATCTT AACCGCCCTG CTATTTGGTA
4081 TTGGTTTAGG CACTACTACC ACCTTCGCAA GCTCCCACTG ACTACTCGCC TGAATAGGCC
4141 TGGAAATAAA TACTCTTGCC ATTATTCCTC TAATAGCTCA ACACCATCAC CCCCGGGCAG
4201 TTGAAGCAGC CACTAAATAT TTCTTGATTC AAGCTGCCGG AGCAGCTATA CTACTATTTG
4261 CCAGCACCAC CAACGCTTGA TTGACTGGGC AATGGGACCT TTTGCAAATT GCCCACCCTT
4321 TCCCAACTGT TCTTGTCACT TTGGCCCTCG CACTAAAGGT GGGACTTGCA CCTGTACACT
4381 CATGGCTACC TGAAGTACTT CAAGGCCTAG ACCTAACCAC AGGACTTATT TTGTCGACCT
4441 GACAAAAACT TGCCCCATTT GCCTTATTAG TCCAAACTCC CTGTGCCAAC ACCACCCTTT
4501 TAATTATCCT CGGACTTACC TCAACCATTG TAGGAGGCTG AGGAGGTCTC AACCAAACCC
4561 AGCTTCGCAA GATTCTTGCC TACTCCTCCA TCGCACACCT CGGCTGGATA GTAATTGTAC
4621 TACAATTCTC TCCCTCCTTG ACTATTCTAA CACTACTCAC ATATTTTATT ATAACATTTT
4681 CAGCATTTCT TATGTTTAAA CTTAATAAAG CAACCAGCAT TAATGCTCTA GCAACCTCAT
4741 GGGCAAAGAC CCCCGCCCTA ACCGCCCTTG CACCCCTTCT ATTGTTATCC TTAGGAGGCC
4801 TCCCCCCACT TACAGGATTT ATGCCAAAGT GACTTATTCT TCAAGAGCTT GCTAAACAAG
4861 ACCTTGCCCC CGCCGCAACA CTGGCGGCAA TAACCGCCCT CCTTAGCCTA TATTTTTATC
4921 TACGACTATC ATACGCAATG GCACTAACTA TTTCGCCAAA TAACCTAACC GCAATTTCCC
4981 CATGACGCCT CCCCTCTTTA CAACTAACAA TGCCACTTGC TACCTCAGCC ATAGCTACGC
5041 TAATGCTTCT ACCCCTAACA CCCGCCGCAA TAGCACTAAT AACCCTTTAA GGGACTTAGG
5101 TTAAAACAAG ACCAAGGGCC TTCAAAGCCC TAAGTGAGGG TGGAAGTCCC CCAGTCCCTG
5161 ATAAGGCTTG CGGGACACTA CCCCACATCT CCTGTATGCA AAACAGGTAC TTTAATTAAG
5221 CTAAAGCCTT CCTAGAAGGG CAGGCCTCGA TCCTGCAAGA TCTTAGTTAA CAGCTAAGCG
5281 CTCAAACCAG CGAGCATCCA TCTATCTTTC CCCCGCCTGA GGGACGGGCG GGGCGGGGGA
5341 AAGTCCCGGC AGACGACTAG TCTGCATCTT CAGATTTGCA ATCTGATATG TAAAACACCT
5401 CAAGACTTCT GGTAAGAAGA GGACTCAAAC CTCTGTTTGT GGGGCTACAA TCCATCGCTT
5461 AAAAACTCAG CCATCCTACC TGTGGCCATC ACACGTTGAT TTTTCTCCAC TAATCACAAA
5521 GACATCGGCA CCCTTTATCT AGTATTTGGT GCCTGAGCCG GTATAGTAGG CACAGCCCTC
5581 AGCCTACTCA TTCGAGCAGA ACTAAGCCAA CCGGGCGCTC TCCTTGGAGA CGACCAAATT
5641 TATAATGTAA TTGTTACAGC ACATGCTTTC GTAATGATTT TCTTTATAGT AATGCCAATT
5701 ATAATTGGAG GTTTTGGAAA CTGATTAATT CCCCTAATGA TTGGAGCCCC AGATATAGCA
5761 TTTCCTCGTA TGAATAACAT AAGTTTCTGA CTTCTTCCCC CTTCTTTCCT ACTACTACTT
5821 GCCTCTTCTG GAGTAGAAGC GGGTGCCGGA ACCGGGTGAA CAGTGTACCC GCCCCTGGCT
5881 GGTAACTTAG CCCACGCAGG AGCATCAGTC GACCTGACAA TCTTTTCACT TCACCTAGCA
5941 GGTATTTCCT CAATCCTAGG GGCAATCAAT TTTATTACCA CAATTATTAA TATGAAGCCC
6001 CCGGCTATCT CTCAGTACCA GACACCCCTA TTTGTGTGAG CTGTCCTAAT TACCGCTGTT
6061 CTTCTCCTTC TCTCTCTACC AGTTCTCGCT GCCGGCATCA CAATGCTCCT TACCGACCGA
6121 AATCTTAATA CCACCTTCTT TGACCCGGCA GGAGGAGGGG ATCCAATCCT TTATCAACAC
6181 TTATTCTGGT TTTTTGGACA CCCGGAAGTA TATATTCTTA TTCTGCCTGG CTTTGGTATG
6241 ATTTCACACA TCGTCGCCTA TTATTCTGGC AAAAAAGAAC CCTTTGGCTA TATAGGCATA
6301 GTATGAGCAA TAATGGCTAT TGGCCTCCTA GGCTTTATTG TATGAGCTCA TCACATATTT
6361 ACAGTTGGCA TGGACGTAGA CACGCGTGCT TATTTCACAT CTGCCACAAT AATCATCGCA
6421 ATTCCCACCG GTGTTAAAGT ATTTAGCTGA CTTGCGACCC TACATGGGGG CTCTATTAAA
6481 TGAGAAACAC CCCTTTTATG AGCCCTTGGC TTTATTTTCC TGTTTACAGT AGGAGGGCTT
6541 ACAGGCATTG TTCTGGCCAA TTCATCTCTA GATATTGTAC TCCACGATAC CTATTATGTA
6601 GTAGCCCACT TCCACTACGT ACTATCTATG GGGGCCGTGT TTGCCATTGT CGCCGCCTTC
6661 GTGCACTGAT TCCCGCTATT TTCAGGCTAC ACGCTTCACA GCACTTGAAC AAAAATCCAC
6721 TTCGGTATTA TGTTCTTAGG GGTAAACTTA ACCTTCTTCC CACAACATTT CCTCGGATTA
6781 GCCGGAATGC CCCGACGATA CTCCGACTAC CCTGACGCCT ATACCCTATG AAATACAGTC
6841 TCCTCAATCG GGTCACTTAT CTCCCTAGTG GCTGTCATCA TGTTCTTATT TATTATTTGA
6901 GAGGCATTCG CCGCCAAACG TGAAGTTCTA GCAACAGATT TAACAACAAC CAATGTAGAA
6961 TGACTACATG GCTGCCCTCC CCCATACCAC ACATTCGAGG AGCCTGCCTT TGTACAAGTA
7021 CAAGCAGACT AACGAGAAAG GGAGGAGTCG AACCCCCATA GGTCAGTTTC AAGCCGACCA
7081 CATAACCGCT CTGCCACTTT CTTTATAAGA CACTAGTAAA AGAGTACATT ACACCGCCTT
```

(Figure 7. Continued)

```
 7141 GTCAAGGCGG AAGTGTGGGT TAGACCCCCG CGTGTCTTGC TTTTAATGGC CCATCCGTCA
 7201 CAGCTTGGAT TTCAAGATGC AGCTTCACCT GTTATAGAAG AACTTCTTCA TTTTCACGAC
 7261 CATGCTTTAA TAATCGTCTT CCTGATTAGC ACACTAGTGC TTTATATTAT TCTTGCTATA
 7321 GTTACCACTA AATTAACGAA CAAATATATT TTAGATTCAC AAGAGATTGA AATTATCTGA
 7381 ACAATTCTCC CAGCTATCAT TTTAATTCTG ATTGCACTCC CTTCCCTTCG CATCCTCTAT
 7441 CTTATAGATG AGATTAACAA CCCTTTATTA ACAATTAAAG CCGTTGGCCA CCAGTGGTAC
 7501 TGAAGCTATG AATATACTGA CTACGAAGAT CTTGGCTTTG ATTCATATAT AATTCCCACC
 7561 CAAGACCTAA CCCCTGGACA ATTCCGCCTA TTAGAAGCCG ACCATCGCAT GGTTATTCCA
 7621 GTTGAATCCC CCATCCGAGT CTTAGTCTCC GCAGACGATG TACTCCACTC ATGAGCAGTC
 7681 CCGGCCCTGG GAGTAAAAAT GGACGCAGTC CCAGGCCGCC TAAACCAAAC AGCCTTTATC
 7741 GCATCCCGAC CAGGCGTATT CTACGGACAA TGCTCTGAGA TCTGCGGAGC AAATCACAGC
 7801 TTTATACCTA TTGTAGTGGA AGCAGTTCCC CTAGAACACT TTGAAAACTG ATCATCTCGA
 7861 ATACTTGAAG ACGCCTCGCT AGGAAGCTAA ATAGGGTATA GCGTTAGCCT TTTAAGCTAA
 7921 AGATTGGTGG CTCCCAACCA CCCCTAACGA CATGCCCCAA CTCAACCCCG CACCTTGATT
 7981 TGCTATTTTA GTCTTTTCGT GAATGGTCTT CCTGGCCGTT ATTCCCGCTA AAGTTACAGC
 8041 CCACACTTTC CCAAATACCC CTAATCTGCA AAGCGCAGAA AAAGCCAAAA CAGACCCCTG
 8101 AACTTGACCA TGACACTAAG CTTTTTTGAC CAGTTTATAA GCCCCACCTA TCTTGGGATC
 8161 CCATTAATAG CCCTTGCCCT TACCCTACCC TGACTCCTTT ACCCCACACC TACAACTCGA
 8221 TGATTAAATA ACCGATTCCT CGCGCTTCAA GGTTGATTTA TTAACCGTTT TACTCAACAG
 8281 CTTCTCCTCC CCTTAAATAT TGGAGGTCAT AAGTGAGCTG CCCTCCTAAC CTCATTAATG
 8341 ATCTTTTTAA TTACCCTAAA TATATTAGGA CTTCTTCCCT ACACTTTTAC TCCTACCACC
 8401 CAATTGTCAC TAAATTTAGG ACTTGCGGTA CCTCTCTGAT TAGCAACTGT TATTATTGGC
 8461 ATACGAAACC AACCAACCCA TGCCCTAGGA CACCTCCTAC CAGAAGGCAC ACCCGGCCCC
 8521 CTTATCCCCG TGCTTATCAT TATCGAAACA ATTAGCCTCT TTATTCGCCC TCTTGCCCTA
 8581 GGAGTACGAC TGACAGCCAA TTTAACAGCT GGTCACCTTT TAATTCAACT AATTGCTACA
 8641 GGCGCCTTCG TACTTCTCCC CTTAATACCA ACGGTGGCAA TCATCACAAC AACAGTACTG
 8701 GTTCTCCTTA CCCTATTAGA AGTTGCTGTA GCAATAATTC AAGCATACGT CTTCGTCCTC
 8761 CTACTAACAC TATACCTACA AGAAAACGTC TAATGGCCCA TCAAGCACAC CCTTACCACA
 8821 TAGTTGACCC CAGCCCTTGA CCCCTAACAG GGGCAATTGC TGCCCTACTG ATAACATCAG
 8881 GCCTCGCGAC CTGATTTCAT TTTCGCTCAA CAACCTTAAT AACCTTAGGA ACAGCTCTAC
 8941 TGCTTCTTAC AATATACCAA TGATGACGAG ATATCGTACG AGAAGGTACA TTCCAAGGAC
 9001 ATCATACGCC CCCCGTACAA AAAGGTCTTC GATACGGAAT AATTCTCTTC ATTACCTCCG
 9061 AAGTATTTTT TTTCCTAGGA TTCTTCTGAG CCTTTTACCA CGCAAGCCTA GCCCCCACTC
 9121 CTGAGCTAGG GGGCTGCTGA CCTCCAACGG GCATTACAAC TCTTGACCCA TTTGAAGTCC
 9181 CCCTCCTTAA TACAGCTGTC CTACTTGCCT CCGGGGTGAC GGTTACCTGA GCCCACCACA
 9241 GCATTATGGA AGGTGAACGA AAACAAACCA TTCAATCGCT AGCCTTAACT ATCCTTCTAG
 9301 GCTTTTATTT CACATTTCTT CAAGGCCTGG AATACTATGA GGCCCCCTTT ACAATTGCAG
 9361 ATGGCGTATA CGGCTCTACC TTTTTCGTAG CCACTGGCTT TCACGGACTA CATGTTATTA
 9421 TTGGCTCCAC ATTTTTAGCT GTTTGCCTCC TACGACAAAT CCAATACCAC TTTACATCCG
 9481 AGCACCACTT CGGGTTCGAA GCAGCTGCCT GATACTGACA TTTCGTAGAC GTTGTGTGAT
 9541 TATTCCTATA TATCTCTATC TACTGATGAG GCTCTTAATC TTTCTAGTAT TAAAACTAGT
 9601 ATAAGTGACT TCCAATCACC CGGTCTTGGT TAAAATCCAA GGAAAGATAA TGAACGTAGC
 9661 AATAGCTGTA ATTACCATCA CTATTTTGCT TTCCGTAATC CTGGCCGTTG TATCCTTCTG
 9721 GCTCCCCCAA ATGACCCCCG ATCACGAAAA GCTCTCCCCC TATGAATGTG GTTTCGACCC
 9781 CTTAGGATCA GCCCGCCTAC CATTTTCCCT CCGCTTCTTC CTAGTCGCCA TTCTTTTCCT
 9841 CCTTTTCGAT TTAGAAATTG CCCTTCTCCT CCCACTCCCT TGAGGGGACC AATTAACTTC
 9901 CCCCTTACTG ACACTCTTCT GAGCCGTGGC CGTGCTTATT CTTCTTACCC TTGGCTTGGT
 9961 TTACGAGTGA ATTCAAGGAG GTTTAGAATG AGCCGAATAG CCAATTAGTT TAAGAAAAAT
10021 ATTTGATTTC GGCTCAAAAG CTTATGGTTA AAGTCCATAA TTGTCTAATG ACTCCCGCTC
10081 ACTTCGCTTT CTCATCGGCC TTTACCCTAG GACTGACAGG CCTAGCATTC CATCGAACCC
10141 ACCTCCTCTC TGCTCTTTTA TGCTTAGAAG GGATGATGCT CTCTTTATTT ATCGGACTTT
10201 CGATTTGAAC CCTCCAACTA GGATCCACAA ATTTCTCTGC GGCTCCTATG CTCCTATTGG
10261 CTTTTTCAGC TTGTGAAGCA AGCGCAGGGC TTGCTTTACT GGTAGCCACA GCTCGCACGC
10321 ATGGTTCAGA TCGCCTTCAA ACCTTAAACC TCTTACAATG CTAAAAATCC TAATTCCCAC
10381 CCTAATGCTT CTTCCCACAG CCTGGCTTGC CCCTGCCAAA TGATTATGAC CTACTACCCT
10441 CTCCCACAGC CTAGTCATTG CACTGGCCAG CCTCACTTGA TTAAAAAATA CATCTGAAAC
10501 AGGCTGATCT TGCCTCACGC CCTTCATAGC CACAGACCCC CTCTCAACAC CCCTCCTTGT
10561 TCTTACCTGC TGACTACTCC CTCTTATAAT TTTGGCAAGC CAAAGCCACA CAGCACTAGA
10621 ACCTATTAAC CGCCAACGAA CCTACATTAG CCTATTAACG TCTCTGCAAG TATTCCTTAT
10681 TATAGCATTC GGTGCCACTG AACTCCTTAT GTTTTATGTT ATATTTGAAG CTACTCTCAT
```

(Figure 7. Continued)

```
10741 CCCCACACTA ATTATTATTA CTCGGTGGGG TAATCAGGCA GAACGCCTTA ATGCAGGAGT
10801 ATATTTTTTG TTTTATACCC TAGCAGGCTC TCTCCCGTTA CTAGTTGCCC TCTTGCTTCT
10861 TCAAAAGGAT ACAGGATCCC TCTCCCTCTT AACCATCCAA TATACTAGCT CTACCCCTCT
10921 TTCATCTTAT GCTGACAAAC TTTGATGAGC AGGCTGCCTA ATTGCATTTT TAGTAAAAAT
10981 ACCTTTATAT GGAGCACATC TCTGATTACC AAAAGCACAT GTAGAAGCCC CAGTTGCAGG
11041 CTCAATGGTT CTAGCTGCAG TTCTTCTAAA ACTAGGAGGC TACGGTATAA TCCGAATGAT
11101 AGTTATATTA GAACCTCTCA CCAAGGAATT AAGCTATCCC TTTATTGTCC TAGCCCTCTG
11161 AGGTGTAATT ATAACTGGCT CCACCTGCCT TCGCCAAACA GATCTTAAAT CCCTCATCGC
11221 CTATTCATCC GTAAGCCATA TGGGCCTGGT CGTTGGAGGT ATTCTTATCC AAACACCCTG
11281 AGGTCTTGCC GGCGCTGTAA TCCTTATGAT TGCACACGGC CTGACGTCCT CGGCCCTCTT
11341 CTGCTTGGCC AATACAAACT ATGAACGCCT CCATAGCCGA ACAATACTAT TAGCCCGGGG
11401 ATTACAGATA GTGCTTCCAC TCATAGCAAC ATGATGATTT ATTGCCAGCC TCGCAAACTT
11461 AGCCCTTCCC CCTCTGCCTA ACCTCATGGG AGAACTTTTA ATTATTACCT CATTATTTGG
11521 TTGATCATGA TGAACTCTCG TACTCACAGG GGCGGGGACC CTTATTACCG CGAGCTATTC
11581 ACTTTATATA TTCCTCATGA CCCAACGGGG TCCCCTCCCA GCACATATTA TTAGCCTCAA
11641 CCCCTCCTAC ACCCGGGAAC ACCTAGTTAT AGCCCTTCAC CTCCTCCCCC TGCTTCTACT
11701 TGTATTAAAG CCCGAATTAG TATGAGGCTG AACCACCTGT AGATATAGTT TAACAAAAAT
11761 ATTAGATTGT GATTCTAAAG ACAGAGGTTA AAATCCCCTT ATCCACCGAG AGAGGCTCGC
11821 CAGCAACGAA GACTGCTAAT CTCCGTGACC TTGGTTGGAC CCCAGGGCTC ACTCGGCCTG
11881 CTCCTAAAGG ATAACAGCTC ATCCATTGGT CTTAGGAACC AAAAACTCTT GGTGCAAATC
11941 CAAGTAGCAG CTATGCACTC CTCATCACTT ATTATGTCAT CCAGCTTAGT CATTATCTTT
12001 TTACTATTAG CATATCCTAT CTTTACGACC CTGGACCCTC GCCCCCGAAA CCCTGACTGG
12061 GCCGTTTCCC ATGTTAAGAC AGCGGTCGCC CTGGCCTTCT TCGTTAGCCT AATCCCTTTA
12121 TTTCTCTTTC TTAACGAGGG GGCGGAAGCA ATCATCACCT CATGAAATTG AATGAATACA
12181 CTAACCTTCG ACGTGAATAT TAGCTTCAAG TTTGACCACT ACTCAGTTAT CTTTGTACCC
12241 ATTGCCCTCT ACGTCACTTG ATCTATTCTA GAGTTTGCAT CATGGTATAT ACACACAGAC
12301 CCATACATAA ACCGATTCTT TAAATACCTC CTAGTTTTCC TTATTGCCAT AATTATTCTT
12361 GTCACAGCAA ACAATCTATT CCAACTTTTC ATTGGTTGGG AGGGAGTAGG CATTATATCA
12421 TTTCTACTCA TTGGCTGATG ATACGGACGA GCGGATGCCA ACACAGCGGC CCTTCAGGCC
12481 GTCGTGTATA ATCGGGTAGG AGACATTGGA TTGCTATTCA CAATAGCTTG AATAGCAACC
12541 AACGCTAACT CCTGAGAGTT ACAACAAATT TTTGTAGCAA CTAAGGACCT CGATCTTACC
12601 CTACCCCTAC TAGGCCTGAT TGTTGCCGCT ACAGGCAAGT CGGCCCAATT TGGTCTTCAC
12661 CCTTGACTCC CCTCTGCTAT AGAGGGTCCT ACACCGGTAT CTGCCCTACT GCATTCGAGC
12721 ACCATAGTCG TTGCCGGTAT TTTTCTCCTA GTACGAACAA GTCCCCTCCT GGAAAATAAT
12781 CAAACTGCCC TCACCACCTG CCTATGCCTA GGTGCCCTAA CGACACTATT TACAGCCACC
12841 TGTGCCCTAA CCCAAAATGA TATCAAAAAG ATCGTAGCAT TCTCCACATC AAGTCAACTA
12901 GGTCTAATAA TAGTTACTAT TGGCTTAAAT CAACCTCAAC TAGCCTTTCT CCACATTTGC
12961 ACCCATGCCT TCTTTAAGGC AATACTATTC CTCTGTTCTG GCTCAATTAT TCACAGCCTC
13021 AACGACGAAC AAGATATCCG AAAAATAGGA GGTATACATC ACCTTGCCCC TTTTACATCC
13081 TCCTGCCTTA CTATTGGTAG TTTAGCCCTC ACAGGCACCC CCTTCCTAGC AGGATTCTTC
13141 TCCAAAGATG CCATTATTGA GGCACTAAAC ACATCTCACC TAAACGCCTG AGCCCTAGTC
13201 CTAACCCTTC TAGCCACCTC ATTCACCGCC ATCTACAGTC TCCGCGTAGT GTATTTTGTT
13261 TCAATGGGCC ACCCACGGTT TAACGCTATT TCCCCCATCA ATGAAAATAA CCCAGCGGTT
13321 ATTAACCCCT TAAAGCGACT TGCATGAGGA AGCATTGTCG CTGGCCTCCT AATTATCTCA
13381 AGCATTACCC CCCTTAAGAC CCCCGTGATA TCTATACCTC CCTTGCTCAA ACTAGCTGCT
13441 CTTGTAGTTA CAATTATAGG ATTACTCATT GCCCTCGAGC TAGCAACACT CACCAATAAA
13501 CAGTACAAAG TTACCCCTAA TCTAGTTACC CATCACTTCT CCAACATGTT ACGCTTTTTC
13561 CCCTCGATTG TTCACCGATT TACCCCCAAA CTAAATCTAG TCTTAGGACA GACACTTGCC
13621 AGCCAACTGA TTGACCAAAC TTGACTAGAG AAAGTTGGTC CCAAAGCAAT CTCTTCATCA
13681 AATATTCCCC TAATTACAAC AACAAGCAAC ACACAACAAG GAATAATTAA GACGTACCTC
13741 ACCCTATTCC TTCTTACCCT GACCCTTGCT GCCCTATTAT TTACCCGTTA AACTGCCCGA
13801 AGGGTCCCCC GACTAAGTCC CCGAGTTAAC TCCAGCACAA CAAACAAGGT GAGAAGCAAG
13861 ACCCACGCAC TAAGTACTAA TAACCCTCCC CCTAATGAGT ACATTAACGC AACCCCTCCC
13921 ATATCGCCTC GCAAGACAGA GAGCTCACTA AGCTCATCAG CCGGCACCCA TGAAAACTCA
13981 TATCACCCCC CTCAAAATAT ACTAGAAGCC ACCCCCACCC CTACTAAGTA TATCAACATG
14041 TCACCTACAA CAGGACCACT TACCCAGCTT TCCGGGTAAG GCTCAGCGGC AAGTGCCGCC
14101 GAGTACGCAA ACATGACTAG TATGCCACCC AAATAAATCA AAAACAGCAC CAGTGATAGA
14161 AAGGGTCCCC CATGTCCAAC CAATACTCCA CACCCCATGC CCGCCACGAC TACTAACCCT
14221 AAGGCAGCAA AGTAAGGAGA AGGGTTAGAG GCAACTGCAA CCAACCCTAG AACTAACCCA
14281 ATTAAAAATA AAGACATAAT GTAAGTCATA ATTCCTGCCA GGACTTTAAC CAGAACTAAT
```

(Figure 7. Continued)

```
14341 GGCTTGAAAA ACCACCGTTG TTATTCAACT ACAAGAACCC ACTAATGGCA AGTCTACGAA
14401 AGACACACCC TCTCCTCAAA ATCGCAAACA ATGCCCTAGT TGACCTACCC GCCCCCTCAA
14461 ATATTTCAGT GTGATGAAAC TTCGGCTCTC TCTTGGGACT CTGCTTAATT ATTCAAATCC
14521 TCACGGGACT ATTTTTAGCC ATACACTATA CCTCTGATAT TGCTACAGCT TTTTCTTCCG
14581 TTGCTCATAT TTGCCGAGAC GTAAATTACG GGTGATTCAT CCGAAACCTT CACGCCAACG
14641 GTGCATCCTT CTTCTTTGTA TGCATCTATG CCCACATTGG CCGCGGACTT TACTACGGCT
14701 CATACCTTTA TAAAGAGACA TGAAACATCG GAGTAGTCCT ACTACTTCTA GTTATAATAA
14761 CTGCTTTCGT CGGTTATGTG TTACCCTGAG GCCAAATGTC CTTTTGAGGT GCCACCGTTA
14821 TTACCAACCT ACTCTCCGCA GTACCCTACG TAGGTAACGC CCTCGTTCAG TGAATTTGAG
14881 GTGGATTCTC AGTAGACAAT GCAACCCTTA CCCGATTCTT CGCCTTCCAC TTTTTATTCC
14941 CCTTTGTAAT TGCAGGCGCG ACCATAGTCC ACCTCCTTTT CCTTCATCAG ACAGGATCAA
15001 ATAATCCCCT CGGCCTAAAT TCAGATGCAG ATAAAATAAG CTTCCACCCC TACTTCTCAT
15061 ACAAAGACTT ATTAGGGTTT GCAGTACTTG TCATTGCCCT TACATGTCTA GCTTTATTTT
15121 CACCCAACCT GCTAGGAGAC CCAGACAACT TCACCCCCGC CAATCCACTA GTTACTCCAC
15181 CCCACATCAA GCCAGAGTGA TATTTTCTGT TCGCATACGC AATTCTACGC TCCATTCCCA
15241 ATAAACTAGG GGGAGTTTTA GCCCTCCTAG CTTCAATCCT TATTCTTATG CTAGTGCCCT
15301 TTCTACACAC GTCTAAACAA CGAAGCCTCA CTTTCCGACC ACTCACACAA TTCTTGTTTT
15361 GAACCCTGAT CGCAGACGTT ATTATTCTCA CCTGAATCGG GGGAATGCCC GTATCGCACC
15421 CGTTCGTCAT TATCGGACAA GTCGCATCCT TTTTATACTT TTTCCTCTTT CTAGTCCTTA
15481 CACCACTAGC AGGCTATGCA GAGGACAAAG CACTTGAATG AGCTTGCACT AGTAGCTCAG
15541 CGTCAGAGCC CTGGTCTTGT AAACCAGATG TCGGAGGTTA GAGTCCTCCC TACTGCTCAA
15601 AGAAAGGAGA TTTTAACTCC CACCCCTGGC TCCCAAAGCC AGGATTCTTA GTTAAACTAT
15661 TCTTTGTAAT ATATGTACAA TAATTTTATA TACATATATG TATTATCACC ATTAATTTAT
15721 ATTAACCATA TCATATAGCA TTCAAGTACA TATATGTATT ATCACCATAT CTAGGGTTTA
15781 ACCATTCAAG TGTTATATTA AACGAATAAT TTACATAAAG CAGAATAATA AAAAACAATA
15841 AACACTTATA AATACCGGGC GAAACTTAAG ACCTATCACA ATAACTCATA AGTCAAGTTA
15901 TACCTTTACT CAAAATCCCG CCAAACTCAA ATATTTAATG TAGTAAGAGC CGACCAACAA
15961 GTCCATTTCT TAATGCCAAC GGTTATTGAA GGTGAGGGAC AACTATTGTG GGGGTTTCAC
16021 ACAGTGATTT ATTCCTGGCA TTTGGTTCCT ATTTCAGGCC CACATATTGT AAACCCCCCC
16081 ATACGTCTAT CGTAGAAAGC ATAAGTTAAT GGTGGAAAAC AATAGCGGGA GCGGCCACCA
16141 TGCCGAGCGT TCTTTCCATA GGGCATTAAG TTTTTTTTTT TTTTTTTCCT TTTCAATGGA
16201 CATTTCACAG TGCACGCGAT CTGATTAACA AGGTGGGAAT AATCTTAGGA AGCAAGGAAA
16261 TAGTATGAGT GGTGAAAGGT CTTTACAAAA GAATTACATA TAAGGATTTC AAGGACATAA
16321 AGTAGTGAAA TTTAGTCGGA AGATATCTAT ATTACCCCCT TTTGGCTTTT TCGCGTTAAA
16381 CCCCCCTACC CCCCTAAACT CGTGAGATAA CTAACGCTCC TGTAAACCCC CCGGCCTTAG
16441 GAAAACTTCG
```

(Figure 7. Continued)

Table 4. Summary of the mitochondrial genome of *S. steindachneri*

*H and L denote heavy and light strands, respectively.

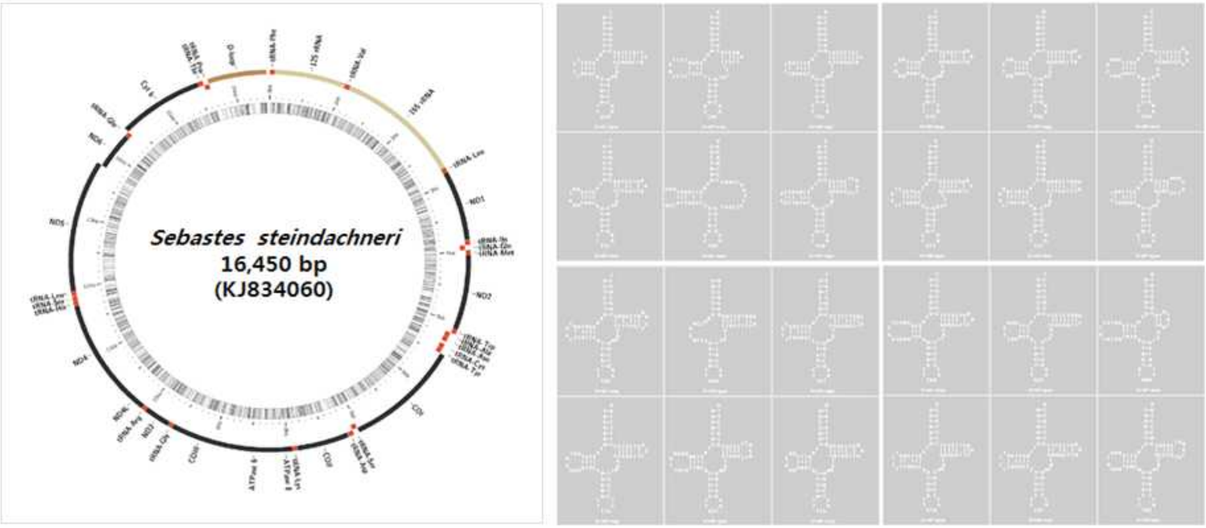


Figure 8. A gene arrangement in the complete mitogenome *Sebastes steindachneri*.

3. 결론 및 토의

동해서식 볼락류의 자원특성을 조사하고 유전특성을 파악하기 위하여 다양한 체색 패턴을 가진 볼락류를 대상으로 연구를 수행하였다. 우럭볼락, 흰꼬리볼락, 황볼락, 노랑볼락의 미토게놈 정보를 분석하여 염기서열 정보를 세계유전자은행인 GenBank에 등록하였고 분석정보는 논문으로 출판하였다. 본 연구를 통하여 확보한 4종의 볼락류 미토게놈 정보는 해양어류의 유전정보 뱅크화에 직접 사용이 가능하다.

제2절 해양어류자원의 기능유전체 정보 이용기술 탐색

1. 연구 재료 및 방법

가. 어류시료 확보

볼락류의 체색관련 기능유전체 분석을 위하여 체색이 다양한 개볼락을 임원항에서 구입하였다. 개볼락 시료는 반점이나 마크를 기준으로 color type과 wild type으로 구분하였으며, 체색관련 기능유전체 분석 시료로 활용하였다.

나. 개볼락 체색타입별 total RNA 추출

개볼락의 근육조직을 각각 100mg씩 취하여 호모게나이저를 사용하여 파쇄하였다. TRIzol® Reagent(Invitrogen, USA) 1㎖을 첨가하여 상온에서 5분간 방치하면서 완전히 섞어 준 다음, chloroform을 0.2㎖ 넣는다. 상온에 3분간 두었다가 원심분리(12,000x*g*, 4℃, 15분) 하였다. 상층액에 isopropyl alcohol을 첨가하여 상온에 10분간 두었다가 원심분리(12,000x*g*, 4℃, 10분)하여 RNA pellet만 취했다. 75% 에탄올로 세정 후에 RNA pellet을 건조시킨 후, RNase-free water를 첨가하여 충분히 녹인다.

추출한 total RNA는 Agilent사의 2100 BioAnalyzer를 이용하여 QC 과정을 거쳤으며, BAConcentration(ng/ul), RIN value, 28s/18s ratio 값을 활용하였다.

다. 개볼락 근육조직 RNA-seq 분석

개볼락 체색 타입별 근육조직 각각의 total RNA 2㎍을 oligo(dT)를 사용하여 mRNA를 분리·정제하였다. Library는 Paired-end 100 bp로 진행되었으며, Illumina사의 TruSeq RNA Sample Prep Kit을 이용하여 Library 준비를 진행하였다. 분리·정제된 mRNA는 fragmentation 단계를 거쳐 random hexamer primer를 사용하여 single-stranded cDNA로 합성하였다. 이를 주형으로 하여 double stranded cDNA를

합성하였다. Blunt-end를 만들기 위한 End Repair, Adapter를 붙이기 위한 A-tailing, Adapter ligation 과정을 순차적으로 거친 후 PCR(Polymerase Chain Reaction)을 이용 증폭하였다. PCR 최종 산물은 2100 BioAnalyzer를 사용하여 확인하였으며, 제작된 library는 KAPA library quantification kit을 이용하여 정량한 후 cluster generation하여 Hiseq 2500을 이용하여 염기서열 해독을 진행하였다(Figure 9).

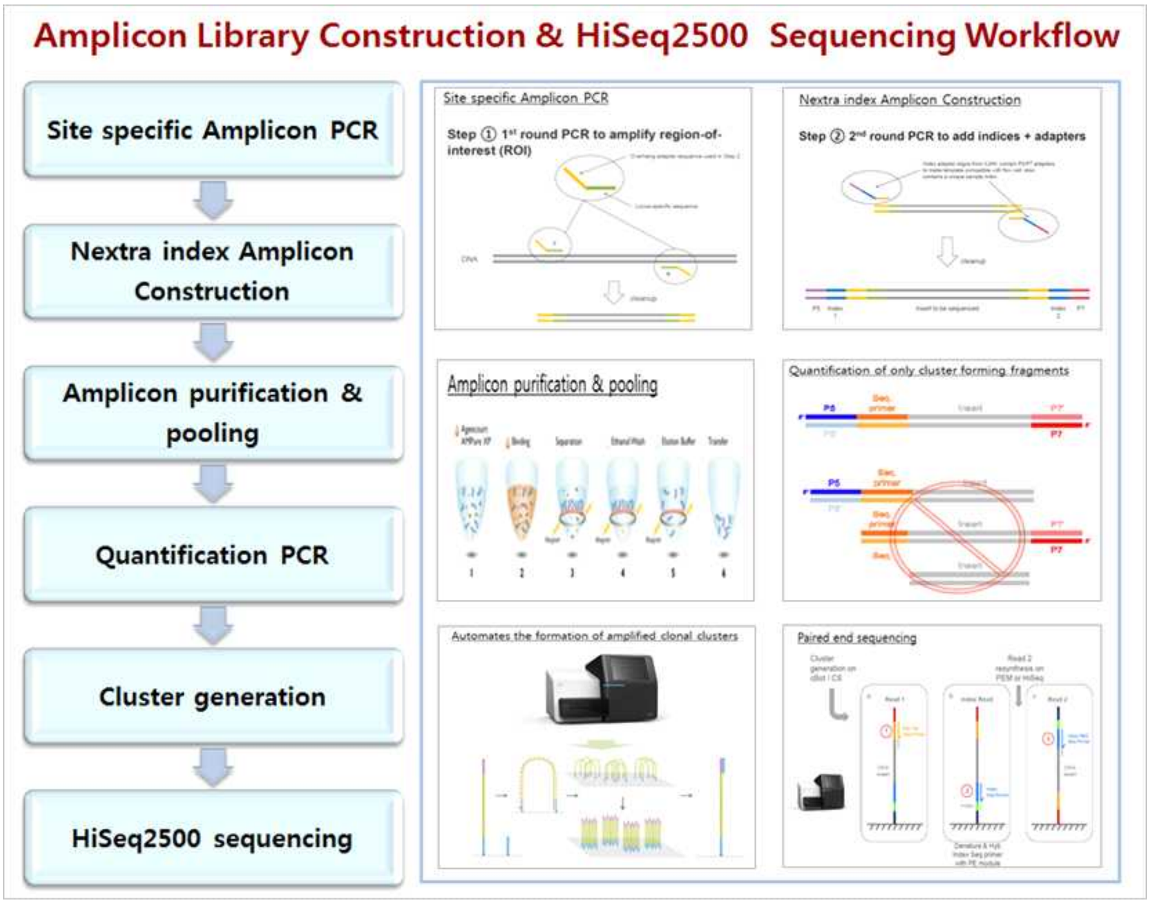


Figure 9. Workflow of amplicon library construction and Hiseq2500 sequencing.

라. 염기서열 조립 및 유전자 확보

전사체 서열을 얻기 위한 조립과정에 앞서 서열 해독과정에서 생긴 잘못된 서열을 가진 리드(Read)나 혹은 어댑터 서열을 제거하고, 낮은 품질의 서열로 인한 조립 오류를

줄이기 위한 서열 필터링 작업을 진행하였다.

서열 조립은 전체 샘플의 서열 데이터를 하나로 합친 뒤, de Bruijin graph (DBG) 알고리즘을 이용하여 RNA-seq de novo assembly를 수행하는 대표적인 RNA Assmebler인 Trinity (Grabherr et al., 2011; Hass et al., 2013) 프로그램을 이용하여 서열을 조립하였다. Trinity의 서열 조립과정은 3단계(Inchworm, Chrysalis, Butterfly)로 이루어져 있고, 첫 번째 단계인 Inchworm에서는 메모리의 효율적인 사용을 위하여 각 서열을 서로 연결하여 서브그룹으로 나누는 작업을 하였다. Inchworm 단계에서는 contig를 구성하였고, Chrysalis 단계에서는 Inchworm에서 구성된 contig를 이용하여 리드에 대한 집단화 (clustering)를 진행하였다.

조립된 전사체의 집단화는 주어진 서열을 1대 1로 비교하여 서열간의 유사도를 계산하고, CAP3 (Huang and Madan, 1999)으로 각 집단에 속한 서열을 재조립하여 집단을 대표하는 서열을 만들어 내는 TGICL (Pertea et al., 2003) 프로그램을 이용하여 수행하였다.

Unigene의 기능을 알기 위하여 단백질 발현 부분인 coding sequence (CDS)를 예측하였다. CDS 예측과정은 TransDecoder (Haas et al., 2013)를 이용하였으며, 주어진 서열에서 가능성 있는 모든 CDS 서열을 예측하였고, GeneID(Blanco et al., 2007)로 가능성을 검증하여 log-likelihood score가 0보다 큰 것을 선택하였다.

마. 유전자 기능 분석

조립된 unigene과 이를 기반으로 예측된 CDS의 기능을 파악하기 위하여 DNA 서열 및 단백질 서열의 상동성 검색을 진행하였다. DNA 서열의 상동성은 BLAST 검색을 하였고, 단백질 서열의 상동성은 InterProScan을 이용하였다.

DNA 서열에 기반한 상동성 검색은 NCBI Blast 2.2.28+와 NCBI Non-redundant(nr) database(2013/07/17)가 사용되었고, Blastx를 사용하여 unigene 서열에 대한 모든 가능한 단백질 서열을 non-redundant database를 검색하여서 그 기능을 예측하였다. 검색 과정에서 상동성의 유의성은 E-value < $1e^{-5}$로 설정하였다.

단백질 서열 기반의 상동성 검색은 Hidden Markov Model을 이용하여 단백질 기능의 단위인 도메인 수준에서 유사성을 검색하여 기능을 예측하는 InterProScan (v5)을 이용하였다. ProDom, PfamA,Panther, SMART, SuperFamily, Gene3d의 총 6개의 단백질 데이터베이스를 E-value < $1e^{-5}$를 기준으로 검색하였다.

바. 유전자 발현량 측정 및 차등발현유전자 기능정보 분석

개볼락 체색타입별 근육조직 발현유전자의 발현량 분석은 유전자의 발현량을 측정하는 도구인 RSEM (Li and Dewey, 2011) 프로그램을 이용하였으며, Bowtie를 사용하여 리드를 전사체에 정렬한 뒤 directed graph model로 전사체의 발현량을 계산하였다.

조직별 발현량을 이용하여 체색타입별로 발현량 차이를 나타내는 차등발현유전자 (Differentially Expressed Gene, DEG)를 탐색하였다. DEG 분석은 TCC (Sun et al., 2013) 프로그램의 DEGES/DESeq 방법을 활용하였다. DEGES/DESeq 방법은 Negative-binomial 분포를 이용하여 DEG 분석을 수행하는 DESeq(Anders and Huber, 2010) 방법에 기반한 분석 방법으로 초기 두 시료 사이의 normalization 과정을 반복 수행함에 따라 DEG 선정의 정확도를 높여준다(Kadota et al., 2012). 본 연구에서는 normalization 과정을 3번 반복 수행하였으며, 기준값을 q-value < 0.05 이하로 설정하여 DEG를 선정하였고, 차등발현유전자의 기능 정보는 Gene Ontology (GO) enrichment 분석으로 확인하였다.

2. 연구결과

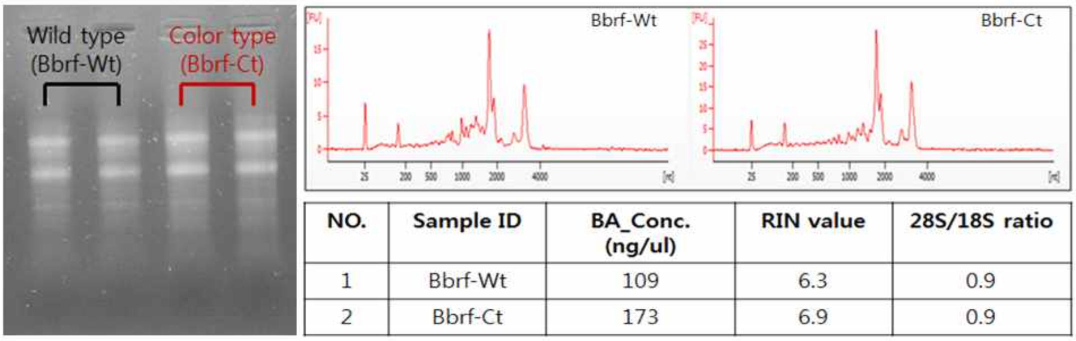
가. 볼락류의 체색관련 기능유전체 분석을 위한 시료 확보

체색이 다양한 개볼락 시료를 확보하여 반점이나 mark가 있는 color type과 반점이나 mark가 없는 wild type으로 구분하여 체색관련 기능유전체 분석 시료로 활용하였다(Figure 10).

Figure 10. The skeletal muscles of *S. pachycephalus* for RNA-seq analysis.

나. 개볼락 체색타입별 근육조직 total RNA 추출

Color type과 wild type 개볼락의 근육조직 total RNA를 추출하여 전기영동으로 확인한 후, RNA-seq 분석에 사용하기 위하여 순도를 조사하였다(Figure 11).



NO.	Sample ID	BA_Conc. (ng/ul)	RIN value	28S/18S ratio
1	Bbrf-Wt	109	6.3	0.9
2	Bbrf-Ct	173	6.9	0.9

Figure 11. The result of quality assessment of total RNA.

다. 볼락류의 체색관련 기능유전체 확보를 위한 RNA-seq analysis

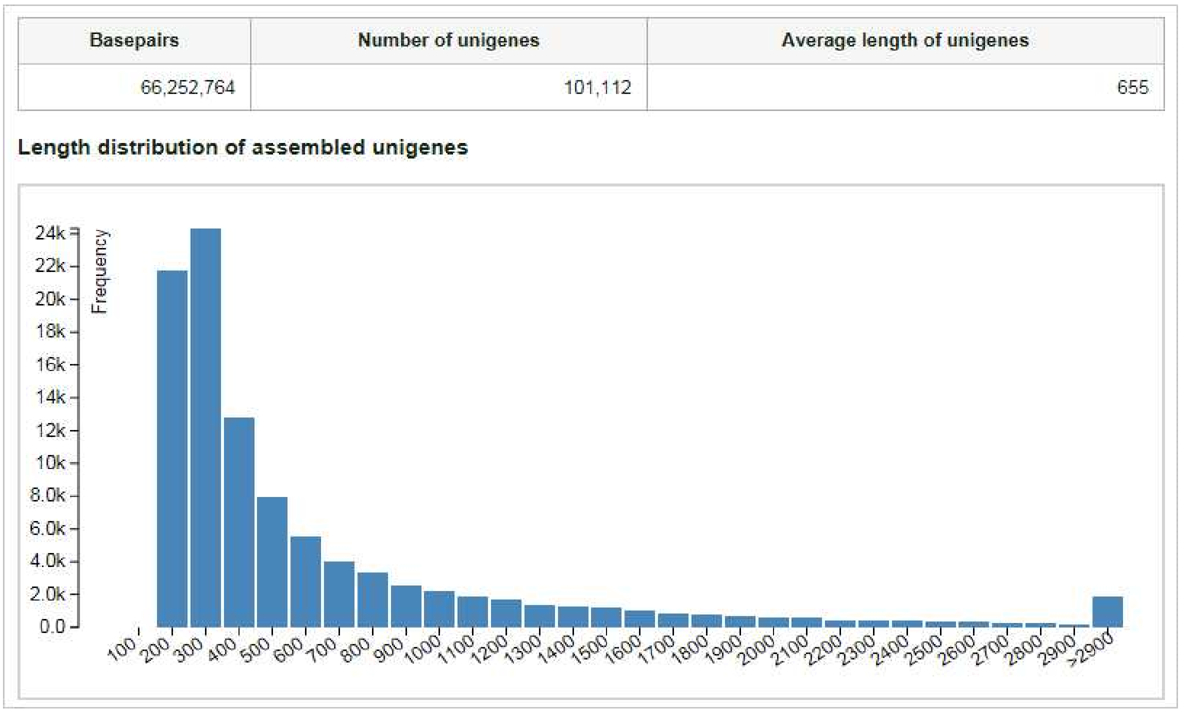
(1) 개볼락 체색타입별 근육조직 전사유전체 서열 조립 및 unigene 확보

개볼락 체색타입별 근육조직 각각의 cDNA 라이브러리를 제작한 후, 전사체 정보를 분석하기 위하여 염기서열을 해독하였다. 염기서열 정보 중 N으로 나타난 염기의 비율이 전체 서열의 10% 이상 포함되어 있거나, Q20 미만의 염기가 20% 이상인 리드, 평균 quality가 Q20 이하인 리드 및 서열의 양끝에 존재하는 Q20 미만의 염기를 제거하여 전사유전체 서열 조립을 위한 필터링을 수행하여(Table 5) unigene을 확보하였다.

Table 5. Summary of short-read data from skeletal muscle of S. pachycephalus

Sample name	Raw		Clean		Low Quality reads
	Reads	Basepair	Reads	Basepair	
Bbrf-Wt	78,199,118	7,819,911,800	73,392,934 (93.9%)	7,316,256,602 (93.6%)	4,806,184 (6.1%)
Bbrf-Ct	77,946,600	7,794,660,000	73,454,304 (94.2%)	7,323,430,566 (94.0%)	4,492,296 (5.8%)

서열 조립은 개볼락 체색 타입별 근육시료의 서열 데이터를 하나로 합친 뒤 Trinity 프로그램을 이용하여 Inchworm, Chrysalis, Butterfly 3단계를 거쳐 진행하였고, 조립된 전사체 서열의 집단화(Clustering)는 TGICL 프로그램을 이용하였으며, 클러스터링을 위한 서열 유사성 기준은 0.94로 설정하여 101,112개의 unigene(평균길이 655bp)을 확보하였다(Figure 12).

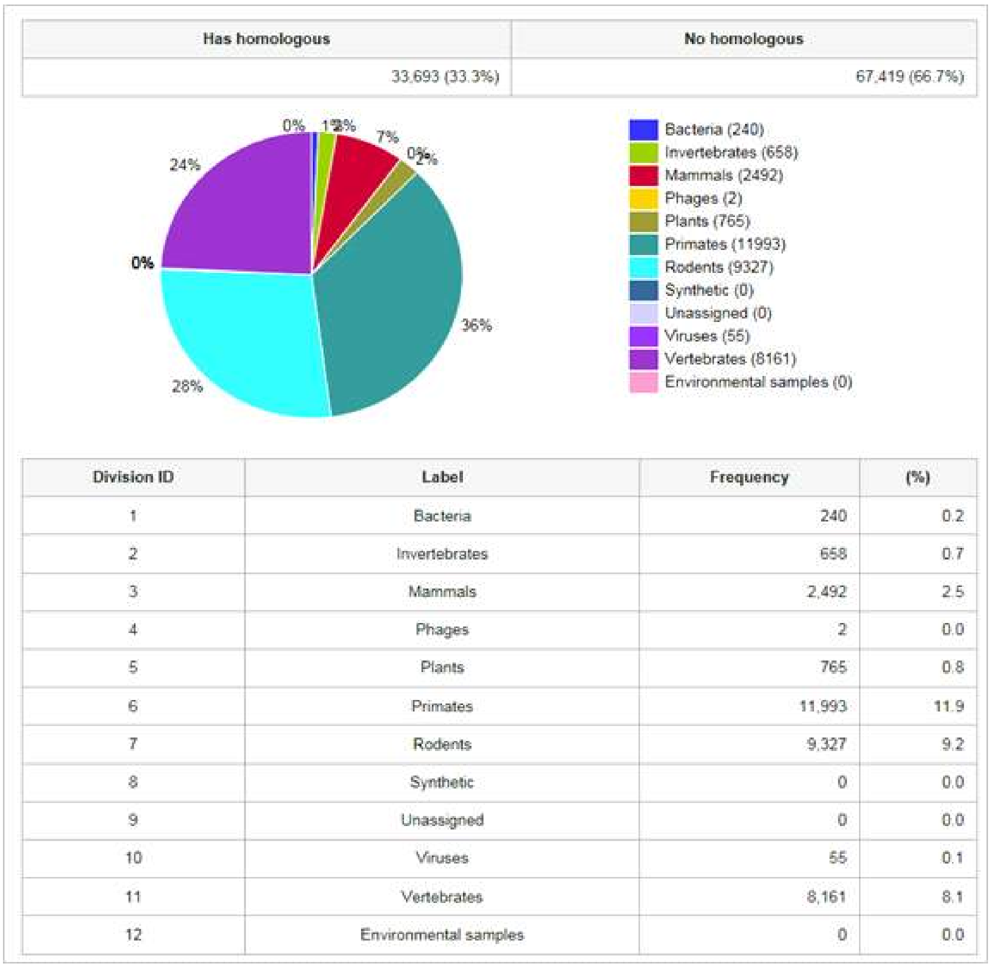


Basepairs	Number of unigenes	Average length of unigenes
66,252,764	101,112	655

Length distribution of assembled unigenes

Figure 12. Results of the blass bloched rockfish (*S. pachycephalus*) de novo assemly

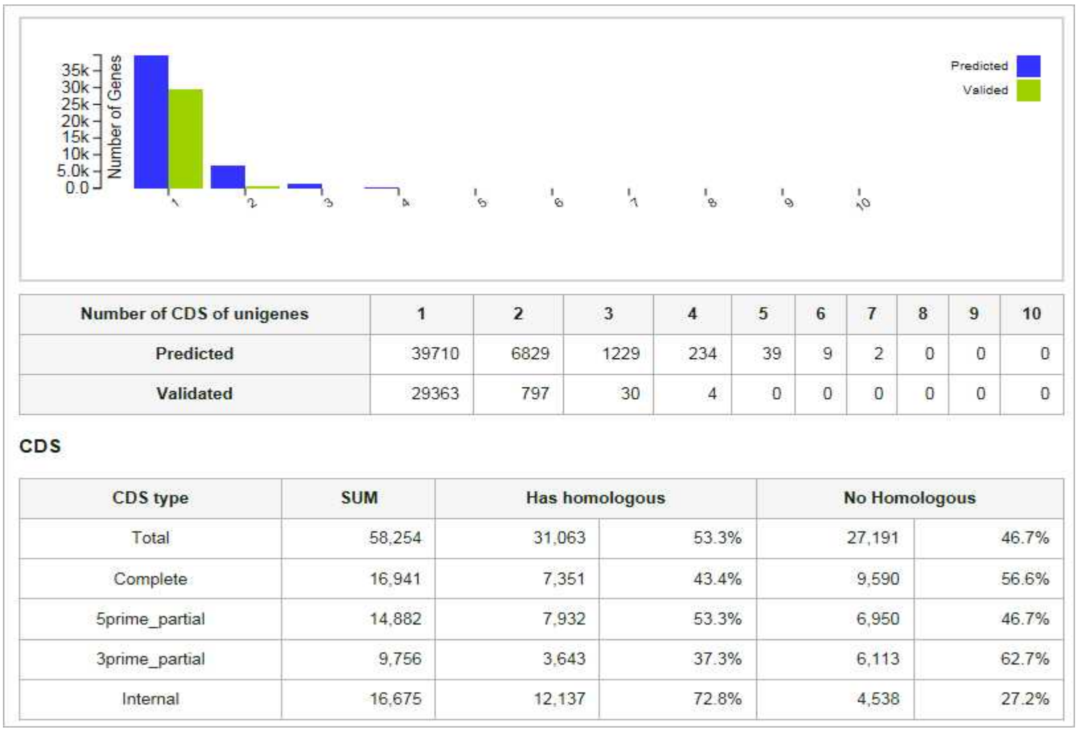
확보한 101,112개 unigene의 DNA 서열을 BLASTn 검색한 결과 67,419개(66.7%) 유전자는 상동성이 확인되지 않았고, 상동성이 확인된 33,693개(33.3%)의 유전자를 annotation 하였다(Figure 13). Unigene의 기능은 아미노산 서열을 기반으로 InterProScan v5을 이용하여 예측하였으며, 6개의 단백질 데이터베이스(ProDom, PfamA, Panther, SMART, SuperFamily, Gene3d)를 검색하였으며, 기준은 e-value < $1e^{-5}$로 정하였다(Figure 14). 개볼락 체색타입별 근육조직 전사체로부터 확보한 unigene을 검색하여 유전자의 기능이 밝혀진 35,482개와 기능이 밝혀지지 않은 65,630개 유전자의 염기서열 정보 및 아미노산 서열정보를 뱅크화 하였다(Figure 15, Table 6, Figure 16, Figure 17).



Has homologous	**No homologous**
33,693 (33.3%)	67,419 (66.7%)

Division ID	**Label**	**Frequency**	**(%)**
1	Bacteria	240	0.2
2	Invertebrates	658	0.7
3	Mammals	2,492	2.5
4	Phages	2	0.0
5	Plants	765	0.8
6	Primates	11,993	11.9
7	Rodents	9,327	9.2
8	Synthetic	0	0.0
9	Unassigned	0	0.0
10	Viruses	55	0.1
11	Vertebrates	8,161	8.1
12	Environmental samples	0	0.0

Figure 13. Frequancy of the blass bloched rockfish unigenes BLAST annotation

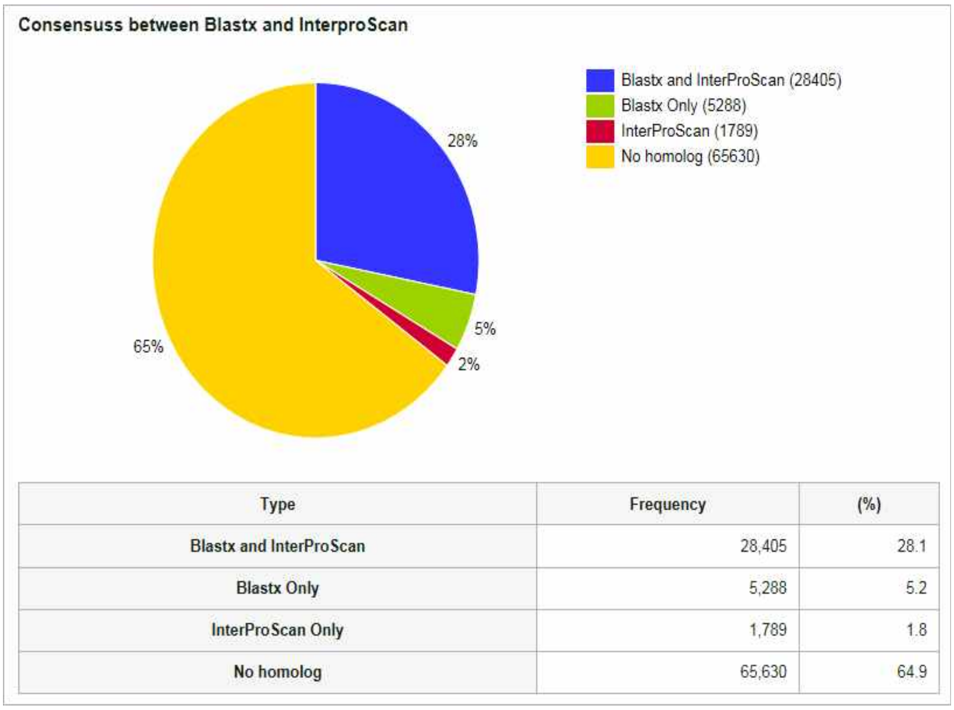


Number of CDS of unigenes	1	2	3	4	5	6	7	8	9	10
Predicted	39710	6829	1229	234	39	9	2	0	0	0
Validated	29363	797	30	4	0	0	0	0	0	0

CDS

CDS type	SUM	Has homologous		No Homologous	
Total	58,254	31,063	53.3%	27,191	46.7%
Complete	16,941	7,351	43.4%	9,590	56.6%
5prime_partial	14,882	7,932	53.3%	6,950	46.7%
3prime_partial	9,756	3,643	37.3%	6,113	62.7%
Internal	16,675	12,137	72.8%	4,538	27.2%

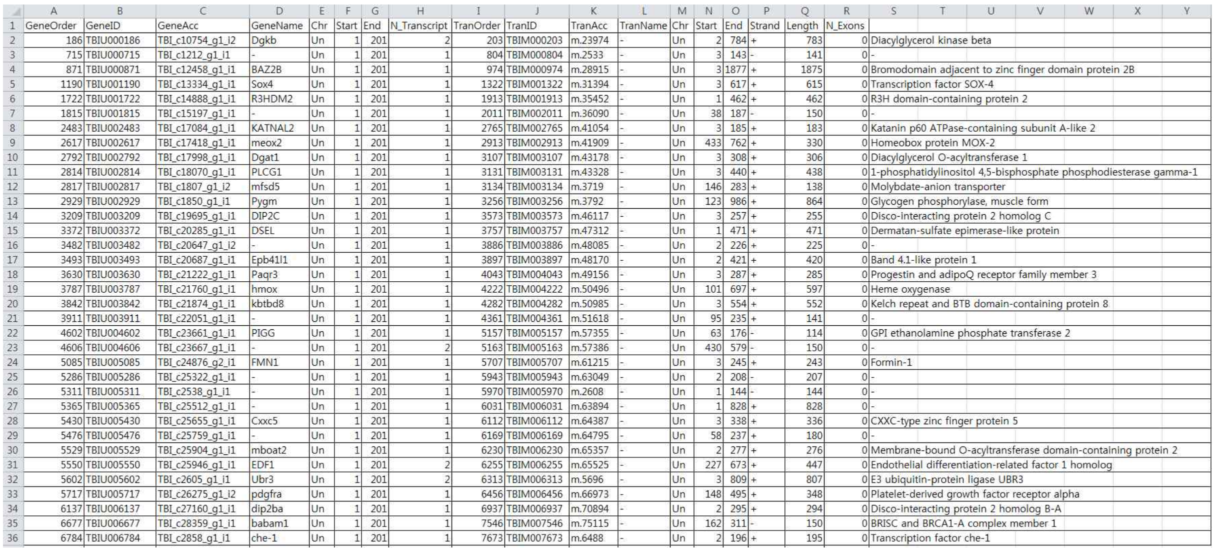
Figure 14. Predict of CDS of unigenes by InterProScan



Type	Frequency	(%)
Blastx and InterProScan	28,405	28.1
Blastx Only	5,288	5.2
InterProScan Only	1,789	1.8
No homolog	65,630	64.9

Figure 15. Annotation consensus between Blastx and InterProScan of the unigenes

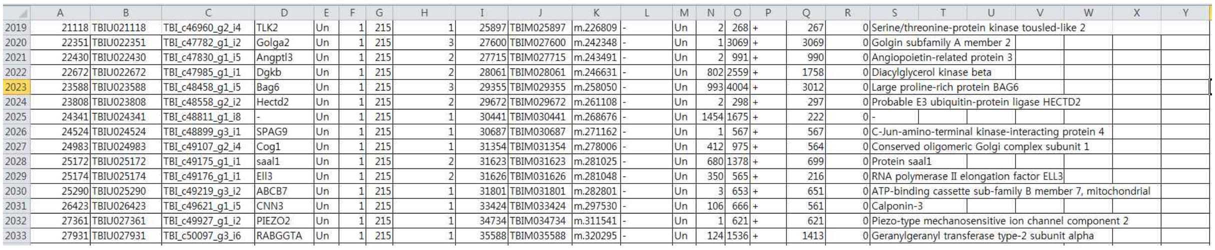
Table 6. Unigenes transcripts detected in this study



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	GeneOrder	GeneID	GeneAcc	GeneName	Chr	Start	End	N_Transcript	TranOrder	TranID	TranAcc	TranName	Chr	Start	End	Strand	Length	N_Exons	
2	186	TBIU000186	TBI_c10754_g1_i2	Dgkb	Un	1	201	2	203	TBIM000203	m.23974	-	Un	2	784	+	783	0	Diacylglycerol kinase beta
3	715	TBIU000715	TBI_c1212_g1_i1	-	Un	1	201	1	804	TBIM000804	m.2533	-	Un	3	143	-	141	0	-
4	871	TBIU000871	TBI_c12458_g1_i1	BAZ2B	Un	1	201	1	974	TBIM000974	m.28915	-	Un	3	1877	+	1875	0	Bromodomain adjacent to zinc finger domain protein 2B
5	1190	TBIU001190	TBI_c13334_g1_i1	Sox4	Un	1	201	1	1322	TBIM001322	m.31394	-	Un	3	617	+	615	0	Transcription factor SOX-4
6	1722	TBIU001722	TBI_c14888_g1_i1	R3HDM2	Un	1	201	1	1913	TBIM001913	m.35452	-	Un	1	462	+	462	0	R3H domain-containing protein 2
7	1815	TBIU001815	TBI_c15197_g1_i1	-	Un	1	201	1	2011	TBIM002011	m.36090	-	Un	38	187	-	150	0	-
8	2483	TBIU002483	TBI_c17084_g1_i1	KATNAL2	Un	1	201	1	2765	TBIM002765	m.41054	-	Un	3	185	+	183	0	Katanin p60 ATPase-containing subunit A-like 2
9	2617	TBIU002617	TBI_c17418_g1_i1	meox2	Un	1	201	1	2913	TBIM002913	m.41909	-	Un	433	762	+	330	0	Homeobox protein MOX-2
10	2792	TBIU002792	TBI_c17998_g1_i1	Dgat1	Un	1	201	1	3107	TBIM003107	m.43178	-	Un	3	308	+	306	0	Diacylglycerol O-acyltransferase 1
11	2814	TBIU002814	TBI_c18070_g1_i1	PLCG1	Un	1	201	1	3131	TBIM003131	m.43328	-	Un	3	440	+	438	0	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase gamma-1
12	2817	TBIU002817	TBI_c1807_g1_i2	mfsd5	Un	1	201	1	3134	TBIM003134	m.3719	-	Un	146	283	+	138	0	Molybdate-anion transporter
13	2929	TBIU002929	TBI_c1850_g1_i1	Pygm	Un	1	201	1	3256	TBIM003256	m.3792	-	Un	123	986	+	864	0	Glycogen phosphorylase, muscle form
14	3209	TBIU003209	TBI_c19695_g1_i1	DIP2C	Un	1	201	1	3573	TBIM003573	m.46117	-	Un	3	257	+	255	0	Disco-interacting protein 2 homolog C
15	3372	TBIU003372	TBI_c20285_g1_i1	DSEL	Un	1	201	1	3757	TBIM003757	m.47312	-	Un	1	471	+	471	0	Dermatan-sulfate epimerase-like protein
16	3482	TBIU003482	TBI_c20647_g1_i2	-	Un	1	201	1	3886	TBIM003886	m.48085	-	Un	2	226	+	225	0	-
17	3493	TBIU003493	TBI_c20687_g1_i1	Epb41l1	Un	1	201	1	3897	TBIM003897	m.48170	-	Un	2	421	+	420	0	Band 4.1-like protein 1
18	3630	TBIU003630	TBI_c21222_g1_i1	Paqr3	Un	1	201	1	4043	TBIM004043	m.49156	-	Un	3	287	+	285	0	Progestin and adipoQ receptor family member 3
19	3787	TBIU003787	TBI_c21760_g1_i1	hmox	Un	1	201	1	4222	TBIM004222	m.50496	-	Un	101	697	+	597	0	Heme oxygenase
20	3842	TBIU003842	TBI_c21874_g1_i1	kbtbd8	Un	1	201	1	4282	TBIM004282	m.50985	-	Un	3	554	+	552	0	Kelch repeat and BTB domain-containing protein 8
21	3911	TBIU003911	TBI_c22051_g1_i1	-	Un	1	201	1	4361	TBIM004361	m.51618	-	Un	95	235	+	141	0	-
22	4602	TBIU004602	TBI_c23661_g1_i1	PIGG	Un	1	201	1	5157	TBIM005157	m.57355	-	Un	63	176	-	114	0	GPI ethanolamine phosphate transferase 2
23	4606	TBIU004606	TBI_c23667_g1_i1	-	Un	1	201	2	5163	TBIM005163	m.57386	-	Un	430	579	-	150	0	-
24	5085	TBIU005085	TBI_c24876_g2_i1	FMN1	Un	1	201	1	5707	TBIM005707	m.61215	-	Un	3	245	+	243	0	Formin-1
25	5286	TBIU005286	TBI_c25322_g1_i1	-	Un	1	201	1	5943	TBIM005943	m.63049	-	Un	2	208	-	207	0	-
26	5311	TBIU005311	TBI_c2538_g1_i1	-	Un	1	201	1	5970	TBIM005970	m.2608	-	Un	1	144	-	144	0	-
27	5365	TBIU005365	TBI_c25512_g1_i1	-	Un	1	201	1	6031	TBIM006031	m.63894	-	Un	1	828	+	828	0	-
28	5430	TBIU005430	TBI_c25655_g1_i1	Cxxc5	Un	1	201	1	6112	TBIM006112	m.64387	-	Un	3	338	+	336	0	CXXC-type zinc finger protein 5
29	5476	TBIU005476	TBI_c25759_g1_i1	-	Un	1	201	1	6169	TBIM006169	m.64795	-	Un	58	237	+	180	0	-
30	5529	TBIU005529	TBI_c25904_g1_i1	mboat2	Un	1	201	1	6230	TBIM006230	m.65357	-	Un	2	277	+	276	0	Membrane-bound O-acyltransferase domain-containing protein 2
31	5550	TBIU005550	TBI_c25946_g1_i1	EDF1	Un	1	201	2	6255	TBIM006255	m.65525	-	Un	227	673	+	447	0	Endothelial differentiation-related factor 1 homolog
32	5602	TBIU005602	TBI_c2605_g1_i1	Ubr3	Un	1	201	2	6313	TBIM006313	m.5696	-	Un	3	809	+	807	0	E3 ubiquitin-protein ligase UBR3
33	5717	TBIU005717	TBI_c26275_g1_i2	pdgfra	Un	1	201	1	6456	TBIM006456	m.66973	-	Un	148	495	+	348	0	Platelet-derived growth factor receptor alpha
34	6137	TBIU006137	TBI_c27160_g1_i1	dip2ba	Un	1	201	1	6937	TBIM006937	m.70894	-	Un	2	295	+	294	0	Disco-interacting protein 2 homolog B-A
35	6677	TBIU006677	TBI_c28359_g1_i1	babam1	Un	1	201	1	7546	TBIM007546	m.75115	-	Un	162	311	-	150	0	BRISC and BRCA1-A complex member 1
36	6784	TBIU006784	TBI_c2858_g1_i1	che-1	Un	1	201	1	7673	TBIM007673	m.6488	-	Un	2	196	+	195	0	Transcription factor che-1

•

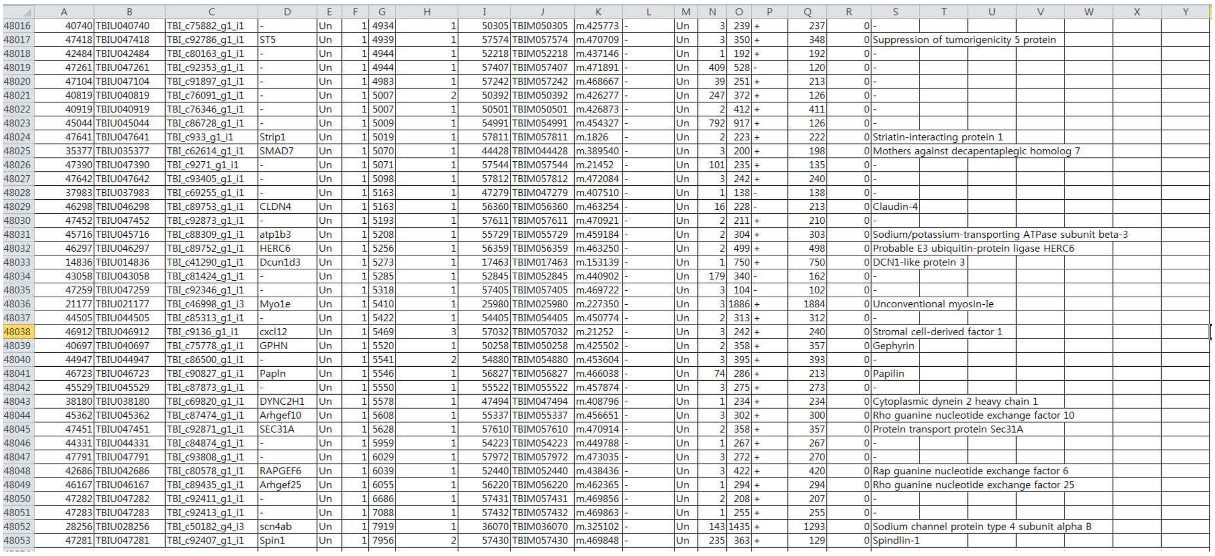
•



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
2019	21118	TBIU021118	TBI_c46960_g2_i4	TLK2	Un	1	215	1	25897	TBIM025897	m.226809	-	Un	2	268	+	267	0	Serine/threonine-protein kinase tousled-like 2
2020	22351	TBIU022351	TBI_c47782_g1_i2	Golga2	Un	1	215	3	27600	TBIM027600	m.242348	-	Un	1	3069	+	3069	0	Golgin subfamily A member 2
2021	22430	TBIU022430	TBI_c47830_g1_i5	Angptl3	Un	1	215	2	27715	TBIM027715	m.243491	-	Un	2	991	+	990	0	Angiopoietin-related protein 3
2022	22672	TBIU022672	TBI_c47985_g1_i1	Dgkb	Un	1	215	2	28061	TBIM028061	m.246631	-	Un	802	2559	+	1758	0	Diacylglycerol kinase beta
2023	23588	TBIU023588	TBI_c48458_g1_i5	Bag6	Un	1	215	3	29355	TBIM029355	m.258050	-	Un	993	4004	+	3012	0	Large proline-rich protein BAG6
2024	23808	TBIU023808	TBI_c48558_g2_i2	Hectd2	Un	1	215	2	29672	TBIM029672	m.261108	-	Un	2	298	+	297	0	Probable E3 ubiquitin-protein ligase HECTD2
2025	24341	TBIU024341	TBI_c48811_g1_i8	-	Un	1	215	1	30441	TBIM030441	m.268676	-	Un	1454	1675	+	222	0	-
2026	24524	TBIU024524	TBI_c48899_g3_i1	SPAG9	Un	1	215	1	30687	TBIM030687	m.271162	-	Un	1	567	+	567	0	C-Jun-amino-terminal kinase-interacting protein 4
2027	24983	TBIU024983	TBI_c49107_g2_i4	Cog1	Un	1	215	1	31354	TBIM031354	m.278006	-	Un	412	975	+	564	0	Conserved oligomeric Golgi complex subunit 1
2028	25172	TBIU025172	TBI_c49175_g1_i1	saal1	Un	1	215	2	31623	TBIM031623	m.281025	-	Un	680	1378	+	699	0	Protein saal1
2029	25174	TBIU025174	TBI_c49176_g1_i1	Ell3	Un	1	215	2	31626	TBIM031626	m.281048	-	Un	350	565	+	216	0	RNA polymerase II elongation factor ELL3
2030	25290	TBIU025290	TBI_c49219_g3_i2	ABCB7	Un	1	215	1	31801	TBIM031801	m.282801	-	Un	3	653	+	651	0	ATP-binding cassette sub-family B member 7, mitochondrial
2031	26423	TBIU026423	TBI_c49621_g1_i5	CNN3	Un	1	215	1	33424	TBIM033424	m.297530	-	Un	106	666	+	561	0	Calponin-3
2032	27361	TBIU027361	TBI_c49927_g1_i2	PIEZO2	Un	1	215	1	34734	TBIM034734	m.311541	-	Un	1	621	+	621	0	Piezo-type mechanosensitive ion channel component 2
2033	27931	TBIU027931	TBI_c50097_g3_i6	RABGGTA	Un	1	215	1	35588	TBIM035588	m.320295	-	Un	124	1536	+	1413	0	Geranylgeranyl transferase type-2 subunit alpha

•

•



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
48016	40740	TBIU040740	TBI_c75882_g1_i1	-	Un	1	4934	1	50305	TBIM050305	m.425773	-	Un	3	239	+	237	0	-
48017	47418	TBIU047418	TBI_c92786_g1_i1	ST5	Un	1	4939	1	57574	TBIM057574	m.470709	-	Un	3	350	+	348	0	Suppression of tumorigenicity 5 protein
48018	42484	TBIU042484	TBI_c80163_g1_i1	-	Un	1	4944	1	52218	TBIM052218	m.437146	-	Un	1	192	+	192	0	-
48019	47261	TBIU047261	TBI_c92353_g1_i1	-	Un	1	4944	1	57407	TBIM057407	m.471891	-	Un	409	528	-	120	0	-
48020	47104	TBIU047104	TBI_c91897_g1_i1	-	Un	1	4983	1	57242	TBIM057242	m.468667	-	Un	39	251	+	213	0	-
48021	40819	TBIU040819	TBI_c76091_g1_i1	-	Un	1	5007	2	50392	TBIM050392	m.426277	-	Un	247	372	+	126	0	-
48022	40919	TBIU040919	TBI_c76346_g1_i1	-	Un	1	5007	1	50501	TBIM050501	m.426873	-	Un	2	412	+	411	0	-
48023	45044	TBIU045044	TBI_c86728_g1_i1	-	Un	1	5009	1	54991	TBIM054991	m.454327	-	Un	792	917	+	126	0	-
48024	47641	TBIU047641	TBI_c933_g1_i1	Strip1	Un	1	5019	1	57811	TBIM057811	m.1826	-	Un	2	223	+	222	0	Striatin-interacting protein 1
48025	35377	TBIU035377	TBI_c62614_g1_i1	SMAD7	Un	1	5070	1	44428	TBIM044428	m.389540	-	Un	3	200	+	198	0	Mothers against decapentaplegic homolog 7
48026	47390	TBIU047390	TBI_c9271_g1_i1	-	Un	1	5071	1	57544	TBIM057544	m.21452	-	Un	101	235	+	135	0	-
48027	47642	TBIU047642	TBI_c93405_g1_i1	-	Un	1	5098	1	57812	TBIM057812	m.472084	-	Un	3	242	+	240	0	-
48028	37983	TBIU037983	TBI_c69255_g1_i1	-	Un	1	5163	1	47279	TBIM047279	m.407510	-	Un	1	138	-	138	0	-
48029	46298	TBIU046298	TBI_c89753_g1_i1	CLDN4	Un	1	5163	1	56360	TBIM056360	m.463254	-	Un	16	228	-	213	0	Claudin-4
48030	47452	TBIU047452	TBI_c92873_g1_i1	-	Un	1	5193	1	57611	TBIM057611	m.470921	-	Un	2	211	+	210	0	-
48031	45716	TBIU045716	TBI_c88309_g1_i1	atp1b3	Un	1	5208	1	55729	TBIM055729	m.459184	-	Un	2	304	+	303	0	Sodium/potassium-transporting ATPase subunit beta-3
48032	46297	TBIU046297	TBI_c89752_g1_i1	HERC6	Un	1	5256	1	56359	TBIM056359	m.463250	-	Un	2	499	+	498	0	Probable E3 ubiquitin-protein ligase HERC6
48033	14836	TBIU014836	TBI_c41290_g1_i1	Dcun1d3	Un	1	5273	1	17463	TBIM017463	m.153139	-	Un	1	750	+	750	0	DCN1-like protein 3
48034	43058	TBIU043058	TBI_c81424_g1_i1	-	Un	1	5285	1	52845	TBIM052845	m.440902	-	Un	179	340	-	162	0	-
48035	47259	TBIU047259	TBI_c92346_g1_i1	-	Un	1	5318	1	57405	TBIM057405	m.469722	-	Un	3	104	-	102	0	-
48036	21177	TBIU021177	TBI_c46998_g1_i3	Myo1e	Un	1	5410	1	25980	TBIM025980	m.227350	-	Un	3	1886	+	1884	0	Unconventional myosin-Ie
48037	44505	TBIU044505	TBI_c85313_g1_i1	-	Un	1	5422	1	54405	TBIM054405	m.450774	-	Un	2	313	+	312	0	-
48038	46912	TBIU046912	TBI_c9136_g1_i1	cxcl12	Un	1	5469	3	57032	TBIM057032	m.21252	-	Un	3	242	+	240	0	Stromal cell-derived factor 1
48039	40697	TBIU040697	TBI_c75778_g1_i1	GPHN	Un	1	5520	1	50258	TBIM050258	m.425502	-	Un	2	358	+	357	0	Gephyrin
48040	44947	TBIU044947	TBI_c86500_g1_i1	-	Un	1	5541	2	54880	TBIM054880	m.453604	-	Un	3	395	+	393	0	-
48041	46723	TBIU046723	TBI_c90827_g1_i1	Papln	Un	1	5546	1	56827	TBIM056827	m.466038	-	Un	74	286	+	213	0	Papilin
48042	45529	TBIU045529	TBI_c87873_g1_i1	-	Un	1	5550	1	55522	TBIM055522	m.457874	-	Un	3	275	+	273	0	-
48043	38180	TBIU038180	TBI_c69820_g1_i1	DYNC2H1	Un	1	5578	1	47494	TBIM047494	m.408796	-	Un	1	234	+	234	0	Cytoplasmic dynein 2 heavy chain 1
48044	45362	TBIU045362	TBI_c87474_g1_i1	Arhgef10	Un	1	5608	1	55337	TBIM055337	m.456651	-	Un	3	302	+	300	0	Rho guanine nucleotide exchange factor 10
48045	47451	TBIU047451	TBI_c92871_g1_i1	SEC31A	Un	1	5628	1	57610	TBIM057610	m.470914	-	Un	2	358	+	357	0	Protein transport protein Sec31A
48046	44331	TBIU044331	TBI_c84874_g1_i1	-	Un	1	5959	1	54223	TBIM054223	m.449788	-	Un	1	267	+	267	0	-
48047	47791	TBIU047791	TBI_c93808_g1_i1	-	Un	1	6029	1	57972	TBIM057972	m.473035	-	Un	3	272	+	270	0	-
48048	42686	TBIU042686	TBI_c80578_g1_i1	RAPGEF6	Un	1	6039	1	52440	TBIM052440	m.438436	-	Un	3	422	+	420	0	Rap guanine nucleotide exchange factor 6
48049	46167	TBIU046167	TBI_c89435_g1_i1	Arhgef25	Un	1	6055	1	56220	TBIM056220	m.462365	-	Un	1	294	+	294	0	Rho guanine nucleotide exchange factor 25
48050	47282	TBIU047282	TBI_c92411_g1_i1	-	Un	1	6686	1	57431	TBIM057431	m.469856	-	Un	2	208	+	207	0	-
48051	47283	TBIU047283	TBI_c92413_g1_i1	-	Un	1	7088	1	57432	TBIM057432	m.469863	-	Un	1	255	+	255	0	-
48052	28256	TBIU028256	TBI_c50182_g4_i3	scn4ab	Un	1	7919	1	36070	TBIM036070	m.325102	-	Un	143	1435	+	1293	0	Sodium channel protein type 4 subunit alpha B
48053	47281	TBIU047281	TBI_c92407_g1_i1	Spin1	Un	1	7956	2	57430	TBIM057430	m.469848	-	Un	235	363	+	129	0	Spindlin-1

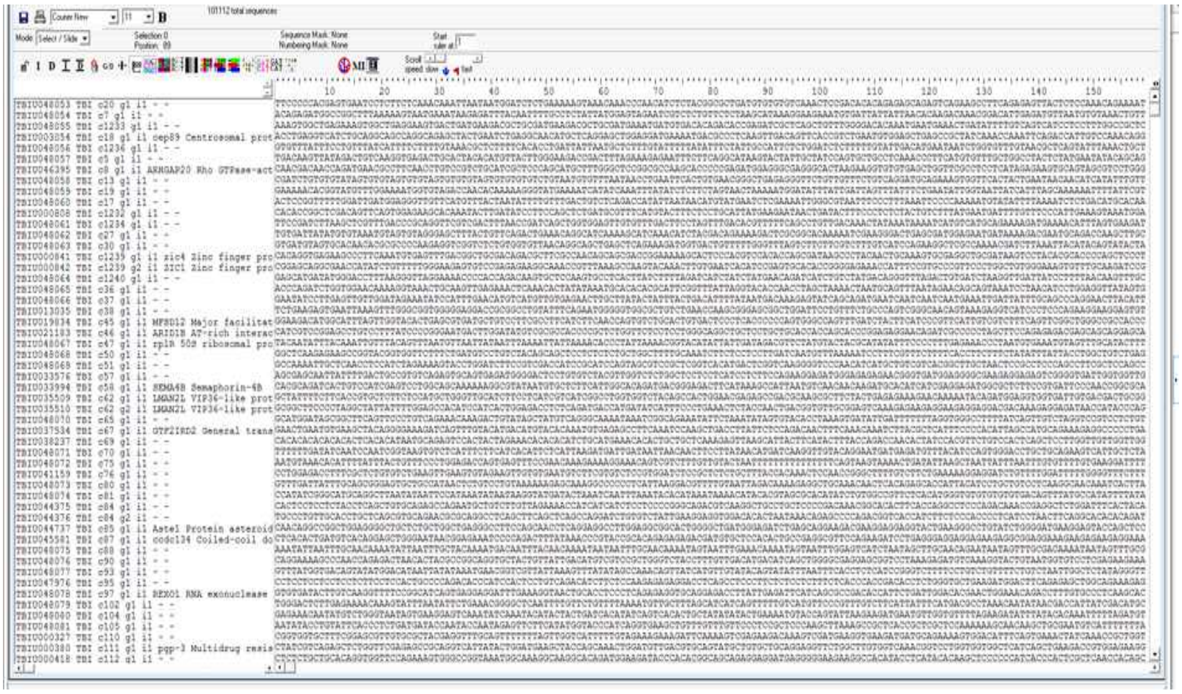


Figure 16. DNA sequences of unigene transcripts

Figure 17. Amino acid sequences of unigene transcripts

(2) 개볼락 체색타입별 근육조직 전사유전체의 발현량 측정

RSEM 프로그램을 이용하여 개볼락 체색 타입별 근육조직 발현량을 분석하였으며 Bowtie를 이용하여 리드를 전사체에 정렬한 뒤 directed graph model로 전사체의 발현량을 측정하였다. 개볼락 근육조직에서 발현되는 유전자 개수가 체색타입별로 차이를 나타내었다. 노란색 반점을 가진 color type (Bbrf-Ct) 근육조직에서는 53,499개의 유전자가 발현되었으며, wild type (Bbrf-Wt) 근육조직에서는 55,508개의 유전자가 발현되었다(Figure 18).

UNIGENES

Sum	Known	Novel
101,112	35,482	65,630

Expression

Genes (> fpkm 1.0)

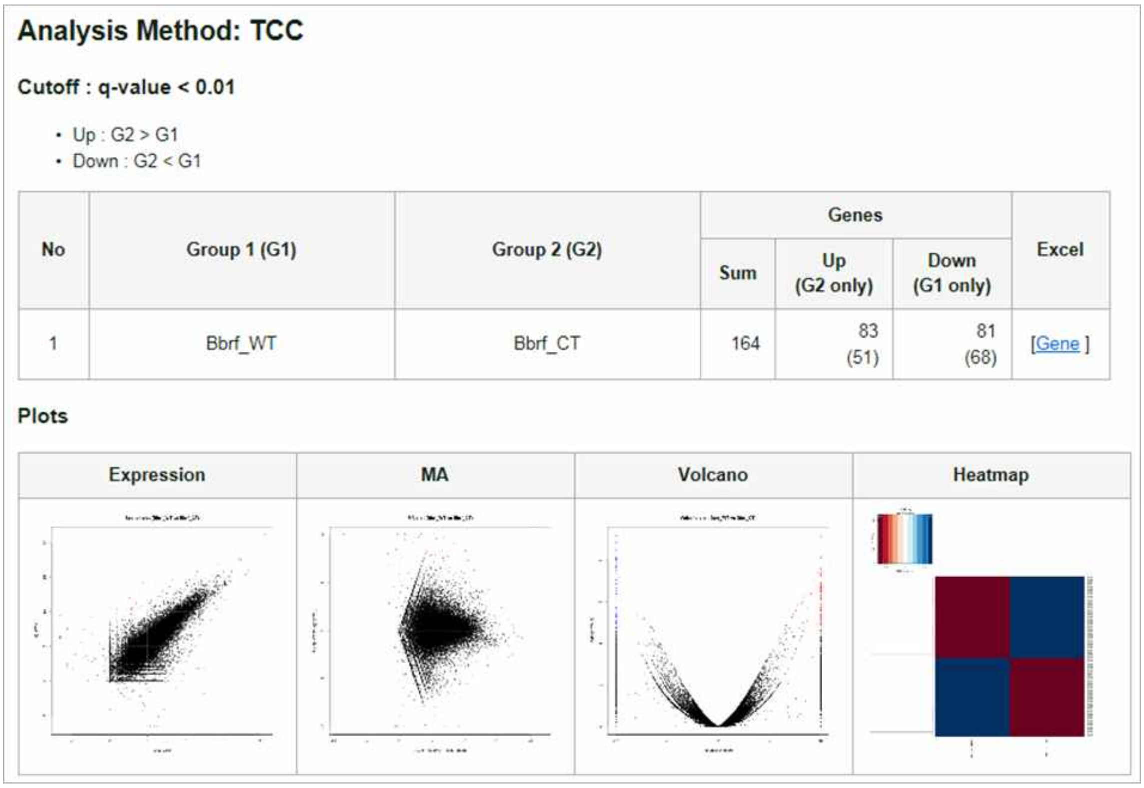
Name	Gene				Gene (> fpkm 1.0)			
	Expressed	Known	Novel	Unexpressed	Expressed	Known	Novel	Unexpressed
Bbrf_WT	85,831	32,094	53,737	15,281	55,508	22,504	33,004	4,611
Bbrf_CT	84,198	31,226	52,972	16,914	53,499	21,695	31,804	6,620

Figure 18. Expressiom levels of unigene transcripts

(3) 개볼락 근육조직의 체색타입별 차등발현유전자 탐색

개볼락의 체색 타입에 따른 DEG를 탐색하고 분석한 결과, 164개의 유전자가 발현량 차이를 나타내었으며(Figure 19), 차등발현유전자중 wild type에서만 발현되는 68개의 유전자와 color type에서만 발현되는 51개의 유전자를 확보하였다(Table 7).

개볼락 color type 근육조직에서 차등발현하는 유전자 개수는 83개 중 51개 유전자는 wild type 근육조직에서는 발현되지않고 color type 근육조직에서만 발현되었으며, 35개는 기능이 밝혀진 유전자이었다(Table 8). 반면, 개볼락 wild type 근육조직에서 차등발현하는 유전자 개수는 81개이었고, 68개 유전자는 color type 근육조직에서는 발현되지않고 wild type 근육조직에서만 발현되었으며, 44개가 기능이 밝혀진 유전자이었다(Table 9).



Analysis Method: TCC

Cutoff : q-value < 0.01

- Up : G2 > G1
- Down : G2 < G1

No	Group 1 (G1)	Group 2 (G2)	Genes			Excel
			Sum	Up (G2 only)	Down (G1 only)	
1	Bbrf_WT	Bbrf_CT	164	83 (51)	81 (68)	[Gene]

Plots

Figure 19. Differentially expressed genes (DEG) analysis of unigene transcripts

Table 7. List of color pattern type-specific differentially expressed genes

GeneId	GeneName	q-value	DEG:UP_DOWN	EXP:Bbrf_WT:FPKM	EXP:Bbrf_CT:FPKM
TBIU062084	-	0.000288592	DOWN	270.35	0
TBIU064175	-	7.42E-05	UP	0	72.53
TBIU064176	-	1.60E-05	DOWN	341.79	0
TBIU065291	-	0.00541945	DOWN	29.63	0
TBIU065468	-	0.000878531	DOWN	82.08	0.1
TBIU067415	rnaset2	0.00327647	DOWN	9.03	0
TBIU067654	-	0.00332399	UP	0	14.36
TBIU068429	-	0.000901415	UP	0.11	55.42
TBIU069081	-	0.00317218	DOWN	23.33	0
TBIU069579	-	0.00541945	UP	0	9.14
TBIU069971	-	0.000136337	UP	0.16	246.51
TBIU069985	-	0.00101459	UP	0	21.7
TBIU069986	-	0.00174934	DOWN	19.56	0
TBIU070060	-	0.00648428	DOWN	20.85	0
TBIU070198	-	0.00877295	DOWN	10.6	0
TBIU070559	-	0.000205273	DOWN	72.34	0
TBIU070668	-	0.00374817	DOWN	20.79	0.02
TBIU070994	-	0.000562856	UP	1.91	978.01
TBIU071086	-	0.00397836	DOWN	9.72	0
TBIU071087	-	0.00399712	UP	0	9.39
TBIU071155	-	0.00225047	DOWN	37.96	0
TBIU071992	-	0.00374817	UP	0	35.14
TBIU073514	-	0.0092286	UP	0	15.61
TBIU072198	-	0.00201045	DOWN	27.79	0
TBIU072278	-	0.00790906	UP	7.13	688.67
TBIU073257	-	0.00631178	UP	0	11.92
TBIU073667	Etnk1	0.000350436	DOWN	11.09	0
TBIU073754	-	3.84E-05	UP	0	158.4
TBIU073969	-	0.000581513	UP	0	15.98
TBIU074890	-	0.00572695	UP	0	19.16
TBIU075697	-	0.00545732	UP	0.14	56.99
TBIU075704	-	0.00397836	UP	0	18.8
TBIU019675	letm2	0.00182905	UP	0.04	14.88
TBIU024777	CUL1	5.88E-05	UP	0	58.69
TBIU011833	-	0.000443256	UP	0	48.18
TBIU027578	NMRK2	1.60E-05	DOWN	453.55	0.02
TBIU004254	Sds	0.00101915	UP	0.22	65.74
TBIU027579	NMRK2	1.60E-05	DOWN	520.22	0
TBIU022000	-	0.00313787	DOWN	122.01	0.46

GeneId	GeneName	q-value	DEG:UP_DOWN	EXP:Bbrf_WT:FPKM	EXP:Bbrf_CT:FPKM
TBIU005394	bcas2	0.00180366	DOWN	21.63	0
TBIU020370	-	0.00273889	UP	0	17.38
TBIU004387	TOMM7	5.94E-05	DOWN	147.43	0
TBIU025303	-	0.000903742	UP	0	13.79
TBIU027572	NMRK2	1.60E-05	DOWN	184.81	0
TBIU018571	LACTB	0.00349544	UP	0	11.22
TBIU026119	MST4	0.000831836	DOWN	18.56	0
TBIU027793	DCAF11	0.000640213	DOWN	14.4	0
TBIU011844	mrp63	0.000831836	UP	0	50.88
TBIU028221	murc	0.000382233	DOWN	17.89	0
TBIU028294	-	0.000500669	DOWN	41.26	0
TBIU027575	NMRK2	3.84E-05	DOWN	129.75	0
TBIU028890	S100a9	0.00371605	DOWN	24.49	0
TBIU010317	vma21	8.86E-05	DOWN	67.08	0
TBIU025670	MRPL18	0.000382233	DOWN	43.5	0
TBIU014060	MYH7	2.43E-05	UP	0	441.29
TBIU027577	NMRK2	0.00660113	DOWN	91.4	0.78
TBIU019691	-	0.000146252	UP	0	24.81
TBIU021247	ANXA6	0.000665209	UP	0	15.31
TBIU023744	Rnf213	0.000355721	DOWN	40.56	0
TBIU015845	SCEL	0.000471375	DOWN	19.74	0
TBIU022338	NBR1	0.00973217	UP	0.04	14.22
TBIU010365	-	0.000351338	UP	0.34	245.76
TBIU014932	MRPS14	0.00310118	UP	0	30.76
TBIU028775	samhd1	0.00131392	DOWN	18.96	0
TBIU010202	Wdr33	0.000382233	UP	0	40.17
TBIU013082	fos	0.00431954	DOWN	38.75	0.24
TBIU020844	YPEL5	0.00522832	DOWN	9.06	0
TBIU026224	ascc3	0.000562856	UP	0	9.65
TBIU023185	tiprl	0.00476572	UP	0	9.89
TBIU001129	C1QTNF3	0.00215564	UP	0.57	105.08
TBIU023347	GIMAP8	0.00973217	DOWN	100.51	1.05
TBIU028776	samhd1	0.000903742	UP	0	20.02
TBIU021383	Ppan	0.00141128	DOWN	27.74	0
TBIU008482	PFKM	0.000760019	UP	0.61	193.8
TBIU017755	isca1	9.91E-05	UP	0	67.85
TBIU008448	HRSP12	0.000382233	DOWN	54.72	0
TBIU028292	-	0.00101915	DOWN	70.09	0
TBIU024593	GIMAP7	0.00840863	DOWN	6.62	0
TBIU010127	SNTX subunit beta	0.000471375	DOWN	1331.32	0.43

GeneId	GeneName	q-value	DEG:UP_DOWN	EXP:Bbrf_WT:FPKM	EXP:Bbrf_CT:FPKM
TBIU016181	ZRSR2	0.00259161	DOWN	8.55	0
TBIU019960	Lgals9	0.000180366	UP	0	38.83
TBIU017754	isca1	9.91E-05	DOWN	68.02	0
TBIU009302	-	7.36E-05	UP	0	126.86
TBIU004388	TOMM7	6.02E-05	UP	0	163.5
TBIU023658	cmas	0.00028491	UP	0	23.93
TBIU012460	-	0.000894994	DOWN	20.32	0
TBIU020950	PCBP2	0.000279397	DOWN	23.78	0
TBIU030108	klhl20	0.00166919	UP	0.22	42.75
TBIU015715	IGFN1	0.00476572	UP	0	6.88
TBIU028527	Emilin1	0.000208004	UP	0	14.56
TBIU025158	-	0.000298688	UP	0	31.45
TBIU025851	TNS	0.000125783	DOWN	65.93	0.06
TBIU017534	ACADM	7.42E-05	DOWN	46.83	0
TBIU026028	slc38a2	0.000124019	DOWN	18.78	0
TBIU016140	unc45b	0.00572695	UP	0	5.72
TBIU021898	RBM12B	0.000740309	UP	0	8.94
TBIU016179	ZRSR2	0.000629851	UP	0	13.07
TBIU021897	RBM12B	0.000424178	DOWN	11.05	0
TBIU023497	SGCA	8.86E-05	DOWN	220.52	0.13
TBIU017010	LEG_ELEEL	0.00729205	UP	3.36	2635.33
TBIU018148	FEL	0.000640213	UP	0.51	160.02
TBIU005395	bcas2	0.00264582	UP	0	19.06
TBIU013080	fos	0.000333069	DOWN	25.34	0
TBIU025160	-	0.000148257	DOWN	39.9	0
TBIU008231	ARL3	0.00190558	UP	0	24.19
TBIU005548	-	0.0003629	DOWN	148.14	0
TBIU025389	-	0.000529899	DOWN	22.94	0
TBIU010126	SNTX subunit beta	0.000124019	UP	0.15	125.42
TBIU026370	-	0.00235698	DOWN	71.57	0
TBIU028859	Dhx8	0.00314893	DOWN	10.11	0
TBIU017395	-	0.00186158	DOWN	35.51	0
TBIU030393	MYOT	5.94E-05	UP	0	20.64
TBIU026051	Nfkb2	0.00374817	UP	0	8.85
TBIU023584	Bag6	0.00724879	DOWN	4.23	0
TBIU027220	CTL	8.86E-05	UP	0	71.44
TBIU015844	SCEL	0.00259161	UP	0.2	37.93
TBIU019692	-	0.00028491	DOWN	18.9	0
TBIU008479	PFKM	0.00374817	UP	0.51	113.64
TBIU022019	tmem66	0.00541772	DOWN	10.57	0

GeneId	GeneName	q-value	DEG:UP_DOWN	EXP:Bbrf_WT:FPKM	EXP:Bbrf_CT:FPKM
TBIU017325	Ube2l3	0.000529899	DOWN	26.21	0
TBIU023496	SGCA	0.000216366	UP	0.23	145.08
TBIU026697	fam73a	0.000476091	DOWN	11.1	0
TBIU025540	BAG4	0.000374894	DOWN	18.58	0
TBIU019961	LGALS9	0.000773707	DOWN	22.97	0
TBIU025183	CAC1S_CYPCA	0.000208004	UP	0	35.2
TBIU029397	Aloxe3	0.00313787	DOWN	15.58	0
TBIU028649	-	0.000443256	DOWN	14.43	0
TBIU031738	GPI	0.000903742	UP	0.69	598.34
TBIU010316	vma21	9.21E-05	UP	0	63.64
TBIU026225	ascc3	0.00141128	DOWN	7.24	0
TBIU026410	-	0.00180366	UP	0	8.23
TBIU029225	brd4	0.00407128	DOWN	5.67	0
TBIU038365	MYSS_CYPCA	0.00401887	UP	2.08	299.64
TBIU025184	CAC1S_CYPCA	0.000873524	DOWN	21.09	0
TBIU027136	CUZD1	0.00122484	UP	0.71	185.84
TBIU028904	ZNF106	5.88E-05	UP	0	18.34
TBIU029809	NEB	0.000119949	UP	0	18.82
TBIU028488	Acot11	0.00186827	UP	0	14.25
TBIU026001	-	0.0022221	DOWN	29.66	0
TBIU020166	-	0.00484085	DOWN	84.23	0
TBIU028706	-	0.00174934	DOWN	19.86	0
TBIU025069	ASL	0.00186825	UP	0	15.92
TBIU023901	V-MIL	0.00322677	UP	0	19.34
TBIU015572	egr1	0.00943071	DOWN	46.82	0.47
TBIU022079	Gm129	9.91E-05	DOWN	23.03	0
TBIU029492	Thrap3	0.0031625	UP	0.1	15.4
TBIU025757	GOT1	0.00837356	UP	0.59	53.97
TBIU020411	FAM57B	0.00440512	DOWN	12.89	0
TBIU026698	fam73a	5.94E-05	UP	0	28.19
TBIU008401	lyrm7	0.00886531	UP	0	21.49
TBIU020043	krt13	0.00652609	UP	2.93	867.66
TBIU020410	FAM57B	0.00724879	UP	0	10.64
TBIU024378	ACVR1	0.00474586	DOWN	16.87	0
TBIU025825	POL	0.00101915	DOWN	14.92	0.03
TBIU029849	TNNI3	0.00141128	UP	0.51	106.61
TBIU027072	SLC4A7	0.000471375	DOWN	20.99	0
TBIU033342	Apcs	0.00392642	UP	2.17	306.99
TBIU029784	b2m	0.000102214	UP	0	57.3
TBIU026562	FLNC	5.88E-05	DOWN	13.49	0
TBIU029975	-	3.84E-05	UP	0	46.64
TBIU024774	CUL1	0.000180366	DOWN	97.11	0.11
TBIU029786	b2m	5.88E-05	DOWN	66.42	0
TBIU045153	-	0.00584705	UP	3.61	398.03
TBIU042518	-	0.000562856	UP	0.33	173.1

Table 8. Known genes of color-type skeletal muscle specific expressed genes

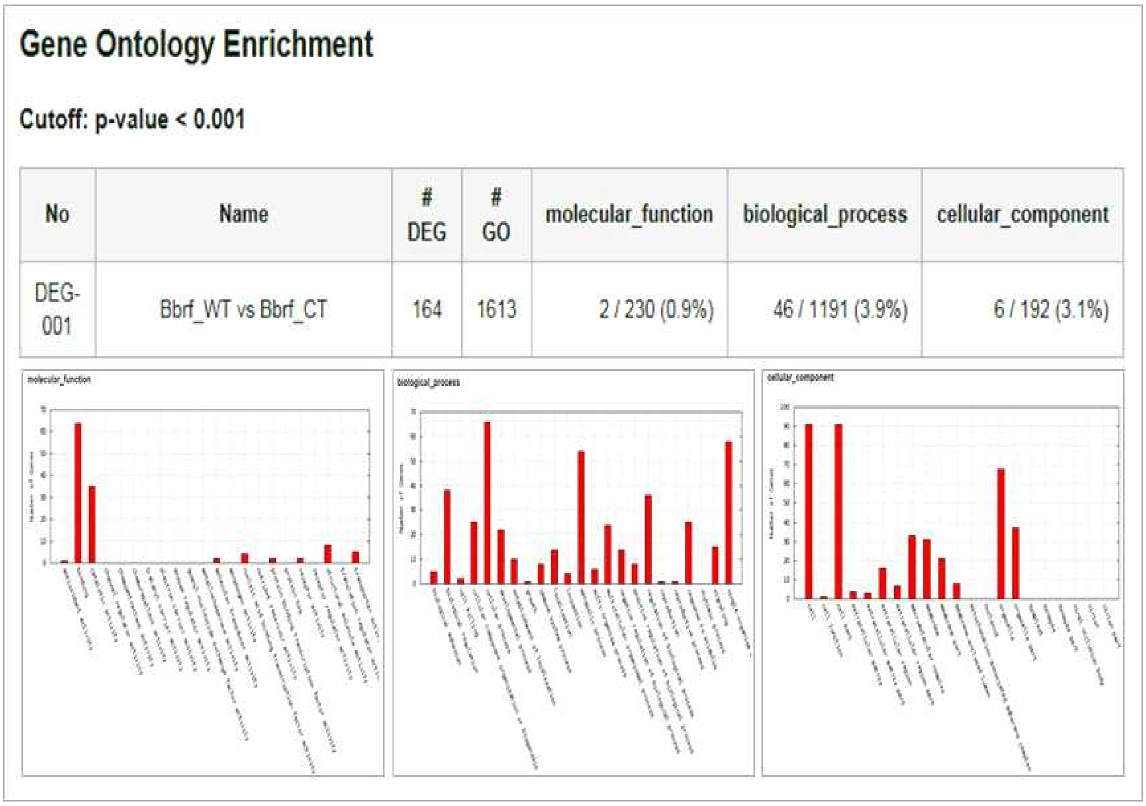
No.	GeneId	GeneName	Description	EXP:Bbrf_WT:FPKM	EXP:Bbrf_CT:FPKM
1	TBIU024777	CUL1	Cullin-1	0	58.69
2	TBIU018571	LACTB	Serine beta-lactamase-like protein LACTB, mitochondrial	0	11.22
3	TBIU011844	mrp63	Ribosomal protein 63, mitochondrial	0	50.88
4	TBIU014060	MYH7	Myosin-7	0	441.29
5	TBIU021247	ANXA6	Annexin A6	0	15.31
6	TBIU014932	MRPS14	28S ribosomal protein S14, mitochondrial	0	30.76
7	TBIU010202	Wdr33	pre-mRNA 3' end processing protein WDR33	0	40.17
8	TBIU026224	ascc3	Activating signal cointegrator 1 complex subunit 3	0	9.65
9	TBIU023185	tiprl	TIP41-like protein	0	9.89
10	TBIU028776	samhd1	Deoxynucleoside triphosphate triphosphohydrolase SAMHD1	0	20.02
11	TBIU017755	isca1	Iron-sulfur cluster assembly 1 homolog, mitochondrial	0	67.85
12	TBIU019960	Lgals9	Galectin-9	0	38.83
13	TBIU004388	TOMM7	Mitochondrial import receptor subunit TOM7 homolog	0	163.5
14	TBIU023658	cmas	N-acylneuraminate cytidylyltransferase	0	23.93
15	TBIU015715	IGFN1	Immunoglobulin-like and fibronectin type Ⅲ domain-containing protein 1	0	6.88
16	TBIU028527	Emilin1	EMILIN-1	0	14.56
17	TBIU016140	unc45b	Protein unc-45 homolog B	0	5.72
18	TBIU021898	RBM12B	RNA-binding protein 12B	0	8.94
19	TBIU016179	ZRSR2	U2 small nuclear ribonucleoprotein auxiliary factor 35 kDa subunit-related protein 2	0	13.07
20	TBIU005395	bcas2	Pre-mRNA-splicing factor SPF27	0	19.06
21	TBIU008231	ARL3	ADP-ribosylation factor-like protein 3	0	24.19
22	TBIU030393	MYOT	Myotilin	0	20.64
23	TBIU026051	Nfkb2	Nuclear factor NF-kappa-B p100 subunit	0	8.85
24	TBIU027220	CTL	Galactose-specific lectin nattectin	0	71.44
25	TBIU025183	CAC1S_CYPCA	Dihydropyridine-sensitive L-type skeletal muscle calcium channel subunit alpha-1	0	35.2
26	TBIU010316	vma21	Vacuolar ATPase assembly integral membrane protein vma21	0	63.64
27	TBIU028904	ZNF106	Zinc finger protein 106	0	18.34
28	TBIU029809	NEB	Nebulin	0	18.82
29	TBIU028488	Acot11	Acyl-coenzyme A thioesterase 11	0	14.25
30	TBIU025069	ASL	Argininosuccinate lyase	0	15.92
31	TBIU023901	V-MIL	Serine/threonine-protein kinase-transforming protein mil	0	19.34
32	TBIU026698	fam73a	Protein FAM73A	0	28.19
33	TBIU008401	lyrm7	Complex Ⅲ assembly factor LYRM7	0	21.49
34	TBIU020410	FAM57B	Protein FAM57B	0	10.64
35	TBIU029784	b2m	Beta-2-microglobulin	0	57.3

Table 9. Known genes of wild color-type skeletal muscle specific expressed genes

No.	GeneId	GeneName	Description	EXP:Bbrf_WT:FPKM	EXP:Bbrf_CT:FPKM
1	TBIU067415	rnaset2	Ribonuclease T2	9.03	0
2	TBIU073667	Etnk1	Ethanolamine kinase 1	11.09	0
3	TBIU027579	NMRK2	Nicotinamide riboside kinase 2	520.22	0
4	TBIU005394	bcas2	Pre-mRNA-splicing factor SPF27	21.63	0
5	TBIU004387	TOMM7	Mitochondrial import receptor subunit TOM7 homolog	147.43	0
6	TBIU027572	NMRK2	Nicotinamide riboside kinase 2	184.81	0
7	TBIU026119	MST4	Serine/threonine-protein kinase MST4	18.56	0
8	TBIU027793	DCAF11	DDB1- and CUL4-associated factor 11	14.4	0
9	TBIU028221	murc	Muscle-related coiled-coil protein	17.89	0
10	TBIU027575	NMRK2	Nicotinamide riboside kinase 2	129.75	0
11	TBIU028890	S100a9	Protein S100-A9	24.49	0
12	TBIU010317	vma21	Vacuolar ATPase assembly integral membrane protein vma21	67.08	0
13	TBIU025670	MRPL18	39S ribosomal protein L18, mitochondrial	43.5	0
14	TBIU023744	Rnf213	E3 ubiquitin-protein ligase RNF213	40.56	0
15	TBIU015845	SCEL	Sciellin	19.74	0
16	TBIU028775	samhd1	Deoxynucleoside triphosphate triphosphohydrolase SAMHD1	18.96	0
17	TBIU020844	YPEL5	Protein yippee-like 5	9.06	0
18	TBIU021383	Ppan	Suppressor of SWI4 1 homolog	27.74	0
19	TBIU008448	HRSP12	Ribonuclease UK114	54.72	0
20	TBIU024593	GIMAP7	GTPase IMAP family member 7	6.62	0
21	TBIU016181	ZRSR2	U2 small nuclear ribonucleoprotein auxiliary factor 35 kDa subunit-related protein 2	8.55	0
22	TBIU017754	isca1	Iron-sulfur cluster assembly 1 homolog, mitochondrial	68.02	0
23	TBIU020950	PCBP2	Poly(rC)-binding protein 2	23.78	0
24	TBIU017534	ACADM	Medium-chain specific acyl-CoA dehydrogenase, mitochondrial	46.83	0
25	TBIU026028	slc38a2	Sodium-coupled neutral amino acid transporter 2	18.78	0
26	TBIU021897	RBM12B	RNA-binding protein 12B	11.05	0
27	TBIU013080	fos	Proto-oncogene c-Fos	25.34	0
28	TBIU028859	Dhx8	ATP-dependent RNA helicase DHX8	10.11	0
29	TBIU023584	Bag6	Large proline-rich protein BAG6	4.23	0
30	TBIU022019	tmem66	Store-operated calcium entry-associated regulatory factor	10.57	0
31	TBIU017325	Ube2l3	Ubiquitin-conjugating enzyme E2 L3	26.21	0
32	TBIU026697	fam73a	Protein FAM73A	11.1	0
33	TBIU025540	BAG4	BAG family molecular chaperone regulator 4	18.58	0
34	TBIU019961	LGALS9	Galectin-9	22.97	0

No.	GeneId	GeneName	Description	EXP:Bbrf_WT:FPKM	EXP:Bbrf_CT:FPKM
35	TBIU029397	Aloxe3	Hydroperoxide isomerase ALOXE3	15.58	0
36	TBIU026225	ascc3	Activating signal cointegrator 1 complex subunit 3	7.24	0
37	TBIU029225	brd4	Bromodomain-containing protein 4	5.67	0
38	TBIU025184	CAC1S_CYPCA	Dihydropyridine-sensitive L-type skeletal muscle calcium channel subunit alpha-1	21.09	0
39	TBIU022079	Gm129	Uncharacterized protein C1orf51 homolog	23.03	0
40	TBIU020411	FAM57B	Protein FAM57B	12.89	0
41	TBIU024378	ACVR1	Activin receptor type-1	16.87	0
42	TBIU027072	SLC4A7	Sodium bicarbonate cotransporter 3	20.99	0
43	TBIU026562	FLNC	Filamin-C	13.49	0
44	TBIU029786	b2m	Beta-2-microglobulin	66.42	0

개볼락 체색 타입별 근육조직 전사체 분석에서 확보한 차등발현하는 유전자의 기능정보를 Gene Ontology(GO)enrichment 분석으로 확인한 결과, 164개의 DEG가 1,613개의 GO ID에 해당하고(Figure 20), DEG 2개는 molecular function, 46개는 biological process, 6개는 cellular component 기능그룹에 해당하였다(Table 10, Table11, Table 12).



Gene Ontology Enrichment

Cutoff: p-value < 0.001

No	Name	# DEG	# GO	molecular_function	biological_process	cellular_component
DEG-001	Bbrf_WT vs Bbrf_CT	164	1613	2 / 230 (0.9%)	46 / 1191 (3.9%)	6 / 192 (3.1%)

Figure 20. Histogram of the Gene Ontology classification

Table 10. Molecular function group of the differentially expressed genes

No.	ACC	Category	# Transcripts	Definition
1	GO:0050262	phosphotransferase activity, alcohol group as acceptor	5	ribosylnicotinamide kinase activity
2	GO:0030628	organic cyclic compound binding	2	pre-mRNA 3′-splice site binding

Table 11. Celluar component group of the differentially expressed genes

No.	ACC	Category	# Transcripts	Definition
1	GO:0030016	cytoplasmic part	10	myofibril
2	GO:0043292	cytoplasmic part	10	contractile fiber
3	GO:0016011	cell part	2	dystroglycan complex
4	GO:0016012	cell part	2	sarcoglycan complex
5	GO:0005742	mitochondrial membrane part	2	mitochondrial outer membrane translocase complex
6	GO:0030017	contractile fiber part	8	sarcomere

Table 12. Biological process group of the differentially expressed genes

No.	ACC	Category	# Transcripts	Definition
1	GO:0045662	regulation of muscle tissue development	5	negative regulation of myoblast differentiation
2	GO:0009435	organonitrogen compound biosynthetic process	5	NAD biosynthetic process
3	GO:0019359	organonitrogen compound biosynthetic process	5	nicotinamide nucleotide biosynthetic process
4	GO:0019363	organonitrogen compound biosynthetic process	5	pyridine nucleotide biosynthetic process
5	GO:0072525	organonitrogen compound biosynthetic process	5	pyridine-containing compound biosynthetic process
6	GO:0019674	organic cyclic compound metabolic process	5	NAD metabolic process
7	GO:0051148	muscle structure development	5	negative regulation of muscle cell differentiation
8	GO:0045661	regulation of muscle tissue development	5	regulation of myoblast differentiation
9	GO:0006094	organic substance metabolic process	5	gluconeogenesis
10	GO:0048742	regulation of muscle tissue development	5	regulation of skeletal muscle fiber development
11	GO:0070070	protein complex subunit organization	2	proton-transporting V-type ATPase complex assembly
12	GO:0070072	protein complex subunit organization	2	vacuolar proton-transporting V-type ATPase complex assembly
13	GO:0019362	organic cyclic compound metabolic process	5	pyridine nucleotide metabolic process
14	GO:0046496	organic cyclic compound metabolic process	5	nicotinamide nucleotide metabolic process
15	GO:0019319	organic substance metabolic process	5	hexose biosynthetic process
16	GO:0072524	organic cyclic compound metabolic process	5	pyridine-containing compound metabolic process
17	GO:0046364	organic substance metabolic process	5	monosaccharide biosynthetic process
18	GO:0045445	cellular developmental process	5	myoblast differentiation
19	GO:0051146	striated muscle cell differentiation	9	striated muscle cell differentiation
20	GO:0006733	cellular metabolic process	5	oxidoreduction coenzyme metabolic process
21	GO:0010830	muscle structure development	5	regulation of myotube differentiation
22	GO:0048641	regulation of muscle tissue development	5	regulation of skeletal muscle tissue development
23	GO:0051188	cofactor metabolic process	7	cofactor biosynthetic process

No.	ACC	Category	# Transcripts	Definition
24	GO:0048741	skeletal muscle organ development	5	skeletal muscle fiber development
25	GO:0051153	muscle structure development	5	regulation of striated muscle cell differentiation
26	GO:0014904	striated muscle cell development	5	myotube cell development
27	GO:0007519	skeletal muscle organ development	7	skeletal muscle tissue development
28	GO:0048747	striated muscle cell development	6	muscle fiber development
29	GO:0060538	skeletal muscle organ development	7	skeletal muscle organ development
30	GO:0070071	protein complex subunit organization	2	proton-transporting two-sector ATPase complex assembly
31	GO:0016202	regulation of muscle tissue development	5	regulation of striated muscle tissue development
32	GO:1901861	biological regulation	5	regulation of muscle tissue development
33	GO:0061061	anatomical structure development	12	muscle structure development
34	GO:0048634	biological regulation	5	regulation of muscle organ development
35	GO:0055002	striated muscle cell development	7	striated muscle cell development
36	GO:0016051	organic substance metabolic process	6	carbohydrate biosynthetic process
37	GO:0007517	anatomical structure development	9	muscle organ development
38	GO:0014706	anatomical structure development	9	striated muscle tissue development
39	GO:0006006	single-organism carbohydrate metabolic process	7	glucose metabolic process
40	GO:0042692	cellular developmental process	9	muscle cell differentiation
41	GO:0051147	muscle structure development	5	regulation of muscle cell differentiation
42	GO:0030150	establishment of protein localization to organelle	2	protein import into mitochondrial matrix
43	GO:0060537	anatomical structure development	9	muscle tissue development
44	GO:0055001	muscle cell development	7	muscle cell development
45	GO:0014902	striated muscle cell differentiation	5	myotube differentiation
46	GO:0051186	cellular metabolic process	7	cofactor metabolic process

제 4 장 연구개발 목표 달성도 및 대외기여도

제1절 연구개발 목표 달성도

연구목표 및 달성도 평가	달성도 (%)
○ 유용 어류자원 탐색 및 볼락류 유전자 정보 탐색 - 동해서식 볼락류(우럭볼락, 흰꼬리볼락, 황볼락, 노랑볼락)의 미토콘드리아 염기서열을 분석하였고, 서열정보는 GenBank에 등록하였음	100
○ 볼락류의 체색 관련 기능유전체 정보 탐색 - 개볼락 체색 타입별 근육조직 전사체 정보를 분석하였음 - 개볼락 체색 타입별 차등발현 또는 특이발현 유전자의 염기서열 및 아미노산 서열 정보를 뱅크화하였고, 기능을 분석하였음	100

제2절 대외 기여도

1. 주요 연구성과의 우수성

동해서식 볼락류의 유전자 정보를 분석하였고, 체색과 관련된 기능유전체를 탐색하여 다음과 같은 결과를 도출하였다.

가. 동해서식 볼락류의 유전자 정보 등록

- 우럭볼락 미토게놈 염기서열(16,453 bp) 정보를 GenBank에 등록

- 흰꼬리볼락 미토게놈 염기서열 (16,445 bp) 정보를 GenBank에 등록
- 황볼락 미토게놈 염기서열(16,465 bp) 정보를 GenBank에 등록
- 노랑볼락 미토게놈 염기서열(16,450) 정보를 GenBank에 등록

나. 동해서식 개볼락의 근육조직 발현 전사체 정보 뱅크화

- 개볼락 체색 타입별 근육조직 전사체의 unigene을 확보(101,112개)하여 35,482개의 기능이 밝혀진 전사체와 65,630개의 기능이 밝혀지지 않은 전사체의 염기서열 및 아미노산 서열 정보를 뱅크화하였음

다. 동해서식 개볼락의 체색타입별 근육조직에서 차등발현 유전자 확보

- 개볼락 color type 근육조직에서 발현되는 53,499개 유전자의 발현량 계산
- 개볼락 wild type 근육조직에서 발현되는 55,508개 유전자의 발현량 계산
- 개볼락 체색 타입별 근육조직에서 차등발현하는 유전자 164개 확보
- 개볼락 color type 근육조직에서 차등발현하는 유전자 83개와 color type 근육조직에서만 특이적으로 발현되는 51개의 유전자 확보
- 개볼락 wild type 근육조직에서 차등발현하는 유전자 81개와 wild type 근육조직에서만 특이적으로 발현되는 68개의 유전자 확보

라. 동해서식 개볼락의 체색관련 차등발현 유전자의 기능 분석

- Molecular function에 해당하는 2개의 차등발현 유전자 확보
- Biological process에 해당하는 46개의 차등발현 유전자 확보
- Celluar component에 해당하는 6개의 차등발현 유전자 확보

2. 연구실적

구 분			최종 목표	(2014)	
				목표	실적
논문 게재	SCI(E)		2	2	4
	기타 국외학술지				
	국내 학술지				
학회 발표	국외				
	국내				
특허	국외	출원			
		등록			
	국내	출원			
		등록			
저서					
세미나(워크샵), 전문가활용 등					

가. SCI 논문

1. Complete mitochondrial genome of Sebastes hubbsi (Scorpaenidae, Scorpaeniformes) from the East Sea, Korea
학술지명 / doi : Mitochondrial DNA / 10.3109/19401736.2014.971311

2. Complete mitochondrial genome of Sebastes longispinis (Scorpaenidae, Scorpaeniformes) from the East Sea, Korea
학술지명 / doi : Mitochondrial DNA / 10.3109/19401736.2014.987240

3. Complete mitochondrial genome of Sebastes owstoni (Scorpaenidae, Scorpaeniformes) from the East Sea, Korea
학술지명 / doi : Mitochondrial DNA / 10.3109/19401736.2014.989501

4. Complete mitochondrial genome of Sebastes steindachneri (Scorpaenidae, Scorpaeniformes) from the East Sea, Korea
학술지명 / doi : Mitochondrial DNA / 10.3109/19401736.2014.1003834

제 5 장 연구개발 결과의 활용계획

동해서식 볼락류의 자원특성 조사 및 유전자 정보 탐색 분야에서 확인한 볼락류의 미토게놈 정보는 유용어류자원의 정보뱅크화 작업에 필요한 기초자료이다. 해양어류의 미토게놈 정보뱅크는 어류의 서식환경과 관련된 유전특성을 파악하고, 개체군 수준에서 서식지 적응 또는 이동에 따른 개체군 수준의 유전자 변이 파악에 이용될 수 있다.

본 연구사업 수행으로 확보한 개볼락의 체색타입별 근육조직 전사체 정보 및 차등발현유전자는 기능유전체 정보 이용 기술 개발 연구를 위한 기초자료로 활용할 계획이다. 또한, 어류의 체색 패턴과 유전현상을 이해하고, 기능유전체 정보를 이용한 유전자마커 등 분자유전학적 도구 개발에도 이용할 예정이다. 향후 연구에서는 어류의 체색 결정 또는 생명현상 유지에 필요한 유전자의 기능 및 작용기작을 밝히고, 다양한 생물의 유전체 빅데이터를 융복합한 목표지향적인 기능유전체 활용 연구와 유전체 정보의 산업화를 위한 이용기술을 개발하고자 한다.

제 6 장 참고문헌

Agilent Technologies. (2002) Quantitation comparison of total RNA using the Agilent 2100 bioanalyzer, ribogreen analysis, and UV spectrometry. Agilent Application Note, Publication Number 5988-7650EN.

Anders S. and Huber W. (2010) Differential expression analysis for sequence count data. Genome Biol. 11(10):R106.

Blanco E., Parra G. Guigó R. (2007) Using geneid to identify genes. Curr Protoc Bioinformatics. Chapter 4:Unit 4.3.

Colihueque N. and Araneda C. (2014) Appearance traits in fish farming: progress from classical genetics to genomics, providing insight into current and potential genetic improvement. Front Genet. 5:Article 251. doi: 10.3389/fgene.2014.00251

Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D.A., Amit I., Adiconis X., Fan L., Raychowdhury R., Zeng Q., Chen Z., Mauceli E., Hacohen N., Gnirke A., Rhind N., di Palma F., Birren B.W., Nusbaum C., Lindblad-Toh K., Friedman N., Regev A. (2011) Nat Biotechnol. 29(7):644-652.

Greenwood A.K., Cech J.N., Peichel C.L. (2012) Molecular and developmental contributions to divergent pigment patterns in marine and freshwater sticklebacks. 14(4):351-362.

Greenwood A.K., Jones F.C., Chan Y.F., Brady S.D., Absher D.M., Grimwood J., Schmutz J., Myers R.M., Kingsley D.M., Peichel C.L. (2011) The genetic basis of divergent pigment patterns in juvenile threespine sticklebacks. 107(2):155-166

Haas B.J., Papanicolaou A., Yassour M., Grabherr M., Blood P.D., Bowden J., Couger M.B., Eccles D., Li B., Lieber M., Macmanes M.D., Ott M., Orvis J., Pochet N., Strozzi F., Weeks N., Westerman R., William T., Dewey C.N., Henschel R., Leduc R.D., Friedman N., Regev A. (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 8(8):1494-1512.

Harms L., Frickenhaus S., Schiffer M., Mark F.C., Storch D., Held C., Pörtner H.O., Lucassen M. Gene expression profiling in gills of the great spider crab Hyas araneus in response to ocean acidification and warming. (2014) BMC Genomics. 12;15:789. doi: 10.1186/1471-2164-15-789.

Henning F., Jones J.C., Franchini P., Meyer A. (2013) Transcriptomics of morphological color change in polychromatic Midas cichlids. BMC Genomics. 13;14:171. doi: 10.1186/1471-2164-14-171.

Huang X. and Madan A. (1999) CAP3: A DNA sequence assembly program. Genome Res. 9, 868-877.

Jiang Y., Zhang S., Xu J., Feng J., Mahboob S., Al-Ghanim K.A., Sun X., Xu P. (2014) Comparative transcriptome analysis reveals the genetic basis of skin color variation in common carp. PLoS One. 25;9(9):e108200. doi: 10.1371/journal.pone.0108200.

Kadota K., Nishiyama T., Shimizu K. (2012) A normalization strategy for comparing tag count data. Algorithms Mol Biol. 7(1):5. doi: 10.1186/1748-7188-7-5.

Kim H.A., Lim C.J., Kim S., Choe J.K., Jo S.H., Baek N., Kwon S.Y. (2014) High-throughput sequencing and de novo assembly of Brassica oleracea var. Capitata L. for transcriptome analysis. PLoS One. 28;9(3):e92087. doi: 10.1371/journal.pone.0092087.

Kuraku S, Meyer A. (2008) Genomic analysis of cichlid fish 'natural mutants'. Curr Opin Genet Dev. 18(6):551-558

Li B. and Dewey C.N. (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics, 4;12:323. doi: 10.1186/1471-2105-12-323.

Li H. and Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform, Bioinformatics. 25(14):1754-1760.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R; 1000 Genome Project Data Processing Subgroup. (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics.

25(16):2078-2079.

Lister J.A., Robertson C.P., Lepage T., Johnson S.L., Raible D.W. (1999) nacre encodes a zebrafish microphthalmia-related protein that regulates neural-crest-derived pigment cell fate. Development. 126(17):3757-3767.

Manfred S., Susanne K., Brigitta W., Toni W., Christiaan V. H., Herman P. S., Svenja M. (2012) Conserved Expression Signatures between Medaka and Human Pigment Cell Tumors. 7(5):e37880. doi: 10.1371/journal.pone.0037880

Martin J.A. and Wang Z. (2011) Next-generation transcriptome assembly. Nat Rev Genet. 12(10):671-682.

Metz H.C., Manceau M., Hoekstra H.E. (2011) Turing patterns: how the fish got its spots. Pigment Cell Melanoma Res. 24(1):12-14.

Parichy D.M., Mellgren E.M., Rawls J.F., Lopes S.S., Kelsh R.N., Johnson S.L. (2000) Mutational analysis of endothelin receptor b1 (rose) during neural crest and pigment pattern development in the zebrafish Danio rerio. Dev Biol. 15;227(2):294-306.

Parichy D.M., Rawls J.F., Pratt S.J., Whitfield T.T., Johnson S.L. (1999) Zebrafish sparse corresponds to an orthologue of c-kit and is required for the morphogenesis of a subpopulation of melanocytes, but is not essential for hematopoiesis or primordial germ cell development. Development. 126(15):3425-3346.

Pertea G., Huang X., Liang F., Antonescu V., Sultana R., Karamycheva S., Lee Y., White J., Cheung F., Parvizi B., Tsai J., Quackenbush J. (2003) TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics, 19(5):651-652.

Qian X., Ba Y., Zhuang Q., Zhong G. (2014) RNA-Seq technology and its application in fish transcriptomics. OMICS. 18(2):98-110

Santos M.E., Braasch I., Boileau N., Meyer B.S., Sauteur L., Böhne A., Belting H.G., Affolter M., Salzburger W. (2014) The evolution of cichlid fish egg-spots is linked with a cis-regulatory change. Nat Commun. 9(5):5149. doi:

10.1038/ncomms6149.

Sun J., Nishiyama T., Shimizu K., Kadota K. (2013) TCC: an R package for comparing tag count data with robust normalization strategies, BMC Bioinformatics, 14:219. doi: 10.1186/1471-2105-14-219.

Wang C., Wachholtz M., Wang J., Liao X., Lu G. (2014) Analysis of the skin transcriptome in two oujiang color varieties of common carp. PLoS One. 6;9(3):e90074. doi: 10.1371/journal.pone.0090074.

Watanabe M., Iwashita M., Ishii M., Kurachi Y., Kawakami A., Kondo S., Okada N. (2006) Spot pattern of leopard Danio is caused by mutation in the zebrafish connexin41.8 gene. EMBO Rep. 7(9):893-897.

Wolf J.B. Principles of transcriptome analysis and gene expression quantification: an RNA-seq tutorial. (2013) Mol Ecol Resour. 13(4):559-572

Yan B., Liu B., Zhu C.D., Li K.L., Yue L.J., Zhao J.L., Gong X.L., Wang C.H. (2013) microRNA regulation of skin pigmentation in fish. J Cell Sci. 1;126(Pt 15):3401-3408.

Yang Y. and Smith S.A. (2013) Optimizing de novo assembly of short-read RNA-seq data for phylogenomics. BMC Genomics. 14:328.

뒷 면

주 의

1. 이 보고서는 한국해양과학기술원에서 수행한 주요사업의 연구결과보고서입니다.

2. 이 보고서 내용을 발표할 때에는 반드시 한국해양과학기술원에서 수행한 주요사업의 연구결과임을 밝혀야 합니다.

3. 국가과학기술 기밀유지에 필요한 내용은 대외적으로 발표 또는 공개하여서는 안됩니다.