

# 멸종위기야생동물 II급 *Satsuma myomphala* (거제외줄달팽이)의 arginine kinase 유전자를 이용한 계통학적 분석

상민규, 황희주, 정종민, 박지은, 송대권, 정준양, 박소영<sup>1</sup>, 박홍석<sup>2</sup>, 조용훈<sup>3</sup>, 이종대<sup>4</sup>, 이준상<sup>5</sup>, 이용석, 강세원<sup>6</sup>

순천향대학교 자연과학대학 생명시스템학과, <sup>1</sup>국립낙동강생물자원관 다양성연구팀, <sup>2</sup>㈜지앤시바이오, <sup>3</sup>전남대학교 농업생명과학대학 식물생명공학부, <sup>4</sup>순천향대학교 자연과학대학 환경보전학과, <sup>5</sup>순천향대학교 기초과학연구소, <sup>6</sup>한국생명공학연구원 생물자원센터

## Phylogenetic analysis of endangered wild animal Class II *Satsuma myomphala* using the arginine kinase gene

Min Kyu Sang, Hee-Ju Hwang, Jong Min Chung, Jie Eun Park, Dae Kwon Song, Jun Yang Jeong, So Young Park<sup>1</sup>, Hong seog Park<sup>2</sup>, Yong Hun Jo<sup>3</sup>, Jong dae Lee<sup>4</sup>, Jun Sang Lee<sup>5</sup>, Yong Seok Lee and Se Won Kang<sup>6</sup>

Department of Life Science and Biotechnology, College of Natural Sciences, Soonchunhyang University, Asan, Chungnam, 31538, Korea

<sup>1</sup>Biodiversity Research team, Nakdonggang National Institute of Biological Resources, Sangju, Gyeongbuk, 37242, Korea

<sup>2</sup>Research Institute, GnC BIO Co., LTD., 621-6 Banseok-dong, Yuseong-gu, Daejeon, 34069, Korea

<sup>3</sup>Department of Applied Biology, Institute of Environmentally-Friendly Agriculture (IEFA), College of Agriculture and Life Sciences, Chonnam National University, Gwangju 61186, Korea

<sup>4</sup>Department of Environmental Health Science, College of Natural Sciences, Soonchunhyang University, Asan, Chungcheongnam-do, Korea

<sup>5</sup>Institute for Basic Sciences, College of Natural Sciences, Soonchunhyang University, Asan, Chungnam, 31538, Korea

<sup>6</sup>Biological Resource Center, Korea Research Institute of Bioscience and Biotechnology, Jeongseup, Jeonbuk 56212, Korea

### ABSTRACT

Arginine kinase (AK) belongs to the phosphagen kinase (PK) super family group of enzyme that has an essential role in maintaining energy homeostasis by catalyzing the re-phosphorylation of ADP. The AK gene is used for phylogenetic analysis in arthropods, mollusks, and Nematoda, and has been reported to branch in the Phylum unit. *Satsuma myomphala* has been reported as an indigenous species in Northeast Asia, including Korea and Japan. Particularly in Korea, it was registered as an endangered wildlife class 2 in 2017. Therefore, the preservation of species and securing genetic resources became more important. The genetic resources of *S. myomphala* registered on the NCBI GenBank have only six nucleotides and two proteins registered. The AK sequence was extracted from the RNAseq data which was obtained from Illumina Hiseq 2500 platform. After sequencing, *de novo* assembly and clustering, 103,774 unigenes were generated. Through annotation step by step using the PANM database and BLAST program, we obtained AK sequence of 354 amino acid residues containing a 1,062 bp coding region. The AK sequence was verified using the BLAST program in the NCBI nr database, and phylogenetic analysis with the AK sequences of Gastropoda, Bivalvia, and Cephalopoda registered in NCBI was conducted. A phylogenetic dendrogram was generated using the MEGA program, and we confirmed the phylogenetic relationships. In addition, the 2D and 3D structures of the AK gene were predicted.

**Key words** : Arginine kinase, *Satsuma myomphala*, Phylogenetics analysis

Received: June 21, 2020; Revised: June 27, 2020; Accepted: June 29, 2020

Corresponding author: Se Won Kang

Tel: +82 (63) 570-5671, e-mail: bioksw@kribb.re.kr  
1225-3480/24764

This is an Open Access Article distributed under the terms of the Creative Commons Attribution Non-Commercial License with permits unrestricted non-commercial use, distribution, and reproducibility in any medium, provided the original work is properly cited.

### INTRODUCTION

Phosphagen kinases (PK) super family group은 ADP의 재인산화를 촉매하여 에너지 항상성 유지에 중요한 역할을 하는 효소이다 (Suzuki and Furukohri, 1994; Jarilla *et al.*, 2014). 이 group에 속하는 Creatine kinase (CK)는 매우 다양한 연구가 진행되고 있고, 유일하게 척추동물에서 발견된다 (Jarilla *et al.*, 2014). 그 외에도 hypotaurocyamine

kinase (HTK), glycoyamine kinase (GK), thallemine kinase (ThK), opheline kinase (OP), lumbricine kinase (LK), taurocyamine kinase (TK), arginine kinase (AK) 등이 PK group에 속하며, 다양한 무척추동물에서 발견된다 (Uda *et al.*, 2005; Jarilla *et al.*, 2014; Min *et al.*, 2018). 이 중 AK는 다른 PK group과 달리 변형되지 않는 아미노산을 사용하며, 전형적인 기능성 단백질이다 (Uda *et al.*, 2006; Jarilla *et al.*, 2014). 또한, 무척추동물에서 가장 널리 분포된 PK group으로 원생동물, 선형동물, 절지동물, 연체동물, 중생동물 및 일부 척추동물에서도 발견되며, 연체동물의 근육에서 다량 존재하는 것으로 알려져 있다 (Suzuki *et al.*, 2000; Suzuki *et al.*, 2000; Jarilla *et al.*, 2014).

PK는 AK 및 HTK를 포함하는 AK group 및 CK, GK, LK, TK, ThK를 포함하는 CK group으로 분류된다 (Suzuki *et al.*, 2000; Uribe *et al.*, 2016). 절지동물, 선형동물, 연체동물 등에서 AK는 유일한 PK group이며, 이 세 분류군에서 AK가 계통발생학적으로 연체동물과 절지동물 및 선형동물로 분지되는 것으로 보고되어 있다 (Suzuki *et al.*, 2000; Uda *et al.*, 2006; Tanaka *et al.*, 2007; Jeong and Lee, 2013; Jarilla *et al.*, 2014).

본 연구에 사용된 *Satsuma myomphala*는 해안가 도서지역의 보존된 숲속에서 서식하며, 국내의 경상남도 거제시와 일본에 분포하는 육산 달팽이로 동북아시아 고유종이다 (Kang, 2016; Kang *et al.*, 2017). 국내에서는 종의 발생 범위와 점유 면적이 매우 적어 2012년 IUCN 적색목록 범주에는 위급 (Critically Endangered, CR)로 표기되어 멸종위기 야생생물 관찰종으로 지정되었다 (Lee and Son, 2012). 이 후 천적에 의한 포식, 관광객에 의한 압사 및 남획, 서식처의 무분별한 개발 등으로 인해 2017년 멸종위기 야생생물 2급으로 지정되었다 (서재화 *외.*, 2018). 현재 NCBI의 GeneBank에 등록된 *S. myomphala*의 유전자는 미토콘드리아 CO I 및 rRNA 서열을 비롯해 6건의 nucleotide와 2건의 protein만이 등록되어 있어, 종의 보존과 유전자원의 확보가 매우 중요하게 여겨진다 (Kang, 2016; Sang *et al.*, 2016).

본 연구에서는 *S. myomphala*에서 확인한 AK 유전자를 이용하여 현재까지 연구되어 NCBI에 등록된 연체동물의 AK 유전자와 함께 계통분류학적 분석을 진행하여 AK 유전자를 통한 보다 세밀한 계통분류의 가능성을 확인하고자 한다.

## MATERIALS AND METHODS

### 1. Materials

본 연구에서 사용된 *S. myomphala*의 분류학적 위치는 연체동물문 (Mollusca), 복족강 (Gastropoda), 병안목

(Stylommatophora), 외줄달팽이과 (Camaenidae), 외줄달팽이속 (*Satsuma*)이며, 국내에서는 거제도에만 분포하는 고립종이다 (Min, 2004). *S. myomphala*는 2014년 8월 경남 거제시 남부면 갈곶리에서 채집하였으며, 샘플을 해부하여 내장낭을 적출하여 분석에 이용하였다.

## 2. Methods

### 1) cDNA library 구축 및 염기서열 분석

*S. myomphala*에서 적출한 내장낭 조직을 냉동상태에서 분쇄하여 Trizol reagent (Thermo Fisher Scientific, USA)를 이용하여 total RNA를 추출하였다. Spectrophotometer 및 Agilent Technologies 2100 Bioanalyser를 사용하여 추출된 RNA의 quality 및 농도를 확인하고, mRNA로 정제 후 cDNA를 합성하였다. 합성된 cDNA를 통해 cDNA library를 구축하였고, Illumina HiSeq 2500 (NGS) 플랫폼을 이용하여 염기서열 분석을 수행하였다.

### 2) *S. myomphala*의 arginine kinase 서열 확보

Illumina HiSeq 2500 플랫폼에서 생산된 raw reads에서 cutadapt (v1.11) (Martin, 2011)를 사용하여 adapt 서열을 제거하고, Sickle (v1.33) (Joshi and Fass, 2011)을 이용하여 low quality sequence를 제거하여 clean reads를 확보하였다. Trinity (v2.0.4) (Grabherr *et al.*, 2011)를 이용하여 clean reads의 *de novo* assembly를 진행하여 contigs를 생성하였으며, TransDecoder를 이용하여 contigs로부터 ORF 서열을 확보하였다. 그 후 TGICL (TIGR gene Indices Clustering tools) (v2.1) (Pertea *et al.*, 2003)로 clustering 하여 unigene을 생성하고, PANM DB version3 (Kang *et al.*, 2019)와 BLAST 프로그램을 사용하여 annotation을 진행하였다. BLAST 결과, arginine kinase로 annotation된 서열들 중 full-length 서열을 확보하여 NCBI (<https://www.ncbi.nlm.nih.gov/>)의 nr (non-redundant) 데이터베이스에서 BLAST하여 arginine kinase 서열을 검증하였고, EMBOSS package (Rice *et al.*, 2000; Madeira *et al.*, 2019)의 sixpack 프로그램을 이용하여 아미노산 서열을 확보하였다.

### 3) 연체동물 (Gastropoda, Bivalvia, Cephalopoda) 내 Phylogenetic 분석

NCBI (<https://www.ncbi.nlm.nih.gov/>)에서 Gastropoda arginine kinase, Bivalvia arginine kinase, Cephalopoda arginine kinase로 검색하여 나오는 서열들 중 full-length 아미노산 서열을 받아, *S. myomphala* AK 유전자 서열 및 *Mirus junensis*의 AK 유전자 서열을 추가하여 MEGA X의

|      |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| 3    | TTT | GGT | CAC | ATT | GTC | ACT | GGG | CCT | TTA | TAC | GTG | CCT | CCT | CCT | GCG | 47   |
| 1    | F   | G   | H   | I   | V   | T   | G   | P   | L   | Y   | V   | P   | P   | P   | A   | 15   |
| 48   | CTG | GAA | GTT | CCT | TTT | TGT | GAA | GTA | TTC | ACC | CGT | GCC | GCA | ACC | TAG | 92   |
| 16   | L   | E   | V   | P   | F   | C   | E   | V   | F   | T   | R   | A   | A   | T   | *   | 30   |
| 93   | CTC | GGA | GTC | TTC | ACT | GAA | CTC | GAG | TAA | CTC | TAA | CTA | GCA | ATG | GCT | 137  |
| 31   | L   | G   | V   | F   | T   | E   | L   | E   | *   | L   | *   | L   | A   | M   | A   | 45   |
| 138  | GAT | GTA | GAT | AGT | CTG | TAC | CAA | CGG | CTC | AAA | GAC | GCC | ACC | GAG | TCC | 182  |
| 46   | D   | V   | D   | S   | L   | Y   | O   | R   | L   | K   | D   | A   | T   | E   | S   | 60   |
| 183  | AAG | TCG | CTG | GTG | AGG | AAG | TAC | CTG | ACT | CCA | AAC | ATT | TAC | ACC | CAG | 227  |
| 61   | K   | S   | L   | V   | R   | K   | Y   | L   | T   | P   | N   | I   | Y   | T   | O   | 75   |
| 228  | CTG | AAG | GGG | CTG | ACC | ACT | AAA | TTC | GGG | GGC | ACC | CTG | GCA | GAT | TGT | 272  |
| 76   | L   | K   | G   | L   | T   | T   | K   | F   | G   | G   | T   | L   | A   | D   | C   | 90   |
| 273  | ATA | CGA | TCA | GGC | TGC | GAG | AAC | ACC | GAC | AGC | GGT | GTG | GGT | GTC | TAC | 317  |
| 91   | I   | R   | S   | G   | C   | E   | N   | T   | D   | S   | G   | V   | G   | V   | Y   | 105  |
| 318  | GCC | TCT | GAT | CCC | GAC | GCC | TAC | ACG | ACC | TTC | GCC | CCT | CTG | TTG | GAC | 362  |
| 106  | A   | S   | D   | P   | D   | A   | Y   | T   | T   | F   | A   | P   | L   | L   | D   | 120  |
| 363  | GCC | GTC | ATC | AAG | GAC | TAC | CAC | AAG | GTC | TCT | GAG | CTC | AAC | CAC | CCC | 407  |
| 121  | A   | V   | I   | K   | D   | Y   | H   | K   | V   | S   | E   | L   | N   | H   | P   | 135  |
| 408  | AAC | CCA | AAC | TTC | GGA | GAT | GTG | GAC | CAT | CTG | GAC | TTT | GGG | GAT | CTC | 452  |
| 136  | N   | P   | N   | F   | G   | D   | V   | D   | H   | L   | D   | F   | G   | D   | L   | 150  |
| 453  | GAC | CCA | TCT | GGA | AAT | CTG | ATT | GTC | TCC | ACC | CGT | GTC | CGC | GTG | GGC | 497  |
| 151  | D   | P   | S   | G   | N   | L   | I   | V   | S   | T   | R   | V   | R   | V   | G   | 165  |
| 498  | CGC | AGC | CAT | GAC | AGC | TAT | GGC | TTC | CCT | CCA | GTA | CTG | ACC | AAG | GAG | 542  |
| 166  | R   | S   | H   | D   | S   | Y   | G   | F   | P   | P   | V   | L   | T   | K   | E   | 180  |
| 543  | GAT | CTT | GTG | GAG | ATG | GAG | AAG | AGG | ACA | GTT | GAG | GCT | TTC | AAG | AAG | 587  |
| 181  | D   | L   | V   | E   | M   | E   | K   | R   | T   | V   | E   | A   | F   | K   | K   | 195  |
| 588  | CTG | GAT | GGA | GAG | TTG | AAG | GGA | ACA | TAT | CAT | CCA | CTG | ACC | GGT | ATG | 632  |
| 196  | L   | D   | G   | E   | L   | K   | G   | T   | Y   | H   | P   | L   | T   | G   | M   | 210  |
| 633  | AGC | AAA | GAA | ACA | CAG | AAG | CAG | CTC | ACT | GAG | GAT | CAT | TTC | CTC | TTC | 677  |
| 211  | S   | K   | E   | T   | O   | K   | O   | L   | T   | E   | D   | H   | F   | L   | F   | 225  |
| 678  | AAC | GAT | AGT | GAC | AGG | TTC | TTG | AGA | GCG | GCC | GGG | GGC | TAC | CAA | AAC | 722  |
| 226  | N   | D   | S   | D   | R   | F   | L   | R   | A   | A   | G   | G   | Y   | O   | N   | 240  |
| 723  | TGG | CCC | TTA | GGG | CGC | GGT | ATT | TTC | TTC | AAT | GAC | AAC | AAG | ACT | TTC | 767  |
| 241  | W   | P   | L   | G   | R   | G   | I   | F   | F   | N   | D   | N   | K   | T   | F   | 255  |
| 768  | TTG | GTG | TGG | GTC | AAT | GAA | GAG | GAT | CAC | TTG | AGG | TTC | ATT | TCT | ATG | 812  |
| 256  | L   | V   | W   | V   | N   | E   | E   | D   | H   | R   | R   | F   | I   | S   | M   | 270  |
| 813  | CAG | AAA | GGT | GGC | AAT | CTG | AAG | GAG | GTT | TAC | GTC | AGA | CTC | GTC | AAG | 857  |
| 271  | O   | K   | G   | G   | N   | L   | K   | E   | V   | Y   | V   | R   | L   | V   | K   | 285  |
| 858  | GCA | ATC | AGA | GCT | CTA | GAG | AGC | AGT | GGT | CTC | TCC | TTT | GCC | AAA | CGA | 902  |
| 286  | A   | I   | R   | A   | L   | E   | S   | S   | G   | L   | S   | F   | A   | K   | R   | 300  |
| 903  | GAA | GGT | CTA | GGT | TAC | CTG | ACC | TTC | TGT | CCA | TCA | AAC | TTG | GGA | ACC | 947  |
| 301  | E   | G   | L   | G   | Y   | L   | T   | F   | C   | P   | S   | N   | L   | G   | T   | 315  |
| 948  | ACT | CTT | CGT | GCC | TCA | GTG | CAT | ATC | AAG | ATC | CCC | AAA | CTG | GCT | GCG | 992  |
| 316  | T   | L   | R   | A   | S   | V   | H   | I   | K   | I   | P   | K   | A   | A   | A   | 330  |
| 993  | ACA | CCA | GAC | TTC | AAG | GCT | TTC | TGT | GAC | AAA | TAC | AAC | ATC | CAG | GCC | 1037 |
| 331  | T   | P   | D   | F   | K   | A   | F   | C   | D   | K   | Y   | N   | I   | O   | A   | 345  |
| 1038 | AGA | GGT | ATC | CAT | GGA | GAG | CAC | ACA | GAG | TCA | GTC | GGG | GGC | GTG | TAC | 1082 |
| 346  | R   | G   | I   | H   | G   | E   | H   | T   | E   | S   | V   | G   | G   | V   | Y   | 360  |
| 1083 | GAC | ATC | TCC | AAC | AAG | AGA | AGA | CTC | GGG | CTC | AGG | GAA | ATT | GAC | GCC | 1127 |
| 361  | D   | I   | S   | N   | K   | R   | R   | L   | G   | L   | R   | E   | I   | D   | A   | 375  |
| 1128 | ATT | CAA | GAG | ATG | CGC | CGA | GGT | GTC | GAG | GCT | GTC | ATT | GCC | GAG | GAG | 1172 |
| 376  | I   | O   | E   | M   | R   | R   | G   | V   | E   | A   | V   | I   | A   | E   | E   | 390  |
| 1173 | AAG | AAA | TTG | GGA | GGT | GGC | AAC | TAG | GCA | ATC | CAG | CAA | ACA | ACA | AGA | 1217 |
| 391  | K   | K   | L   | G   | G   | G   | N   | *   | A   | I   | O   | O   | T   | T   | R   | 405  |
| 1218 | TGC | ACT | ACT | CAA | CCA | ACC | TAG | CCC | AGT | TTT | CCA | CCG | TCT | CCT | CCT | 1262 |
| 406  | C   | T   | T   | O   | P   | T   | *   | P   | S   | F   | P   | P   | S   | P   | P   | 420  |

Fig. 1. The arginine kinase nucleotide sequence and deduced amino acid sequence of *S. myomphala*.

**Table 1.** List of species used in phylogenetic analysis

| Class                           | Species                                 | Accession number |
|---------------------------------|---|------------------|
| Bivalvia                        | <i>Scapharca broughtonii</i>            | BAD11949         |
|                                 | <i>Pholas orientalis</i>                | ACP43447.1       |
|                                 | <i>Crassostrea gigas</i>                | EKC24881         |
|                                 | <i>Azumapecten farreri</i>              | AEX08673         |
|                                 | <i>Mizuhopecten yessoensis</i>          | OWF45839         |
|                                 | <i>Hyriopsis schlegelii</i>             | AEO94538         |
|                                 | <i>Corbicula japonica</i>               | BAB91357.1       |
|                                 | <i>Archivesica packardana</i>           | AXE71657.1       |
|                                 | <i>Ensis directus</i>                   | AAM90698.1       |
|                                 | <i>Paphia undulata</i>                  | ACP43446.1       |
|                                 | <i>Meretrix lyrata</i>                  | ACP43445.1       |
|                                 | <i>Meretrix meretrix</i>                | ACP43444.1       |
|                                 | <i>Calyptogena kaikoi</i>               | BAE16974.1       |
|                                 | <i>Solen strictus</i>                   | BAB91358.1       |
| Cephalopoda                     | <i>Nautilus pompilius</i>               | BAA95594         |
|                                 | <i>Amphioctopus fangsiao</i>            | AEK65120         |
|                                 | <i>Octopus vulgaris</i>                 | BAA95609         |
|                                 | <i>Sepiella maindroni</i>               | AEK26855         |
|                                 | <i>Sepia pharaonis</i>                  | AKS26488         |
|                                 | <i>Sepioteuthis lessoniana</i>          | BAA95610         |
| Gastropoda                      | <i>Aplysia kurodai</i>                  | BAB41095         |
|                                 | <i>Aplysia californica</i>              | XP_005099408     |
|                                 | <i>Pomacea canaliculata</i>             | AYH91743         |
|                                 | <i>Haliotis diversicolor supertexta</i> | AJW60778         |
|                                 | <i>Haliotis madaka</i>                  | BAA05100.1       |
|                                 | <i>Conus araneosus</i>                  | AQM52449         |
|                                 | <i>Conus litteratus</i>                 | ARS01451         |
|                                 | <i>Conus frigidus</i>                   | ARU12142         |
|                                 | <i>Conus ebraeus</i>                    | ASF90538         |
|                                 | <i>Conus lividus</i>                    | ATG85037         |
|                                 | <i>Nesiohelix samarangae</i>            | AHC02701         |
|                                 | <i>Mirus junensis</i>                   | SCH 2018         |
|                                 | <i>Satsuma myomphala</i>                | -                |
|                                 | <i>Turbo cornutus</i>                   | BAA22870.1       |
| <i>Biomphalaria glabrata</i>    | ADH59421                                |                  |
| <i>Semisulcospira libertina</i> | AGN95434                                |                  |
| <i>Cellana grata</i>            | BAB41096                                |                  |

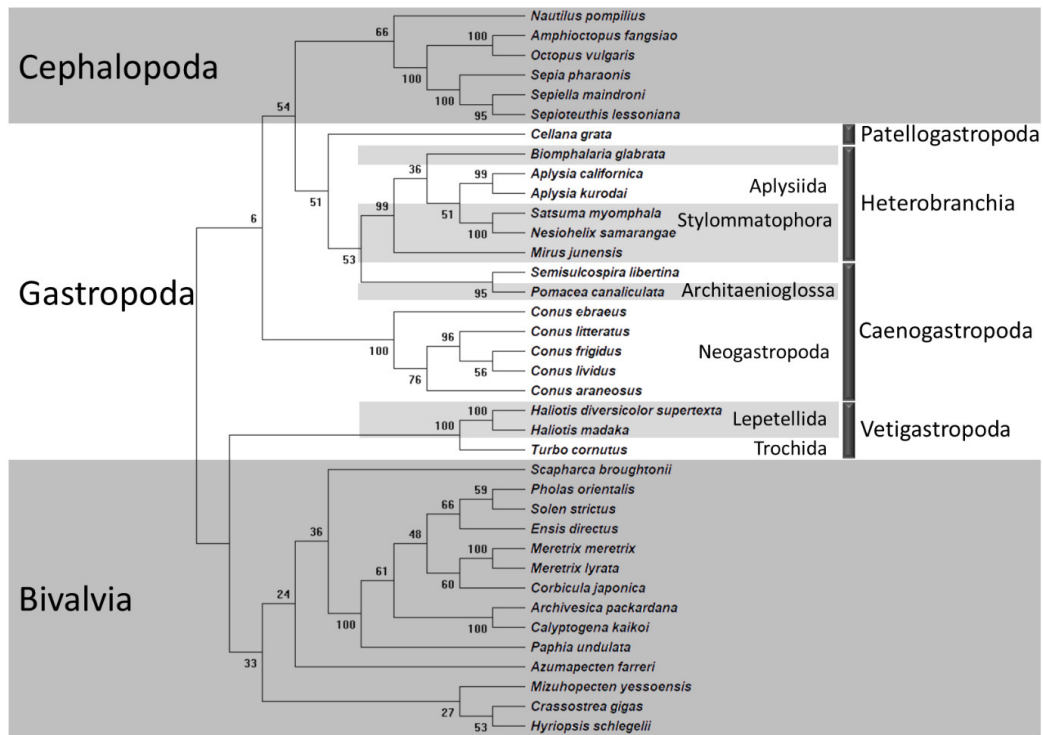
neighbor-joining 알고리즘 (1000 bootstrap replication) 으로 phylogenetic 분석을 진행하여 dendrogram을 완성하였다 (Kumar *et al.*, 2018).

#### 4) Arginine kinase의 단백질 구조 예측

확보한 AK 유전자의 단백질 구조를 예측하기 위해 Psipred 4.0 (McGuffin *et al.*, 2000; Buchan and Jones, 2019) 및 I-Tasser (Roy *et al.*, 2010; Grabherr *et al.*, 2011; Yang *et al.*, 2015; Yang and Zhang, 2015) 를 이용하여 단백질의 2차 구조 및 3차 구조 예측을 진행하였다.

## RESULTS AND DISCUSSION

*S. myomphala*의 전사체 분석을 통해 얻은 103,774개 unigene 중 PANM DB에 BLAST를 수행하여 arginine kinase로 추정되는 4개의 서열을 확보하였으며, 정확도를 확인하기 위해 NCBI에 BLAST를 수행하고 sequence alignment를 확인하여 검증하였다. 그 결과 *S. myomphala* AK 유전자의 coding region은 methionine으로 시작하여 종 결코돈까지 1,062개 염기서열 (354개 아미노산) 로 이루어져 있었으며, 염기서열의 GC contents는 52.5%였다 (Fig. 1).



**Fig. 2.** Results of phylogenetic analysis of evolutionary relationships using arginine kinase. The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei, 1987). The bootstrap consensus tree inferred from 1000 replicates (Felsenstein, 1985) is taken to represent the evolutionary history of the taxa analyzed (Felsenstein, 1985). Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein, 1985). The evolutionary distances were computed using the JTT matrix-based method (Jones *et al.*, 1992) and are in the units of the number of amino acid substitutions per site. The analysis involved 40 amino acid sequences. All ambiguous positions were removed for each sequence pair. There were a total of 774 positions in the final dataset. Evolutionary analyses were conducted in MEGA X (Kumar *et al.*, 2018).

연체동물문의 class에 따른 분류를 확인하기 위하여 NCBI에서 Bivalvia, Gastropoda, Cephalopoda의 AK 유전자들 중 full-length를 가지는 AK 유전자의 아미노산 서열과 *M. junensis*의 AK 유전자 (Min *et al.*, 2018) 까지 포함하여 MEGA X 프로그램으로 bootstrap 1000, Jones-Taylor-Thornton (JTT) model을 이용한 neighbor-joining method로 총 Bivalvia 14종, Cephalopoda 6종, Gastropoda 17종에 대하여 phylogenetic tree를 도식화 하였다. 그 결과 각 phylum은 이전 AK 유전자 phylogenetic 연구와 같이 잘 묶이는 것을 확인할 수 있었다. 또한 6종의 Cephalopoda는 Nautilida, Octopoda, Sepiida, Teuthida의 4개 order로 분지되었으며, 14종의 Bivalvia는 Veneroidea, Myoidea, Unionida, Ostreoida, Pectinoidea, Arcoidea의 6개 order로 분지되었다. 마지막 Gastropoda의

경우 4개의 subclass와 6개의 order로 분지되어 묶이는 것을 확인할 수 있었다 (Table 1 and Fig. 2).

계통분류학적 분석 결과 Gastropoda에 속하는 5개 order (Stylommatophora, Aplysiida, Architaenioglossa, Neogastropoda, Lepetellida) 에 포함되는 종들의 AK 유전자 2D 구조를 PSIPRED 4.0 프로그램을 이용하여 예측하여, 비교하였다. 대상으로 한 10종의  $\alpha$ -helix 구조는 육상에 서식하는 *S. myomphala*와 *Nesiohelix samarangae*의 경우 16개, 그 외 8종의 수생 Gastropoda는 13-14개로 예측되었으며,  $\beta$ -pleated sheet 구조는 10종 모두 10-12개로 비슷하게 예측되었다 (Fig. 3). 이를 통해 Gastropoda 중 육상 생활을 하는 종들의 AK 유전자에는 수생 종들의 AK 유전자에 비해 2개 이상의  $\alpha$ -helix 구조를 더 가지는 것을 확인할 수 있었다.

또한 *S. myomphala*의 3D 구조를 예측하여 이전에 보고된

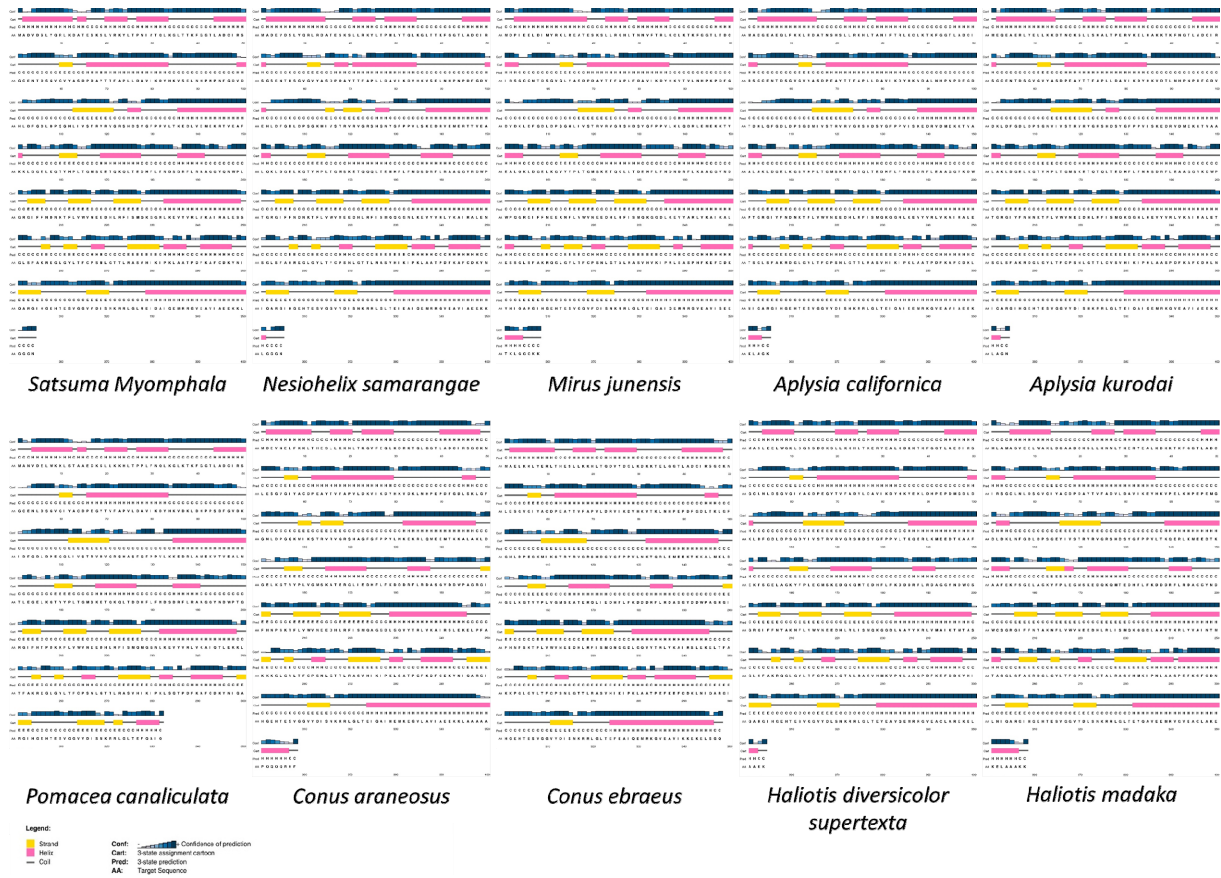


Fig. 3. The predicted 2D structure of arginine kinase by order included in Gastropoda using PSIPRED 4.0.

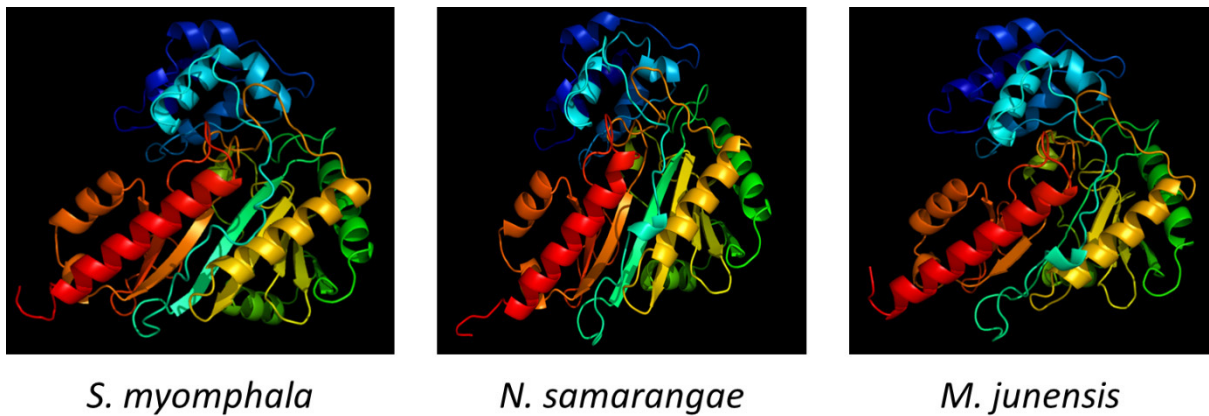


Fig. 4. The predicted 3D structure of arginine kinase belonging to Stylommatophora by I-Tasser.

*N. samarangae*, *M. junensis*의 3D 구조와 비교하였다 (Fig. 4). 예측된 *S. myomphala*의 3D 구조는 figure 3에 예측된 2D 구조와 같은  $\alpha$ -helix 및  $\beta$ -pleated sheet의 개수와 같았다. 또한 이전에 보고된 *N. samarangae*, *M. junensis*와

비슷한 모양의 입체구조를 가지고 있으나,  $\alpha$ -helix 및  $\beta$ -pleated sheet의 위치와 개수 및 coil의 형태에 약간의 차이가 있었다.

AK 유전자를 이용한 phylogenetic analysis는 지속적으로

이루어져 왔는데, 그 수준은 AK 유전자와 같은 PK super family group에 속하는 CK 유전자, HTK 유전자 등과의 진화적 상관관계를 파악하는 것과 Mollusca 내에서의 class 또는 subclass 수준에서 계통분류를 확인하는 연구가 주를 이루었다 (Suzuki *et al.*, 1994; Suzuki *et al.*, 1997; Suzuki *et al.*, 2000; Suzuki *et al.*, 2000; Uda *et al.*, 2006; Mahon and Neigel, 2008; Uda *et al.*, 2008). 본 연구에서 나타난 결과는 이러한 한계에서 벗어나 AK 유전자를 비롯한 PK super family group에 속하는 유전자들이 Mollusca의 중간 특징을 설명하고, 서식 환경을 유추하는데 사용될 수 있을 것으로 생각된다.

## CONCLUSIONS

*S. myomphala*의 AK 유전자는 354개의 아미노산 (1,062 bp 염기서열) 으로 이루어져 있으며, 52.5%의 GC contents 를 가진다. NCBI GenBank에 등록된 Mollusca AK 유전자 서열과 Min *et al.*, 2018에 보고된 *M. junensis*의 AK 유전자 서열을 포함하여 39종의 AK 유전자 서열을 참고서열로 하여 phylogenetic analysis를 진행하였다. 그 결과 선행된 연구와 같이 Cephalopoda, Gastropoda, Bivalvia의 3개의 calss로 분류되는 것을 확인할 수 있었으며 (Uda *et al.*, 2008; Jarilla *et al.*, 2014), subclass 및 order 단위의 분류군으로도 묶이는 것을 확인하였다. 또한 AK 유전자의 2D 구조를 예측하여 Gastropoda에서 육산종과 수생종간의 AK 유전자가  $\alpha$ -helix 구조상에서 구조적으로 차이가 나는 것을 확인하였다. 이러한 결과를 통해 PK super family에 속하는 유전자에 대한 연구가 지속된다면, AK 유전자를 이용한 Mollusca의 세부적인 분류 가능성을 확인하였다.

## ACKNOWLEDGEMENT

본 연구는 KRIBB 기관고유사업과 교육부 (한국연구재단, NRF-2017R1D1A3B06034971) 의 연구비 지원에 의해 수행되었습니다.

## REFERENCE

- Buchan, D.W.A., and Jones, D.T. (2019) The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Res.*, **47**: W402-W407.
- Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *evolution*, **39**: 783-791.
- 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., and Regev, A. (2011) Full-length transcriptome assembly by RNA-Seq data without a reference genome. *Nat. Biotechnol.*, **29**: 644-652.
- Jarilla, B.R., Uda, K., Suzuki, T., Acosta, L.P., Urabe, M., and Agatsuma, T. (2014) Characterization of arginine kinase from the caenogastropod *Semisulcospira libertina*, an intermediate host of *Paragonimus westermani*. *Journal of Molluscan Studies*, **80**: 444-451.
- Jeong, J.E., and Lee, Y.S. (2013) Identification, sequence characterization and expression analysis of the arginine kinase gene in response to laminarin challenge from the Oriental land snail, *Nesiohelix samarangae*. *The Korean Journal of Malacology*, **29**: 171-179.
- Jones, D.T., Taylor, W.R., and Thornton, J.M. (1992) The rapid generation of mutation data matrices from protein sequences. *Bioinformatics*, **8**: 275-282.
- Joshi, N., and Fass, J. (2011) Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33)[Software].
- Kang, S.W. (2016) Genome and transcriptome characterization of the four Korean land snails using NGS and bioinformatics databases. Department of Biology, Graduate School of Soonchunhyang University
- Kang, S.W., Park, S.Y., Hwang, H.J., Chung, J.M., Sang, M.K., Min, H.R., Park, J.E., Cho, H.C., Patnaik, B.B., and Lee, Y.S. (2019) PANM DB ver 3.0 : An update of the bioinformatics database for annotation of large datasets from sequencing of species under Protostomia clade. *The Korean Journal of Malacology*, **35**: 73-75.
- Kang, S.W., Patnaik, B.B., Hwang, H.J., Park, S.Y., Chung, J.M., Song, D.K., Patnaik, H.H., Lee, J.B., Kim, C., Kim, S., Park, H.S., Park, S.H., Park, Y.S., Han, Y.S., Lee, J.S., and Lee, Y.S. (2017) Sequencing and *de novo* assembly of visceral mass transcriptome of the critically endangered land snail *Satsuma myomphala*: Annotation and SSR discovery. *Comp Biochem. Physiol. Part D Genomics Proteomics*, **21**: 77-89.
- Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018) MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.*, **35**: 1547-1549.
- Lee, J.S., and Son, M.-H. (2012) Red Data Book of Endangered Mollusks in Korea pp. 40-41. National Institute of biological resources.
- Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R.N., Potter, S.C., Finn, R.D., and Lopez, R. (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.*, **47**: W636-W641.
- Mahon, B.C., and Neigel, J.E. (2008) Utility of arginine kinase for resolution of phylogenetic relationships among Brachyuran genera and families. *Mol.*

- Phylogenet Evol.*, **48**: 718-727.
- Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011*, **17**: 3.
- McGuffin, L.J., Bryson, K., and Jones, D.T. (2000) The PSIPRED protein structure prediction server. *Bioinformatics*, **16**: 404-405.
- Min, D.-K. (2004) Mollusks in Korea pp. 566. Min Molluscan Research Institute.
- Min, H.R., Hwang, H.J., Chung, J.M., Sang, M.K., Cho, H.C., Park, J.E., Jung, K.Y., Park, H.S., Han, Y.S., and Lee, Y.S. (2018) Molecular Phylogenetic studies of *Mirus junensis* using Arginine kinase gene sequence. *The Korean Journal of Malacology*, **34**: 107-114.
- Pertea, G., Huang, X., Liang, F., Antonescu, V., Sultana, R., Karamycheva, S., Lee, Y., White, J., Cheung, F., Parvizi, B., Tsai, J., and Quackenbush, J. (2003) TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics*, **19**: 651-652.
- Rice, P., Longden, I., and Bleasby, A. (2000) EMBOSS: the European molecular biology open software suite. Elsevier current trends.
- Roy, A., Kucukural, A., and Zhang, Y. (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.*, **5**: 725-738.
- Saitou, N., and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution*, **4**: 406-425.
- Sang, M.K., Kang, S.W., Hwang, H.-J., Chung, J.M., Song, D.K., Min, H.R., Park, J.E., Ha, H.C., Lee, H.J., Hong, C.E., Ahn, Y.M., Park, S.Y., Park, Y.-S., Park, H.S., Han, Y.S., Lee, J.S., and Lee, Y.S. (2016) Molecular Phylogenetic Study of the Endangered Land Snail *Satsuma myomphala* Based on Metallothionein Gene. *The Korean Journal of Malacology*, **32**: 263-268.
- Suzuki, T., Fukuta, H., Nagato, H., and Umekawa, M. (2000) Arginine kinase from *Nautilus pompilius*, a living fossil. Site-directed mutagenesis studies on the role of amino acid residues in the Guanidino specificity region. *J. Biol. Chem.*, **275**: 23884-23890.
- Suzuki, T., and Furukohri, T. (1994) Evolution of Phosphagen Kinase: Primary Structure of Glycocyamine Kinase and Arginine Kinase from Invertebrates. *Journal of Molecular Biology*, **237**: 353-357.
- Suzuki, T., Inoue, N., Higashi, T., Mizobuchi, R., Sugimura, N., Yokouchi, K., and Furukohri, T. (2000) Gastropod arginine kinases from *Cellana grata* and *Aplysia kurodai*. Isolation and cDNA-derived amino acid sequences. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, **127**: 505-512.
- Suzuki, T., Kawasaki, Y., and Furukohri, T. (1997) Evolution of phosphagen kinase1: Isolation, characterization and cDNA-derived amino acid sequence of two-domain arginine kinase from the sea anemone *Anthopleura japonicus*. *Biochemical Journal*, **328**: 301-306.
- Tanaka, K., Ichinari, S., Iwanami, K., Yoshimatsu, S., and Suzuki, T. (2007) Arginine kinase from the beetle *Cissites cephalotes* (Olivier). Molecular cloning, phylogenetic analysis and enzymatic properties. *Insect Biochem. Mol. Biol.*, **37**: 338-345.
- Uda, K., Fujimoto, N., Akiyama, Y., Mizuta, K., Tanaka, K., Ellington, W.R., and Suzuki, T. (2006) Evolution of the arginine kinase gene family. *Comp. Biochem. Physiol. Part D Genomics Proteomics*, **1**: 209-218.
- Uda, K., Iwai, A., and Suzuki, T. (2005) Hypotaurocyamine kinase evolved from a gene for arginine kinase. *FEBS Lett.*, **579**: 6756-6762.
- Uda, K., Yamamoto, K., Iwasaki, N., Iwai, M., Fujikura, K., Ellington, W.R., and Suzuki, T. (2008) Two-domain arginine kinase from the deep-sea clam *Callyptogena kaikoi*-evidence of two active domains. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.*, **151**: 176-182.
- Uribe, J.E., Colgan, D., Castro, L.R., Kano, Y., and Zardoya, R. (2016) Phylogenetic relationships among superfamilies of Neritimorpha (Mollusca: Gastropoda). *Mol. Phylogenet Evol.*, **104**: 21-31.
- Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015) The I-TASSER Suite: protein structure and function prediction. *Nat. Methods*, **12**: 7-8.
- Yang, J., and Zhang, Y. (2015) I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res.*, **43**: W174-181.
- 서재화, 권선만, 이수연, 박창득, 이소희, 이정연, 유정선, and 서민환 (2018) 환눈에 보는 멸종위기 야생생물 (2017년 개정) pp. 332-333. 국립생물자원관.