

# PANM Database Version 6.0: An Updated and Expanded Resource for Efficient Annotation of Arthropoda, Nematoda, and Mollusca Genes

So Young Park<sup>1</sup>, Jun Yang Jeong<sup>2,3</sup>, Chan-Eui Hong<sup>2,3</sup>, Hyeonjun Shin<sup>2,3</sup>,  
Hyeok Lee<sup>2,3</sup>, Kyoung Won Lee<sup>2,3</sup>, Min Kyu Sang<sup>4</sup>, Jie Eun Park<sup>4</sup>, Dae Kwon Song<sup>4</sup>,  
Hee-Ju Hwang<sup>5</sup>, Se Won Kang<sup>6</sup>, Hong Seog Park<sup>7</sup> and Yong Seok Lee<sup>1,2,3</sup>

<sup>1</sup>Diversity Conservation Research Department Biological Specimen Conservation Division, Nakdonggang National Institute of Biological Resources, Sangju, Gyeongbuk, 37242, Korea,

<sup>2</sup>Department of Biology, College of Natural Sciences, Soonchunhyang University, Asan, Chungnam 31538, Korea,

<sup>3</sup>Korea Native Animal Resources Utilization Convergence Research Institute

<sup>4</sup>Research Support Center(Core-Facility) for Bio-Bigdata Analysis and Utilization of Biological Resources

<sup>5</sup>Soonchunhyang Institute of Medi-bio Science, Soonchunhyang University, Cheonan 31151, Korea.

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

<sup>7</sup>Research Institute, GnC BIO Co., LTD., 621-6 Banseok-dong, Yuseong-gu, Daejeon, 34069, Korea.

---

## ABSTRACT

The updated PANM database version 6.0 includes approximately 1.3 times more amino acid sequences compared to its previous version. Although it contains only about 3.6% of the sequences compared to the commonly used NCBI-nr database for annotation, PANM 6.0 was specifically constructed using amino acid sequences of genes from Arthropoda, Nematoda, and Mollusca. This allows for reduced annotation time and improved accuracy for genes, genomes, or transcriptome sequences of species belonging to these three phyla. Like previous versions, PANM database version 6 (<http://panm.sch.ac.kr/>) freely accessible to all researchers and is expected to be useful for researchers studying invertebrates in the future.

---

**Keywords:** PANM-DB, NGS analysis, invertebrate

## Introduction

With the rapid advancement of Next-Generation Sequencing (NGS) technologies, large-scale genomic and transcriptomic sequence data can be obtained at lower costs and in less time, leading to a rapid increase in the amount of NGS data stored in NCBI's Sequence Read Archive (SRA) (NCBI, <https://www.ncbi.nlm.nih.gov/genbank/statistics>, 2024). Recently, the increased use

of second-generation NGS sequencers, specifically Long read sequencers (such as PacBio SMRT and Oxford Nanopore), has resulted in a significant increase in the length and volume of Whole Genome Shotgun (WGS) data, emphasizing the importance of gene prediction and annotation (Altschul *et al.*, 1990).

In the past, it was common to use NCBI's BLAST and the NCBI nr database to annotate data obtained from NGS (McGinnis and Madden, 2004). However, due to the rapid increase in sequence data resulting from NGS advancements, using the NCBI nr database for BLAST analysis has become extremely time-consuming. To overcome this issue and improve the accuracy of annotation results, our research team established a mollusk-specific BLAST server in 2004 (Lee *et al.*, 2004). However, this server was not suitable for analyzing genetic information related to

---

Received: June 18, 2024; Revised: June 23 2024; Accepted: June 30, 2024

Corresponding author: Yong Seok Lee

Tel: +82 (41) 530-3040, e-mail: yslee@sch.ac.kr  
1225-3480/24868

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.

insects and other invertebrates. Therefore, using the similar interface, we developed the PANM (Arthropoda, Nematoda, Mollusca) database version 1.0 in 2015 (Kang *et al.*, 2015).

Due to continuous data growth, PANM DB was updated to version 2.0 in September 2016, and version 3.0 in March 2019, by incorporating additional amino acid sequences (Kang *et al.*, 2016 ; Kang *et al.*, 2019). Subsequently, in February 2021, version 4.0 was released, in February 2021, incorporating all updated amino acid sequences for Arthropoda, Nematoda, and Mollusca updated in NCBI since March 2019 (Sang *et al.*, 2021). The PANM database's rapid annotation speed has led to its active use and adoption among international researcher (Capt *et al.*, 2018). Since the update to PANM version 5.1, we have confirmed a 127% increase in the number of amino acid sequences for Arthropoda, Nematoda, and Mollusca genes registered in NCBI by

of June 2024. To enhance result quality, including annotation accuracy, we have updated the PANM database by incorporating the latest genetic information.

## Materials and Methods

### 1. Database Update

To update PANM database version 5.1, we downloaded all amino acid sequences for genes within Arthropoda, Nematoda, and Mollusca registered in NCBI from October 2022 to June 2024. The acquired amino acid sequences of genes were integrated to PANM database version 5.0. and formatted using the formatdb program to facilitate BLAST execution.

### 2. Web Interface Improvement

The PANM database version 6.0 was newly applied

**Table 1.** Status of updated amino acid sequence in PANM DatabBase version 6.0

		PANM-DB	Arthropoda	Nematoda	Mollusca
Total Sequences	Ver. 1	4,051,323	3,111,849	652,125	287,349
	Ver. 2	7,571,246	6,178,888	964,027	428,331
	Ver. 3	11,612,243	9,313,972	1,585,378	712,893
	Ver. 4	15,889,991	12,943,654	1,858,096	1,088,241
	Ver. 5	21,276,123	17,384,642	2,014,594	1,876,887
	Ver. 6	27,028,682	21,639,412	2,865,367	2,523,903
	Rate of Increase (1 → 5)	525%	558%	308%	653%
	Rate of Increase (2 → 5)	281%	280%	312%	438%
	Rate of Increase (3 → 5)	183%	186%	127%	263%
	Rate of Increase (4 → 5)	133%	134%	108%	172%
	Rate of Increase (5 → 6)	127%	124%	142%	134%
Total Letters	Ver. 1	1,552,609,669	1,196,730,565	226,168,391	99,710,713
	Ver. 2	3,114,590,190	2,590,040,078	266,349,624	158,200,488
	Ver. 3	4,813,465,883	3,939,129,978	569,496,601	304,239,304
	Ver. 4	7,245,891,909	6,032,532,729	697,011,463	516,347,717
	Ver. 5	11,470,671,405	9,573,324,156	935,627,103	961,720,146
	Ver. 6	14,758,481,449	11,990,202,736	1,166,687,363	1,601,591,350
	Rate of Increase (1 → 5)	739%	800%	414%	965%
	Rate of Increase (2 → 5)	368%	370%	351%	608%
	Rate of Increase (3 → 5)	238%	243%	164%	316%
	Rate of Increase (4 → 5)	158%	159%	134%	186%
	Rate of Increase (5 → 6)	128%	125%	124%	166%

to the existing web server where PANM database version 5.1 was established, and the interface was modified to enable BLAST execution. Similar to version 5.0, PANM database version 6.0 is available for download, allowing many researchers to utilize it

## Results and Discussion

For the update of the PANM database, a total of 5,752,559 amino acid sequences from Arthropoda, Nematoda, and Mollusca genes registered in NCBI from September 2022 to June 2024 were downloaded. These were added to version 5.1, resulting in the construction of PANM database version 6.0, which consists of a total of 27,028,682 sequences and 14,758,481,449 amino acids. The data in Version 6.0, like the previous version, is composed of amino acid sequences from three phyla: Arthropoda, Nematoda, and Mollusca. The amino acid sequences for each phylum are as follows: Arthropoda accounts for the largest portion with 21,639,412 sequences (approximately 80.06% of the total), followed by Nematoda with 2,865,367 sequences (about 10.60%), and Mollusca with 2,523,903 sequences (about 9.34%). Compared to the previous version, this represents an increase of approximately 127% in the number of sequences and about 128% in the number of amino acids.

Examining the amino acid sequences of Arthropoda, Nematoda, and Mollusca registered in NCBI, we found that they were first registered in 1999. By June 2015, when PANM database version 1.0 was released, 4,051,323 amino acid sequences had been registered. Since then, an average of about 2 million amino acid sequences have been registered annually. This increase indicates a rise in genomic and transcriptomic research on biological resources, due to the rapid development of next-generation sequencing technologies and the implementation of the Nagoya Protocol, with research purposes and methods becoming increasingly diverse.

## Acknowledgement

This research was supported by Korea Basic Science Institute (National research Facilities and Equipment Center) grant funded by the Ministry of Education (2022R1A6C101B794), the National Research Foundation (NRF-2021R1A6A1A03039503 / NRF-2017R1D1A3B06034971) and Soonchunhyang University Research Fund

## REFERENCES

- S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**: 403-410.
- C. Capt, S. Renaut, F. Ghiselli, L. Milani, N. A. Johnson, B. E. Sietman, D. T. Stewart, and S. Breton. (2018) Deciphering the Link between Doubly Uniparental Inheritance of mtDNA and Sex Determination in Bivalves: Clues from Comparative Transcriptomics. *Genome Biology and Evolution*, **10**: 577-590.
- J. M. Chung, H. R. Min, M. K. Sang, J. E. Park, H. C. Cho, S. W. Kang, S. Y. Park, B. B. Patnaik, Y. S. Lee, K. A. Lutaenko, E. H. Shin, J.-S. Lee, Y. S. Han, H. J. Hwang, and Y. S. Lee. (2019) Molecular phylogenetic study of bivalvia from four countries (China, Japan, Russia and Myanmar) using 3 types. *The Korean Journal of Malacology*, **35**: 137-148.
- S. W. Kang, S. Y. Park, H. J. Hwang, J. M. Chung, M. K. Sang, H. R. Min, J. E. Park, H. C. Cho, B. B. Patnaik, and Y. S. Lee. (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.
- S. W. Kang, S. Y. Park, B. B. Patnaik, H. J. Hwang, J. M. Chung, D. K. Song, Y.-S. Park, J. S. Lee, Y. S. Han, H. S. Park, and Y. S. Lee. (2016) The Protostome database (PANM-DB): Version 2.0 release with updated sequences. *The Korean Journal of Malacology*, **32**: 185-188.
- S. W. Kang, S. Y. Park, B. B. Patnaik, H. J. Hwang, C. Kim, S. Kim, J. S. Lee, Y. S. Han, and Y. S. Lee. (2015) Construction of PANM Database (Protostome DB) for rapid annotation of NGS data in Mollusks. *The Korean Journal of Malacology*, **31**: 243-247.
- Y.-S. Lee, Y.-H. Jo, D.-S. Kim, D.-W. Kim, M.-Y. Kim, S.-H. Choi, J.-O. Yon, I.-S. Byun, B.-R. Kang, K.-H. Jeong, and H.-S. Park. (2004) Construction of BLAST Server for Mollusk. *The Korean Journal of Malacology*, **20**: 165-169.
- S. McGinnis, and T. L. Madden. (2004) BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Research* **32**: W20-25.
- M. K. Sang, J. E. Park, D. K. Song, J. Y. Jeong, C.-E. Hong, Y. T. Kim, H. J. Hwang, S. W. Kang, S. Y. Park, J. S. Lee, Y. S. Han, H. S. Park, and Y. S.

PANM Database Version 6.0: An Updated and Expanded Resource for Efficient Annotation of Arthropoda, Nematoda, and Mollusca Genes

Lee. (2021) PANM DB ver 4.0 An update of the bioinformatics database for annotation of large datasets from sequencing of species under Invertebrates. *The Korean Journal of Malacology*, **37**: 33-36.