

Division-4-02

Sophora Genomes Provide Insight into the Evolution of Alkaloid Metabolites Along with Small-scale Gene Duplication

Yang Jae Kang^{1,2}, Halim Park¹, Yejin Lee¹, Sanghwa Yoon¹, Myoungjai Kwak^{3*}

¹Division of Bio & Medical Bigdata Department (BK4 Program), Gyeongsang National University, Jinju 52828, Republic of Korea

²Division of Life Science Department at Gyeongsang National University, Jinju, Republic of Korea

³National Institute of Biological Resources, Incheon 22755, Republic of Korea

[Introduction]

The genus *Sophora* (Fabaceae) is one of the most famous medicinal plants traditionally used in East Asian countries for a long time. Krishna et al. have found that *sophora* plants can be used for sedative, depressant, analgesic, hypothermic, antitumor, antipyretic, and cardiostimulant activities. The genus *Sophora* includes approximately 70 species, distributed widely in tropical and temperate regions. Among them, *S. flavescens* is a perennial herb, native to China, India, Japan, Korea, and Russia, and its dried roots show anti-oxidant, anti-inflammatory, anti-bacterial, apoptosis modulation, and anti-tumor activities. The sister species of *S. flavescens*, *Echinosophora koreensis*, a deciduous shrub, is endemic to Korea, and its ethnopharmacological effectiveness has not been well known, possibly due to its rarity and protection by the law in South Korea. The *E. koreensis* genome, which is phylogenetically related to *S. flavescens* but is more than 50% smaller, can help with the assembly and validation of *S. flavescens*'s much larger genome. Also, the phylogenetic closeness between *S. flavescens* and *E. koreensis* enables genome comparison, which would show recent evolutionary divergence with respect to their polymorphic phenotypes.

[Materials and Methods]

NextDenovo and Falcon are utilized in the process of assembling PacBio long reads into contigs. Using quickMerge, the assembled contigs were combined into a single assemble set. The short reads that were generated by Illumina were utilized to correct the assemblies by nextPolish. Additionally, PurgeHaplotig was used to remove redundant assemblies that may be allelic variants.

[Results and Discussion]

Here, we used the PacBio sequencing platform to sequence and assemble the *E. koreensis* and *S. flavescens* genomes. The *E. koreensis* genome assembly confirmed the expected size of the genome and predicted almost all of the possible genes. The genome and gene catalog were profiled based on various bioinformatics pipelines. The expected genome size was also obtained from the assembly of the *S. flavescens* genome and a gene catalog was prepared. Based on the genome and gene content of the sister species, the evolution of gene families was observed, and especially transporter gene families showed distinct evolution among legume species. Moreover, based on Kegg analysis, we found pathways that may control pharmaceutical metabolites in *S. flavescens* and *E. koreensis*. The *Sophora* species' genome assembly would further be used for comparative genomics and knowledge mining for drug development.

[Acknowledgement]

본 연구는 식물자원의 유전자 다양성 연구(4단계 2차년도)사업 (사업번호: NIBR201905102)의 지원에 의해 이루어진 결과로 이에 감사드립니다.

*Corresponding author: E-mail, mhkwak1@korea.kr Tel, +82-32-590-7102