REVIEW ARTICLE

Soybean [Glycine max (L.) Merrill]: Importance as A Crop and Pedigree Reconstruction of Korean Varieties

Chaeyoung Lee1, Man-Soo Choi2, Hyun-Tae Kim2, Hong-Tai Yun2, Byungwook Lee3, Young-Soo Chung4, Ryan W. Kim3†, Hong-Kyu Choi4†*

1Department of Medical Bioscience, Dong-A University, Busan 604-714, Republic of Korea
2National Institute of Crop Science, Rural Development Administration, Daegu 711-822, Republic of Korea
3Korea Bioinformation Center, Korea Research Institute of Bioscience & Biotechnology, Daejeon 34141, Republic of Korea
4Department of Genetic Engineering, Dong-A University, Busan 604-714, Republic of Korea

ABSTRACT  Soybean [Glycine max (L.) Merrill] is one of the most important crops in the world and in Korea as well. Since the official start of soybean breeding program in Korea at which a landrace ‘Jangdanbaekmok’ was first released to promote cultivation in 1913, approximately one century has elapsed. Currently, a total of 178 soybean varieties are registered at two representative Korean national institutes, the RDA-Genebank Information Center (http://www.genebank.go.kr) and the Korea Seed & Variety Service (http://www.seed.go.kr). Of these, 155 varieties (87.1%) have been developed through hybridization-based breeding technologies, of which most cultivars (133 varieties, 85.8%) have been released in the last twenty five years. In this review, we attempted to integrate all the information for individual cultivars and to rebuild a breeding pedigree including the entirety of registered Korean soybean varieties. The analysis has resulted in a total of four pedigrees involving 168 cultivars (94.4% out of 178 cultivars), which form the broadest network of pedigrees. Each of pedigrees highlights different key varieties within the context of progenitor networks derived from crossing of various elite parental lines as follows; pedigree I-‘Kwangkyo’, ‘Hwangkeumkong’, ‘Paldalkong’ and ‘Sinpaldalkong2’, pedigree II-‘Baegunkong’, ‘Jangyeobkong’ and ‘Keunolkong’, pedigree III-‘Danyeob’, ‘Pangsa’ and ‘Eunhakong’. These pedigrees also reveal purpose (i.e., desirable traits)-driven development of characteristic soybean varieties during the past century of breeding history in Korea. We expect that the pedigree reconstructed in this study will provide breeders with information useful to design breeding schema and guidance towards the genomics-assisted soybean improvement in the future.

Keywords  Soybean, Pedigree analysis, Breeding, Genetic diversity

INTRODUCTION  Soybean [Glycine max (L.) Merrill] is apparently one of the most important cultivated crops worldwide in its agro-economic value and diverse utilities in both agriculture and industry. The legume family, which contains this crop, is composed of approximately 20000 species, which is the third largest group among flowering plants, and includes other agriculturally important legume crops such as common bean (Phaseolus vulgaris), mung bean (Vigna radiata) and pigeon pea (Cajanus cajan). Among many other evolutionary branches within the family, Phaseoloid clade harbors most of important crop legumes of agricultural importance, within which soybean is a member of this clade (Choi and Cook 2011).

It is generally known that distribution of the wild soybean (G. soja), which is the ancestor of current cultivated soybean, is limited to the East Asia regions including China, Korea and Japan. Historical records have addressed that the first cultivation of soybean originated in China, which was approximately 4500 years ago (Qiu and Chang 2010). “Shu”, which is ancient Chinese character meaning soybean, has been frequently found in ancient Chinese books. In addition, carbonized remain of soybean seeds, which was estimated to be 2600 years old, were discovered in an excavation site of the Eastern Zhou
Dynasty (Qiu and Chang 2010). Although its cultivation history dates back to ancient age, soybean actually came to prominent crop during the last 200 years (Singh and Shivakumar 2010). Because of such a long history of cultivation and natural/artificial selection, which was the natural breeding process itself, China had been the world’s top producer until the first half of the 20th century. However, the situation had been reversed by the USA in the 1950s. It is known that soybean was first cultivated in the USA as early as 1765 (Hymowitz 1984). During 1927-1931, the USA sent scientist to collect soybean germplasm in China, Korea and Japan, and some of germplasm played a pivotal roles as primary parents to breed the current USA cultivars (Qiu and Chang 2010). Using those collected accessions, the USA rapidly developed breeding program and accelerated the production of soybean, and thereafter the country has now become not only the highest producer but also largest exporter all across the world. In the USA, soybean is now the second largest crop, right after corn, in production.

Korea also has a significantly long history of domestication and cultivation of soybean, which is comparable to that of China and dates back to the ancient Chulmun period (8000-1500 BC). A recent archeological study shows that charred soybean seeds discovered in ‘Pyeonggeodong’ around the Nam River valley are as old as 4840-4650 years (Lee et al. 2011). Thereafter soybean continued to remain an important crop in Korea throughout the ancient Mumun period followed by the Three Kingdom periods. Interestingly, the study also addresses that soybean domestication occurred independently in multiple regions of the East Asia and landraces with larger seeds were adapted in Korea and Japan far earlier than in China (Lee et al. 2011). Although Korea is one of the East Asian countries of soybean origin and cultivation, start of systematic breeding programs were relatively delayed. The first Korean cultivar bred by hybridization method ‘Kwangkyo’ was recently developed in 1969. Soybean is still important as one of main crops in Korea, which is the third after rice and wheat, and its production in 1975 could almost fill up 98% of domestic need. Since then, the rate of self-sustenance has gradually decreased and is now only 8.7%, thereby causing severe dependence on the GMO soybean imported from the USA.

Soybean is cultivated in enormous area of arable land worldwide, which accounts for 90.2 million ha, resulting in a total production of about 276 million ton (MT) in 2013 (FAO 2013). Among major producing countries, the USA is the top producer (28.2% and 32.2%), by both area and production, followed by Brazil (23.7% and 27.5%), Argentina (18.5% and 21.2%) and China (9.7% and 7.0%), respectively (Soytech Inc. 2007). In comparison to the total production of the world, Korea produced relatively small amount of soybean (0.16MT) in 2013 (FAO 2013).

Presumably, the most prominent purposes for soybean cultivation should be its high contents of protein and oil, which make up approximately 40% and 20%, respectively. In the Western world of North America and Europe, soybean is mainly regarded as oil crop or protein source for animal feed. In other countries of the Eastern World, this crop is consumed mainly by human in various forms of foods, such as bean curd (tofu), soy milk, sprout, soy paste/sauce. Soybean also has a diverse array of utilities in industry and has been used for the production of lubricants, toner ink, cosmetics and for many other purposes. In more recent years, this crop is increasingly drawing interest of scientists as a useful biofuel source.

Although soybean production in Korea has gradually decreased during the past 40 years, its importance as a major crop has not diminished probably because soybean has a long history of cultivation associated intimately with traditional food culture of Korea. Recent completion of soybean whole genome sequencing (Schmutz et al. 2010) may promote the development of new technologies for soybean molecular breeding. Moreover, recently developed technology of the next generation sequencing (NGS) and a variety of bioinformatic tools can play a critical role in processing and analyzing biological big data. NGS-driven bioinformatic data processing enables researchers to analyze population level of resequencing and transcriptome data. Hence, collection and organization of useful germplasm and their phenotypic characterization in precision are becoming more and more important. To effectively select germplasm for the data analyses, it is essential to know relationships among a variety of cultivars, landraces and wild accessions. Towards this end, this review intends to integrate all available data and information, and provide a