We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

4,200
Open access books available

116,000
International authors and editors

125M
Downloads

154
Countries delivered to

TOP 1%
Our authors are among the most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
Wild Soybeans: An Opportunistic Resource for Soybean Improvement

Muhammad Amjad Nawaz, Seung Hwan Yang and Gyuhwa Chung

Abstract

Reduced genetic diversity in cultivated soybean coupled with changing dietary expectations, climate change, and increase in population demands expansion of current gene pool. Wild soybeans are an opportunistic resource and a rational choice to discover novel genes and gene families for alternative crop production systems and to improve soybean. Multiple agronomic traits, lineage-specific genes, and domestication-related traits have been studied in wild soybeans in contrast to cultivated soybeans, and it has been proved that wild soybeans are an essential genomic resource containing unique and useful genetic resources that have been lost during domestication to expand the gene pool in order to improve soybean. Wild soybean is very often a plant of disturbed habitats of Southeast Asia. The vulnerability of these habitats to agriculture systems and urban expansion causes a reduction in the area of distribution and hence the diversity. To capture the wild soybean genetic diversity in its main distribution areas, a unique and comprehensive germplasm collection, characterization, and conservation platform is direly needed. Chung’s Wild Legume Germplasm Collection is preserving and maintaining a representative wild soybean germplasm collection guided by the principles of conservation genetics. These wild legumes and particularly wild soybean is a promising genetic resource for soybean breeders.

Keywords: CWLGC, genetic diversity, genetic resource, Glycine soja, rediscovery of landraces, wild legumes, germplasm conservation

1. Introduction: rediscovery of crop wild relatives

Challenged by limited land and water resources and a concomitant increase in population, changing dietary expectations and climatic change are demanding escalated food supplies...
The nutritional value of grain legumes is far better than cereals even if their production is low, making them a unique and essential component of balanced diet [2]. Grain legumes have suffered a reduction in genetic diversity largely due to plant breeding activities aimed at artificial selection of desirable traits. The new varieties, as well as land races in farmer's field, have desirable characteristics which have become genetically diverged from their ancestors or wild progenitors [3].

In order to cope with the global warming led climatic variations and limited water supplies, there is a constant need of crop improvement; the crop potential has been reduced due to domestication, genetic bottlenecks, and artificial selection [2]. To explore more genes and gene families for alternative production systems, crop wild relatives are a rational choice mainly due to limited or no breeding barriers [4]. The wild progenitors of crops are sometimes easily available, but this is not the case for all species as some of the wild species have gone extinct, or in other cases, multiple progenitors contributed to the genome of the domesticated plants, e.g., wheat. In some cases, some species are indirectly expanding the genomes of the domesticated crops as they may be related species of wild progenitors or wild cousins [5]. Wild crop relatives are mostly adapted to larger climatic variations and are evolved to withstand biotic and abiotic stresses [6]. Therefore, for crop improvement, we have two possibilities, namely, genetic modification or introduction of genetic materials through breeding with crop wild relatives. Of course the use of genetic engineering to create genetically modified plants is relatively quick and efficient but the acceptance of genetically modified plants among the consumers is still controversial. On the other hand, the desirable traits including resistance to biotic and abiotic stresses, nutritional values can be incorporated into the current agricultural crop by using conventional and new breeding technologies. This practice is sometimes quite challenging mainly due to linkage drag; however, recent advances in genetics and genomic approaches have expanded our understanding of evolution, linkage, and heredity of complex traits [4–6] (Table 1).

2. Wild soybean

The Glycine genus comprises of two subgenera, namely, Glycine Willd and Soja (Moench) F. J. Herrmann. Among 28 species classified under two subgenera, only two annual species G. soja Sieb & Zucc (wild) and G. max (cultivated) are consumed as food or feed either directly or indirectly [13]. On the basis of cytological, proteomic, and genomic evidence, the wild species G. soja is considered as the progenitor of cultivated species G. max. Both herbaceous annual species are Asian inborn; mainly distributed in Southeast and Far East Asia including China, Korea, Japan, Taiwan, and Russia [14]. Wild soybean harbors treasured genetic resource and extraordinarily important gene pool; genes and gene families responsible for higher oil and protein contents, resistance to drought and high temperatures, disease resistance, and insect pest resistance [15].

Wild soybeans grow on roadsides, riversides villages, lakeshores, wastelands, and fertile valleys. Apart from numerous phenotypic distinctions among both species, their annual growing habit with similar ploidy level and ability to produce fertile offspring without genetic isolation
<table>
<thead>
<tr>
<th>Crop</th>
<th>Wild relatives</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rice</td>
<td>\textit{O. glaberrima} Steud.</td>
<td>[7]</td>
</tr>
<tr>
<td>\textit{Oryza sativa} L.</td>
<td>\textit{O. nivara} Sharma et Shastry \textit{O. rufipogon} Griff. \textit{O. breviligulata} \textit{O. barthii} \textit{O. meridionalis} \textit{O. glumaepatula}</td>
<td></td>
</tr>
<tr>
<td>Wheat</td>
<td>\textit{T. booticum} \textit{T. monoecium} \textit{T. dicoccoides} \textit{T. araraticum} \textit{T. dicoccum} \textit{T. palaecolchicum} \textit{T. timopheveii} \textit{T. durum} \textit{T. turgium} \textit{T. persicum} \textit{T. carthlicum} \textit{T. polonicum} \textit{T. turanicum} \textit{T. orientale} \textit{T. parvicoocum} \textit{T. spelta} \textit{T. macha} \textit{T. caviolii} \textit{T. aestivum} \textit{T. compactum} \textit{T. sphaerococcum}</td>
<td>[8]</td>
</tr>
<tr>
<td>\textit{Triticum aestivum} L.</td>
<td>\textit{T. compactum} \textit{T. aestivum} \textit{T. carthlicum}</td>
<td></td>
</tr>
<tr>
<td>Corn</td>
<td>\textit{Z. luxurians} \textit{Z. diploperennis} \textit{Z. Mexicana} \textit{Z. nicaraguensis} \textit{Z. perennis}</td>
<td>[9]</td>
</tr>
<tr>
<td>\textit{Zea mays} L.</td>
<td>\textit{G. soja} Sieb. &amp; Zucc. \textit{G. soja}</td>
<td>see Table 2 for wild cousins of soybean</td>
</tr>
<tr>
<td>Soybean</td>
<td>\textit{G. soja} \textit{G. soja} \textit{G. soja} \textit{G. soja}</td>
<td>[10]</td>
</tr>
<tr>
<td>\textit{Glycine max} L.</td>
<td>See Table 2 for wild cousins of soybean</td>
<td></td>
</tr>
<tr>
<td>Common bean</td>
<td>\textit{P. coccinus} \textit{P. dumosus} \textit{P. maculatifolius} \textit{P. costaricensis} \textit{P. acutifolius} \textit{P. lunatus} \textit{P. parsiolus} \textit{P. costaricensis} \textit{P. albescens} \textit{P. filiformis}</td>
<td>[11]</td>
</tr>
<tr>
<td>\textit{Phaseolus vulgaris} L.</td>
<td>\textit{P. coccinus} \textit{P. dumosus} \textit{P. maculatifolius} \textit{P. costaricensis} \textit{P. acutifolius} \textit{P. lunatus} \textit{P. parsiolus} \textit{P. costaricensis} \textit{P. albescens} \textit{P. filiformis}</td>
<td></td>
</tr>
</tbody>
</table>
results in a flow of certain characteristics from wild to cultivated populations [16]. Wild soybeans exhibit distinct geographical patterns as well as interspecific horizontal mechanisms of flow of genetic information to cultivated species mainly because of sharing the same gene pool and close proximity [10] (Table 2).

Many parts of China and South Korea which were previously regarded as habitat for wild soybean are now being used for agricultural, commercial purposes (roads, buildings or dams) or are now part of the sea. Destruction of natural soybean habitats due to land clearance for agricultural or industrial purposes has led to decreased wild germplasm resources [19]. Furthermore, reduction in genetic diversity has been witnessed due to the domestication of soybean during past three decades. Progenitor wild species exhibit discrete geographic patterns with greater genetic diversity but the selection and allele frequency changes during domestication has curtailed genetic variability. Various studies have reported a reduction of genetic diversity up to 50% in domesticated/improved cultivars as compared to wild progenitors [3, 15]. Artificial selection and domestication mainly focused dominant selection of desirable traits such as oil content, seed size, and seed coat luster, imposing selection pressure on particular traits and ignoring other important traits. Another factor involved in diminishing genetic diversity is habitat fragmentation [20] (Figure 1).

### 2.1. Wild soybean genome: rediscovering the lost diversity

Whole genome sequencing of wild soybean genome started a new era for soybean functional and comparative genomics and has substantially increased our understanding about soybean domestication history, bottlenecks, lost diversity and has created a way forward towards its potential use in expanding the gene pool of soybean. The wild and cultivated soybeans have significant and useful genomic differences which highlight the phenotypic differences and as well as the domestication-related traits. Kim et al. [21] aligned 915.4 Mb genomic sequence of wild soybean with soybean reference genome excluding the gaps and found that wild soybean genome covered 97.65% of soybean genome with a difference of 35.2 Mb (3.76% of 937.5 Mb). The difference region consisted of 0.267% substitution bases, 0.043% insertion/deletions (indels), and 3.45% of large deleted sequences. Single nucleotide polymorphisms (SNPs) and

<table>
<thead>
<tr>
<th>Crop</th>
<th>Wild relatives</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chickpea</td>
<td>C. reticulatum</td>
<td>[12]</td>
</tr>
<tr>
<td>Cicer arietinum L.</td>
<td>C. echinospermum</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C. pinnatifidum</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C. judaicum</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C. bijugum</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C. cuneatum</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C. chorassanicum</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C. yamashitae</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. List of wild relatives of three major cereals and three major legumes.
insertions/deletions (indels) in precisely aligned areas differed by 0.31% between cultivated and wild soybean. The complex genome rearrangement is mainly caused by indels, inversions, and translocations (up to thousands of base pairs); along with SNPs and indels [22]. The wild soybean genome has greater allelic diversity than that of soybean. Resequencing of 17 wild and 14 cultivated soybean genomes to an average of ×5 depth and >90% coverage identified higher allelic diversity [15].

Based on whole-genome SNP analysis using the parameter $\theta_\pi$, a higher level of genetic diversity was found in wild soybeans ($2.97 \times 10^{-3}$) as compared to cultivated soybeans ($1.89 \times 10^{-3}$).

Table 2. Members of genus *Glycine*.
Similar findings were also reported when 302 wild and cultivated soybeans were whole-genome sequenced to an average depth of > x11; the genetic diversity ($\pi$) decreased from $2.94 \times 10^{-3}$ in G. soja to $1.40 \times 10^{-3}$ in landraces and to $1.05 \times 10^{-3}$ in improved cultivars suggesting that nearly half of the annotated resistance-related sequences were lost during the domestication [3]. Noticeably, total number of SNPs in wild (5,924,662) and cultivated soybeans (4,127,942) was comparable (35 and 5%, respectively) and the ratio of nonsynonymous SNP to synonymous SNPs was higher in cultivated (1.38) than wild soybeans (1.36). It was observed that the number of fixed loci were lower in the wild (463,409) as compared to cultivated soybeans (2,148,585). These findings suggest that there should be low-frequency alleles in domesticated soybean owing to domestication bottleneck. However, contrary to this, low-frequency alleles were abundant in wild soybean suggesting that the wild soybean habitat has reduced and the cultivated soybean population has expanded [15].

2.2. Expanding the gene pool for soybean improvement

Wild soybeans are a potential genetic resource for the improvement of cultivated soybean and aid greatly in exploring alternative production systems. Wild soybeans, as in case of another wild relative of cultivated crop species, contain higher genetic diversity as they had a long time opportunity to evolve and withstand under varied environmental conditions without inference by humans [4, 15]. Wild soybeans are interfertile with cultivated soybeans and represent an easily accessible or primary gene pool for soybean improvement [10]. However, the global climate change and increase in human population have developed a scenario of securing, conserving, characterizing, and using wild soybeans as a resource for soybean improvement. Loss of genetic diversity during the process of soybean domestication and presence of a domestication bottleneck, i.e., domestication syndrome has led towards changes in growth habits, loss of germination inhibition and mechanisms of seed dispersal [4]. This domestication has also enabled the crop plants to withstand and adapt to modern agriculture and farming system, which is very encouraging. However, loss of diversity in cultivated soybeans calls for revisiting natural diversity reservoirs, i.e., wild soybeans in search of potential genes/alleles.
for higher yield. Multiples sequences that are unique to wild soybeans have been discovered but a report from Korea by Chung et al. [23] also witnessed gene loss events in wild soybeans. However, this discrepancy might be due to diversity of wild collections [4]. Multiple agronomic traits, lineage-specific genes, and domestication-related traits have been studied in wild soybeans in contrast to cultivated soybeans, and it has been proved that wild soybeans are an essential genomic resource containing unique and useful genetic resources that have been lost during domestication to expand the gene pool in order to improve soybean [3, 15, 24, 25]. One recent example is the salt-resistant gene GmCH1X identified in wild soybean. The salt-resistant gene originally did not have a Ty1/copia retrotransposon insertion into its exon 3 in wild soybean and controls 80% salt tolerance in wild soybean (W05) as compared to its counterpart C08 (cultivated soybean) which had retrotransposon insertion possibly due to recent round of whole genome duplication [24], strongly implying that wild soybeans’ genetic diversity must be explored.

2.3. Traits from the wild

Recent development in high-throughput sequencing technologies is clearly promoting a revolution in the comparative genomic sequencing of major crops. A rapid growth in the number of sequenced genomes of crops and their wild relatives has established that wild species tend to have higher genetic diversities, making the wild relatives promising natural resources of novel genes/alleles for crop improvement. Many studies have provided the details on wild soybean specific genes/alleles controlling major abiotic and biotic stress tolerance-related traits. Contrastingly, cultivated soybeans also have unique genes/alleles which have been possibly lost during the evolution of wild soybeans. However, the results of each comparative study must be based on the genetic diversity present within the subject population [4, 21, 26].

2.3.1. Domestication-related traits

Identification of genes for domestication-related traits is an important task to maintain diversity in crops for improvement. Such a knowledge provides essential understanding of how and what genetic signatures have brought necessary changes in plant phenotype and physiology during the process of domestication. In soybean, the domestication-related traits are the increased size of inflorescence, grain yield, seed size, seed color, hilum color, pubescence form, apical dominance, stem determinacy, and plant height. Many of the domestication QTLs have been identified, such as twining habit (Ch. 02 and 18), hard seededness (Ch. 02 and 06), determinate habit (Ch. 17), maximum internode length (Ch. 06, 18 and 19), flowering time (Ch. 06 and 16), pod dehiscence (Ch. 16), seed weight (Ch. 17), stem determinacy (Ch. 17), oil content (Ch. 03, 11, 12, 13, 15, 17), flower color (Ch. 13), seed coat color (Ch. 08), pubescence form (Ch. 01, 12, 18, 19, 20), and plant height (Ch. 18) [3, 15].

2.3.2. Other traits

Many useful QTLs/genes have been obtained by characterizing wild soybeans or using genetic populations resulting from crosses between wild and cultivated soybeans. These
genes/QTLs have been characterized to understand the stress resistance mechanisms and various biochemical pathways related to plant development, yield, and local breeding traits. (1) Multiple genes/alleles responsible for flower color, i.e., pinkish-white and white flowers have been identified from wild soybeans. Flavonoid 3′5′-hydroxylase (F3′5′H) and dihydroflavonol-4-reductase (DFR) are responsible for anthocyanin production. Different loci control the anthocyanin content and decide the fate of flower color. Different loci, i.e., W1, W3, W4, w1-s1, w1-s2, w1-ip, and w1-p2 have been reported to control white color and pinkish-white [27, 28]. (2) Other studies on seed antioxidant, phenolics, and flavonoid contents have identified GmMATE1, 2, 4 genes [29]. Astringent taste in soy products is caused by group A saponins. Glyma15g39090 has been successfully characterized as sg-5 gene in natural wild soybean mutant line CWS5095. The gene oxygenates the C-21 position of soyasapogenol B or other intermediate which results in the production of saponin A [30, 31]. Another gene Sg-1 of this pathway has also been characterized from Korean wild soybean natural mutant which controls Ab series of saponins [32]. (3) Soybean cyst nematode is a global threat to soybean and host plant resistance is an ideal way of managing the damages. Wild soybean was used to discover SNPs and candidate genes significantly associated with soybean cyst nematode resistance, i.e., 10 SNPs and genes related to disease resistance-related proteins with leucine-rich region. Two genes, namely, a mitogen-activated protein kinase gene (Ch. 18) and a MYB transcription factor (Ch. 19) were found to be strong candidates [33]. (4) Many genes and transcription factors have also been identified from wild soybeans which play important role in drought stress tolerance [34]. (5) A salt-tolerant gene, i.e., GmCHX1 has been identified from wild soybean through whole-genome de novo sequencing approach. (6) A phosphatase 2C-1 (PP2C-1) allele has been reported from wild soybean to be involved in seed weight and seed size. Apart from these genes many QTLs related to (7) linolenic acid production, (8) yield, height and maturity, (9) soybean cyst nematode, (10) seed yield, seed weight, seed filling period, maturity, plant height, and lodging, (11) salt tolerance, (12) sclerotinia stem rot, (13) root traits, (14) oil and local breeding, (15) shoot fresh weight, (16) seed antioxidant, phenolics and flavonoids, and many other traits have been mapped and identified from wild soybeans or populations developed by crossing wild and cultivated soybeans [24, 29, 34–41]. Taking the advantage of higher genetic diversity and identified genes and QTLs for important traits from wild soybean will gear up the soybean yield improvement in changing climatic conditions and modern dietary demands. Joint ventures guided by the principles of plant breeding, genetics, genomics, and modern biotechnology are underway in many parts of the world to improve soybean in terms of resistance to biotic and abiotic stresses, adaptation to low water and higher temperature conditions, as well as intensive agricultural systems.

3. Wild soybean germplasm conservation

Wild soybean (the presumed progenitor of soybean) is very often a plant of disturbed habitats of Southeast Asia. Such habitats are mostly on the roadsides, intensive agricultural lands, and areas with higher human disturbances in terms of land use. The adaptation to these disturbed areas actually predisposes the wild soybeans to agricultural systems; this is one of the reasons for its domestication in East Asia [2, 19]. The vulnerability of these habitats to agriculture
systems and urban expansion causes reduction in area of distribution and hence the diversity.

As discussed earlier, wild soybean is an efficient resource for identification and characterization of furnished and important genes for soybean improvement [10]. Economically, gene-banks and wild genetic resources have been unambiguously reported to have led towards higher economic return by increasing soybean productivity [5]. Wild soybean germplasm preservation is underway in many countries mainly China, Korea, Japan, and the United States of America [10, 19]. Surely, the collections are growing by following the principles of conservation genetics; however, the complete representative collections are yet to be achieved as there remain many unexplored uninhabited natural habitats of wild soybeans which might carry many useful genes, alleles, or mutations. The undiscovered variations are greatly in demand by plant breeders to increase soybean production for an ever-increasing population [18]. Wild soybean germplasm should be collected mainly to (a) to understand the taxonomy and phylogenetic relationships, (b) to understand the biogeometrics of certain yield-related pathways, (c) characterize and conserve germplasm, and (d) make it available for soybean breeders across the globe [10, 18–19]. Currently, there are many gene banks which are working on wild soybean germplasm conservation. In Southeast Asia, China holds the largest wild soybean germplasm collection of 6172 accessions under Chinese Crop Germplasm Information system [42], followed by Chung’s Wild Legume Germplasm Collection (CWLGC) holding 6012 accessions and National Institute of Agrobiological Sciences Genebank in Japan which holds 1131 accessions [43]. Outside East Asia, the largest collection of wild soybean germplasm is USDA Soybean Germplasm Collection holding nearly 21,810 accessions belonging to 21 species of genus Glycine [44]. Almost 1179 accessions belong to 20 wild relatives of cultivated soybean including the wild soybean. N. I. Vavilov Institute of Plant Genetic Resources (VIR) holds ~350 accessions [45]. All of these germplasm collections are either focused on cultivated soybean or limited to a particular country/region. There is a dire need of germplasm center particularly focused on the collection, characterization, and dissemination of wild soybean accessions from the main distribution area of the species i.e., Southeast Asia. Out of the above-mentioned germplasm collections, CWLGC is primarily focused on wild soybean germplasm collections from China, Korea, Japan, and Far East Russia near Chinese border.

3.1. Chung’s wild legume germplasm collection

Chung’s wild legume germplasm collection strives to develop a comprehensive conservation program resourcefully and efficiently to conserve and promote the genetic diversity within wild legumes with main focus on wild legumes. Guided with the principles of conservation genetics, CWLGC focuses on (a) direct collection, (b) acquirement, (c) conservation, (d) evaluation and characterization, and (e) documenting and distribution of wild soybean germplasm. CWLGC was established in 1983 by Professor Gyuhwa Chung at Department of Biotechnology, Chonnam National University, Yeosu campus, Republic of Korea. CWLGC holds 10,314 different legume genera. However, particular emphasis is on the germplasm collection, multiplication, evaluation, and utilization of *G. soja*, *Amphicarpaea edgeworthii*, *Vigna vexillata*, *Rhychnosia volubilis*, and *Phaseolus nipponensis*. The CWLGC efforts incredibly focused on wild soybeans, native to main centers of diversity, particularly East Asia. Wild soybean seeds are the main collection at CWLGC, and the seeds are considered as the currency of germplasm which is safeguarded for global food availability and security as well as
to preserve natural genetic diversity of wild soybeans and other legumes. This diverse collection of pinhead to large-sized seeds is important than ever as plant breeders, geneticists, conservationists, and biotechnologists need to cope with the changing climate, reduction in arable land area, occurrence of natural disasters, environmental degradation, and rising expectations in nutritional standards (Figures 2 and 3).

**Conflict of interest**

The authors declare that there exists no conflict of interest.
Author details

Muhammad Amjad Nawaz, Seung Hwan Yang and Gyuhwa Chung*

*Address all correspondence to: chung@chonnam.ac.kr

Department of Biotechnology, Chonnam National University, Yeosu, Republic of Korea

References


[26] Stupar RM. Into the wild: The soybean genome meets its undomesticated relative. PNAS. 2010;107:21947-21948. DOI: 10.1073/pnas.1016809108


[34] Li D, Pfeiffer TW, Cornelius PL. Soybean QTL for yield and yield components associated with Glycine soja alleles. Crop Science. 2007;48:571-581. DOI: 10.2135/cropsci2007.06.0361


[38] Zhang H, Li C, Davis EL, Wang J, Griffin JD, Kofsky J, Song BH. Genome-wide association study of resistance to soybean cyst nematode (Heterodera glycines) HG Type 2.5.7 in wild soybean (Glycine soja). Frontiers in Plant Science. 2016;17:1214. DOI: 10.3389/fpls.2016.01214


