Compendium of Plant Genomes
*Series Editor:* Chittaranjan Kole

Henry T. Nguyen
Madan Kumar Bhattacharyya *Editors*

# The Soybean Genome

Springer

# Compendium of Plant Genomes

**Series editor**

Chittaranjan Kole
Nadia, West Bengal
India

More information about this series at http://www.springer.com/series/11805

Henry T. Nguyen
Madan Kumar Bhattacharyya
Editors

# The Soybean Genome

Springer

*Editors*
Henry T. Nguyen
Division of Plant Sciences and National Center for Soybean Biotechnology
University of Missouri
Columbia, MO
USA

Madan Kumar Bhattacharyya
Department of Agronomy
Iowa State University
Ames, IA
USA

ISSN 2199-4781 ISSN 2199-479X (electronic)
Compendium of Plant Genomes
ISBN 978-3-319-64196-6 ISBN 978-3-319-64198-0 (eBook)
DOI 10.1007/978-3-319-64198-0

Library of Congress Control Number: 2017947456



Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

*This book series is dedicated to*
*my wife Phullara, and our children Sourav, Carena,*
*and Devleena*

Chittaranjan Kole

# Preface to the Volume

Soybean [*Glycine max* (L.) Merr.] is the most important legume crop, which is a great source of both protein and oil. Soybean seeds contain approximately 40% protein and 20% oil. Soybean is an important source of protein in animal and fish feed in addition to human nutrition. In recent years, soybean is also becoming a source of biodiesel. Soybean root fixes nitrogen through symbiosis with a rhizobacterium, *Bradyrhizobia japonicum,* improving soil health. More than 30% of the world's soybean crop is produced in the USA and is valued at around $40 billion annually. Brazil and Argentina are other two major soybean-growing countries, followed by China and India.

Soybean was originated in East Asia. It was first domesticated over 8000 years ago in China, 5000 years ago in Japan, and 3000 years ago in Korea. It was introduced to Asian countries, such as India, Indonesia, the Philippines, Vietnam, Thailand, Cambodia, Malaysia, Burma, and Nepal between 1 AD and 1600 AD. To the major soybean-growing countries, soybean was introduced only in the recent years viz to the USA in 1765, Argentina in 1882, and Brazil in early 1950s. Over the last few years, there is a growing interest in expanding soybean cultivation in Africa. The soybean has a very narrow genetic base; most cultivars can be traced to the same handful progenitor lines. In recent years, wild species are utilized through hybridization to broaden the genetic base of modern soybean cultivars.

Considering economic importance and narrow genetic base of soybean, molecular genetics and genomics approaches are becoming vital to ensure steady increases in yield potential to meet the food and nutritional demands of over 9 billion people by 2050. Fortunately, with the advent of second- and now third-generation sequencing platforms, we are able to identify and use the genetic potentials of available germplasm in designing new soybean cultivars that are expected to meet the everincreasing nutritional demands of billions of people under the changing growing conditions, anticipated from climate change. The objective of this book is to bring attention of the readers including students to the recent advances in soybean genetics, breeding, and genomics along with resources essential for highly needed genetic improvement in soybean.

The book comprises 13 chapters with Chaps. 1 and 2 describing the economic importance and botanical aspects of soybean, respectively; with the last chapter (Chap. 13) providing the description and navigation of the

SoyBase; the toolbox for molecular markers, genetic and physical maps, mutants, metabolic pathway, gene expression, a Gbrowse for the soybean reference genome, literature searches to facilitate soybean genomic, genetic, and breeding research. The next six chapters (Chaps. 3–8) provide in-depth reviews on molecular markers, molecular maps, structural and comparative genomics, genome-wide association, and application of molecular resources in breeding soybean. Molecular markers and molecular maps essential for developing predictive selection programs for complex traits such as seed, protein, and oil yield are described in Chap. 3. The chapter reviews classical markers, then RFLP, RAPD, AFLP, SSR, and then SNP molecular markers and maps. The chapter also reviews the use of second-generation sequencing in generating tens of thousands of SNP markers for developing SNP maps from crosses involving wild species and cultivars.

Chapter 4 describes detection, description, and development of structural variation in the soybean genome and how to incorporate these polymorphisms in ongoing soybean research and genetic improvement. Genome assemblies and structural variations among a large collection of soybean genotypes obtained from resequencing are presented in Chap. 5. The Chap. 6 reviews comparative genomics in soybean and that of cultivars with accessions of a wild relative for identifying genes involved in defense response, cell growth, and photosynthesis.

In the USA, soybean germplasm collection is composed of nearly 22,000 accessions including 19,648 modern and landrace cultivars (*G. max*), 1168 wild relatives of soybean (*G. soja*), and 1184 perennial wild species. Most lines of the collection have been genotyped for 52,041 SNP loci making it feasible to conduct genome-wide association studies (GWAS). Chapter 7 reviews GWAS conducted for identification of candidate genes for various agronomical traits, seed composition, seed weight, nitrogen traits, photochemical reflectance index, resistance to soybean cyst nematode, and brown stem rot. The Chap. 8 describes progresses made in implementing tightly linked molecular markers in breeding and introgressing quantitative trait loci/genes that control important agronomic traits.

The next four chapters (Chaps. 9–12) review the recent advances in functional genomics of soybean through analyses of mutants created by the chemical mutagen ethyl methanesulfonate, gene silencing, or transposon-induced mutation. The Chap. 9 reviews application of targeted induced local lesions in genomes (TILLING) to identify mutations within soybean genes of interest. TILLING facilitates conducting both forward and reverse genetics in plant species, and this chapter reviews what has been accomplished in soybean with an example of functional characterization of a soybean cyst nematode resistance gene *Rhg4* encoding serine hydroxymethyltransferase (SHMT) involved in one carbon metabolism. The chapter also documents—the identification of suitable mutants through TILLING for improving quality of oil in soybean and the approach can be applicable also to any traits of interest. Chapter 10 describes the recent advances in virus-induced gene silencing, gene silencing through RNAi in stable transgenic plants, and gene editing systems in soybean.

In Chap. 11, landscape of the transposable elements (TEs) including retrotransposons and type II or DNA transposons in soybean is described. As in other crop species, retrotransposons with long terminal repeat retrotransposons are a major component of the soybean genome that are preferentially accumulated in the pericentromeric regions of all chromosomes and are inactive; but they can be activated under certain stressful conditions. In Chap. 12, application of heterologous transposon systems and an endogenous type II transposon element, *Tgm9*, in tagging and functional characterization of soybean genes is described.

The 13 chapters included in this book have been prepared by experts. We greatly appreciate their contributions. We expect that this book will be a useful reference—for graduate students as well soybean researchers and also researchers of other crop species.

We are grateful to all our colleagues for their contribution. We wish to record our thanks and appreciations for Prof. Chittaranjan Kole, the Series Editor, for his assistance and guidance right from the inception till publication of this book.

Columbia, USA Henry T. Nguyen
Ames, USA Madan Kumar Bhattacharyya

# Contents

# The Economic Evolution of the Soybean Industry

1

Chad Hart

**Abstract**

Since the crop's humble beginnings in China, the global soybean industry has grown to be a multi-billion dollar enterprise. Since World War II, the USA has been the dominant market for both production and consumption. For the 2015 growing season, US farmers planted over 80 million acres and produced nearly 4 billion bushels of soybeans. With market prices averaging $10 per bushel, that translates to $40 billion in raw production value each harvest. However, the soybean plant was not always one of the dominant crops in the US agricultural landscape. Large-scale soybean production in the USA is a relatively recent phenomenon. Roughly 55% of the world's soybean crop is directed to feed use currently. Over the course of the past 20 years, another use for soybeans has emerged in the energy sector, biofuel, specifically biodiesel production. Soybean production costs have changed dramatically over the past 20 years. In general, crop production costs can be broken down into four major categories: machinery, land, labor, and crop inputs (seed, chemicals, and fertilizer). Land costs make up nearly half of the total cost of soybean production. Soybean, like many row crops, is a bulk commodity. This means that soybeans produced by one farmer/practice/variety are not differentially marketed or priced from soybeans produced by others (with some rare exceptions for food-grade non-genetically modified varieties or organic practices). Or to put it another way, a soybean is a soybean no matter who produces it or how it is produced. So soybean producers are transacting in a competitive market, where producers can enter and exit the market fairly easily and there is very little room for differentiation. But there is a similar situation for the entities that purchase soybeans as well. So the soybean market is made up of many sellers (producers) and buyers (country elevators, crushing facilities, river terminals, and exporters). In competitive

C. Hart (✉)
Iowa State University, 478F Heady Hall, 518 Farm
House Lane, Ames, IA 50011, USA
e-mail: chart@iastate.edu





H.T. Nguyen and M.K. Bhattacharyya (eds.), *The Soybean Genome*,
Compendium of Plant Genomes, DOI 10.1007/978-3-319-64198-0_1

markets like this, prices are generally equal to production costs. If prices exceed costs, the resulting profits will inspire additional production from existing and new producers, leading to larger supplies and lower prices. If costs exceed prices, the resulting losses will drive some producers out of the business, leading to supply reduction and price improvement.

## 1.1 Introduction

Since the crop's humble beginnings in China, the global soybean industry has grown to be a multi-billion dollar enterprise. Since World War II, the USA has been the dominant market for both production and consumption. For the 2015 growing season, US farmers planted over 80 million acres and produced nearly 4 billion bushels of soybeans. With market prices averaging $10 per bushel, that translates to $40 billion in raw production value each harvest. However, the soybean plant was not always one of the dominant crops in the US agricultural landscape. Large-scale soybean production in the USA is a relatively recent phenomenon. Like many things in US history, soybeans are not native, but imported and adopted by US producers over time.

The soybean plant was domesticated roughly 3000 years ago in China. And for the first 2900 years, it remained mainly a Chinese crop. Soybeans were first brought to the USA just prior to the Revolutionary War. However, significant production of soybeans in the USA did not begin until the mid-1900s. The change in production coincides with a change in use for the crop. Prior to World War II, soybeans were mainly seen as a forage crop, providing feed for grazing animals. One of the side effects of the war was the significant disruption of agricultural trade throughout the world. At the start of World War II, the USA imported nearly half of the edible fats and oils it used. The war severely curtailed that trade, creating pressure to develop domestic sources for edible fats and oils. Soybean production grew to fill the void. Similar pressure led to the development of canola in Canada (Gibson and Benson 2005).

But the need for edible fats and oils is not the only feature of soybean that helped the crop become the 2nd largest crop in the USA, and the soybean plant has several other aspects that made it attractive to US farmers over the past 75 years. Crushing soybeans for oil also provides meal, which has become a staple of livestock rations (continuing the connection between soybean and livestock, but moving the relationship from forage to feed). The timing and production practices for soybean are similar to that of corn, the largest crop in the US, and farmers developed a rotation with the two crops. That rotation benefits corn, as soybean's legume properties maintain nitrogen levels in the soil.

The United States Department of Agriculture (USDA) first began tracking soybean acreage and production in 1924. At that time, roughly 1.5 million acres were planted to soybeans. And since much of the crop was used as forage, less than 30% of the crop was harvested, resulting in production of approximately 5 million bushels. Early yield figures were in the 11–13 bushel per acre range. Acreage and production slowly grew through the 1930s. As noted earlier, the first major shift for soybeans occurred during World War II. Soybean plantings topped 10 million acres for the first time, and crop usage shifted from forage to harvested production. In 1940, just under 46% of the US soybean crop was harvested. By 1945, over 82% of the crop was harvested. The need for edible fats and oils changed the dynamics for soybeans. Since then, forage usage of soybean has become a minimal activity. As Fig. 1.1 shows, US farmers have continued to devote more acreage to soybeans, exceeding 50 million acres in the early 1970s and expanding to over 80 million acres now.

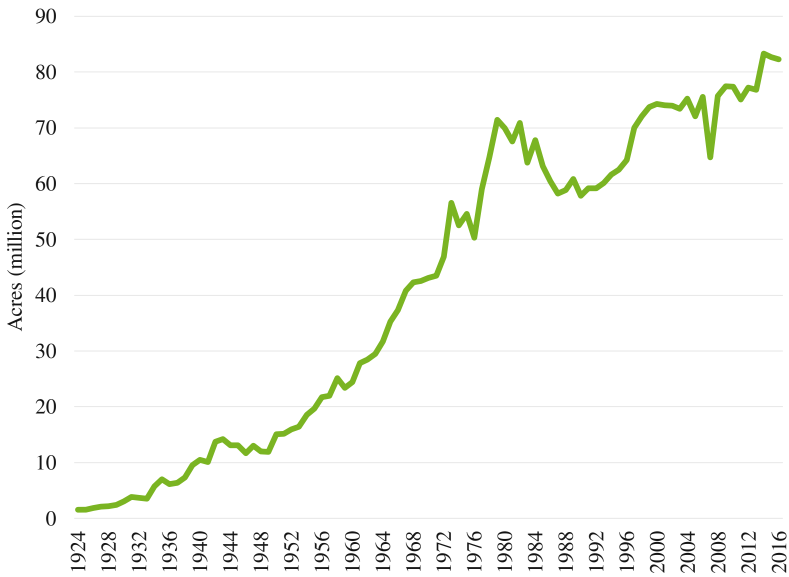


**Fig. 1.1** US soybean planted area (*Source* USDA-NASS 2016a, b)

As soybeans were first introduced in the southeast USA, most of the early production originated there. However, over time, soybean production has shifted to the upper Midwest. By the time USDA started tracking soybeans, Illinois was already firmly established as a leading production state. But a majority of the top 10 soybean-producing states in the 1920s were in the southeast (Tennessee, North Carolina, Alabama, Virginia, Mississippi, and Georgia). Iowa did not enter the top 10 until 1930. Gradually, soybean acreage moved northwest and that shift continues today. While Illinois and Iowa compete for the top spot, Minnesota and North Dakota have emerged as strong soybean production states and the further state southeast in the soybean top 10 states is Missouri.

As Fig. 1.2 shows, soybean yields have steady increased over the past 90 years. USDA first estimated national soybean yields in 1924. At that time, the national average yield was 11 bushels per acre. Since then, the national average soybean yield has had a rough annual growth rate of 0.35 bushels per year. The growth in yield, combined with the increase in planted area, resulted in tremendous growth in soybean production. In 2015, the national soybean crop set a record yield of 48 bushels per acre, with annual production approaching 4 billion bushels.

In terms of global production, the USA has been the dominant producer of soybeans for quite some time. Consistent official global agricultural statistics were first captured in the 1960s. By that time, the USA was already the top soybean-producing country in the world, harvesting over half of the global crop. Commercial soybean production did not really begin in South America until the late 1960s. But since then, South American production has ramped up significantly. While the USA is currently still the top producing country, Brazil will likely surge past in the next 10 years. Argentina is the 3rd largest source of soybeans. And Bolivia, Paraguay, and Uruguay are also strong soybean-producing countries. And China is the 4th largest producer, but their production is well behind their current usage.

As Fig. 1.3 shows, the global production of soybean is very concentrated among the USA, Brazil, and Argentina. However, consumption (shown in Fig. 1.4) is more distributed throughout the globe. The USA was the largest consumer of soybeans until 2007. Since then, China has been the dominant buyer in the global soybean market. China consumes roughly 30% of global production and is the largest soybean buyer for the USA, Brazil, and Argentina. So the global soybean market is driven by three major

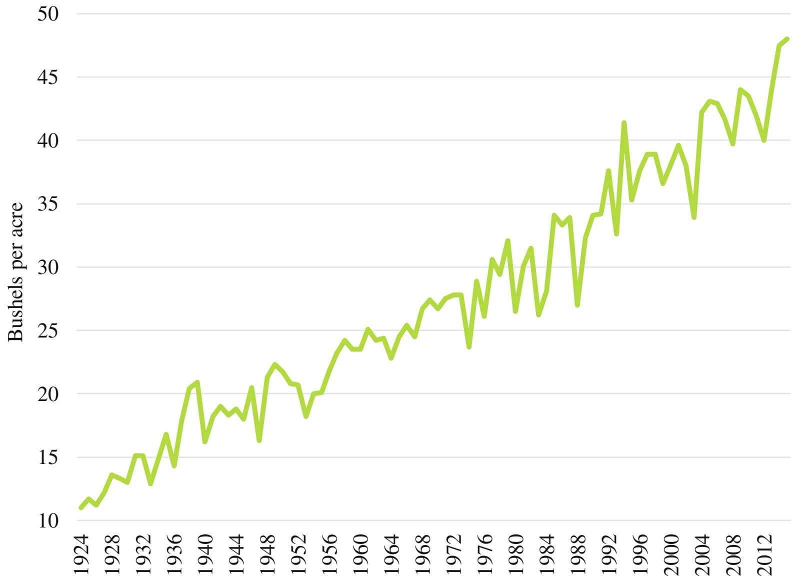


**Fig. 1.2** US soybean yields (*Source* USDA-NASS 2016a, b)

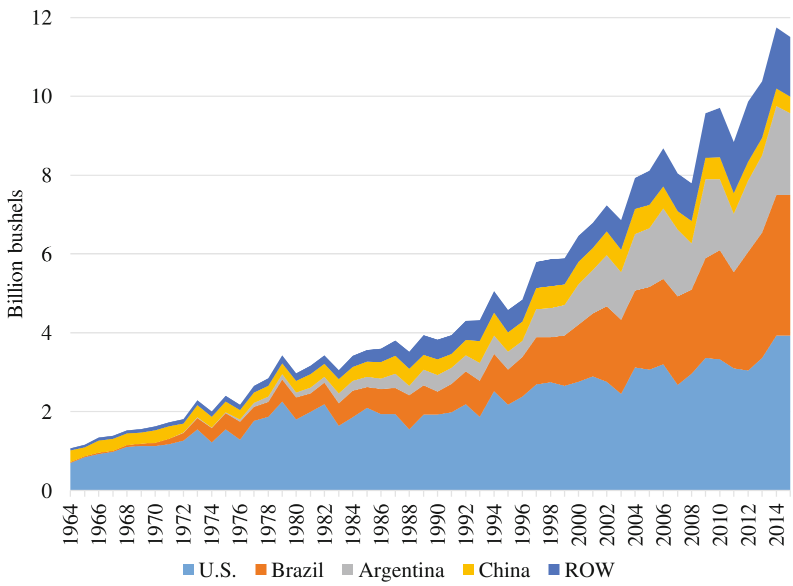


**Fig. 1.3** World soybean production (*Source* USDA-FAS 2016)

producers (USA, Brazil, and Argentina) and one major consumer (China).

From the 1960s to the late 1980s, food demand for soybeans led the market. However, as global demand for meat has risen, feed demand for soybeans has now become the leading use. Roughly 55% of the world's soybean crop is currently directed to feed use. So soybean usage has circled back to its original use as livestock feed. Over the course of the past 20

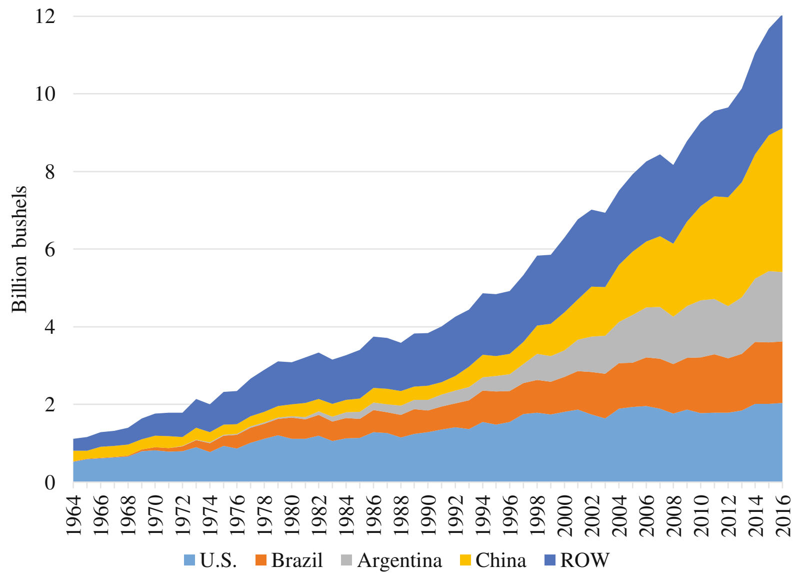


**Fig. 1.4** Global soybean consumption (*Source* USDA-FAS 2016)

years, another use for soybeans has emerged in the energy sector, biofuel, specifically biodiesel, production. Biodiesel can be produced from a variety of feedstocks: animal fats, palm oil, rapeseed or canola oil, etc. In the USA, biodiesel production is largely dependent on soybean oil as the feedstock. Currently, roughly 25% of the soybean oil used in the USA is converted to biodiesel.

Over the course of 100 years, soybean has evolved from a minor forage crop to the 2nd largest row crop in the USA. And as demand for soybeans continues to grow, it may, 1 day, challenge corn as the most planted crop.

## 1.2 Production Economics

Soybean production costs have changed dramatically over the past 20 years. The growing popularity of the crop and the competition for farmland among alternative crops have led to significant cost increases. In the early 1970s, Iowa soybean producers could raise an acre of soybeans for less than $100. Given current conditions, soybean production costs exceed $500 per acre. In general, crop production costs can be broken down into four major categories: machinery, land, labor, and crop inputs (viz. seed, chemicals, and fertilizer). Land costs make up nearly half of the total cost of soybean production. Many farming operations rent land for row crop production, representing a direct land cost for the crop. On owned land, the rental rate represents the opportunity cost foregone by not renting the land to another farmer. As a number of crops could be grown on the land, land costs per crop are consistent across crops (i.e., corn land costs are equal to soybean land costs). Crop input costs are the next largest category, making up roughly 30% of production costs. Seed costs have risen over time due to increased plant populations on soybean fields (farmers are planting more soybean seeds per acre) and new innovations in seed technology (via plant breeding and genetic modification). Fertilizer and chemical costs have varied significantly as prices and usage shift. Machinery costs have grown more slowly, but still account for roughly 15–20% of costs. Labor costs have been the most consistent, but represent less than 10% of costs.

Production costs have nearly doubled over the past 10 years. The largest increase occurred in land costs as farm incomes rose to record levels,

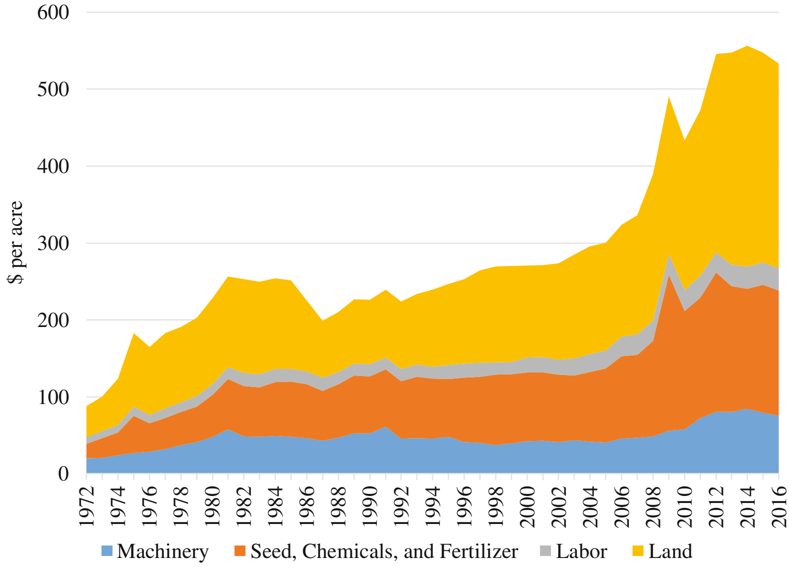


**Fig. 1.5** Soybean production costs (*Source* Johanns and Plastina 2016)

driving strong competition among farmers for rented land. A typical Iowa land rent in 2006 was $145 per acre per year. By 2014, rents had risen to $287 per acre per year. Over the same time period, crop input costs rose from $107 per acre to $156 per acre. As Fig. 1.5 shows, there was a spike in crop input costs for the 2009 growing season. Fertilizer prices skyrocketed that year as global supplies were short and demand was strong.

Seed and chemical costs have been impacted by the large-scale adoption of genetically modified soybean varieties. The first genetically modified varieties entered the marketplace in the mid-1990s. The National Agricultural Statistics Service of the USDA began tracking adoption in 2000 by capturing the percentage of the crop planted to the varieties. By that time, over half of the US soybean crop came from genetically modified varieties. In 2007, that percentage rose to over 90%. And since 2014, the percentage of the soybean crop from genetically modified varieties has held steady at 94%. For soybeans, the major modification was the inserted tolerance to herbicides, specifically glyphosate (or as it is commonly called by its product name, Roundup). This change simplified weed control during soybean production and reduced chemical costs. However, as weeds develop resistance, those cost savings are eroding.

Figure 1.6 outlines soybean production costs for the 2016 crop year in Iowa. Given a yield target of 50 bushels per acre, the total cost per acre to produce soybeans is $533.30. That equates to $10.67 per bushel. Pre-harvest machinery costs (covering the pre-planting preparation and tillage practices) are $38.50 per acre. Seed costs are $53.60 per acre. Fertilizer and lime expenses add up to nearly as much as seed, $53.05 per acre. Other pre-harvest expenses contribute, such as crop insurance and interest payments, roughly $56 per acre. Once harvest begins, additional machinery costs of $37 per acre are incurred. Approximately 2.25 h of labor per acre are required over the course of the production cycle adding $29.25 to the costs. But the largest cost component is the land charge. For 2016, that is $266 per acre. As the figure shows, production costs change with the yield target, mainly for two reasons. First, the quality of the soil impacts potential yields. More productive

**Fig. 1.6** Soybean production budget for Iowa during the 2016 crop year (*Source* Plastina 2016)

**Herbicide Tolerant Soybeans Following Corn**

| | 45 bu. per acre | | 50 bu. per acre | | 55 bu. per acre | |
|---|---|---|---|---|---|---|
| | Fixed | Variable | Fixed | Variable | Fixed | Variable |
| Preharvest Machinery [1/] | $21.10 | $17.40 | $21.10 | $17.40 | $21.10 | $17.40 |
| Seed, Chemical, etc. | Units | | Units | | Units | |
| Seed @ $53.60 per 140 k. | 140 | $53.60 | 140 | $53.60 | 140 | $53.60 |
| Phosphate @ $0.45 per lb. | 36 | 16.20 | 40 | 18.00 | 44 | 19.80 |
| Potash @ $0.35 per lb. | 68 | 23.80 | 75 | 26.25 | 83 | 29.05 |
| Lime (yearly cost) | | 8.80 | | 8.80 | | 8.80 |
| Herbicide [2/] | | 32.20 | | 32.20 | | 32.20 |
| Crop insurance | | 6.90 | | 7.80 | | 8.70 |
| Miscellaneous | | 9.00 | | 10.00 | | 11.00 |
| Interest on preharvest variable costs (8 months @ 5.15%) | | 5.76 | | 5.98 | | 6.20 |
| Total | | $156.26 | | $162.63 | | $169.35 |
| Harvest Machinery | | | | | | |
| Combine | $15.90 | $6.80 | $15.90 | $6.80 | $15.90 | $6.80 |
| Grain cart | 6.20 | 2.70 | 6.20 | 2.70 | 6.20 | 2.70 |
| Haul | 1.93 | 1.34 | 2.15 | 1.49 | 2.36 | 1.64 |
| Handle (auger) | 0.78 | 0.75 | 0.87 | 0.83 | 0.95 | 0.91 |
| Total | $24.81 | $11.58 | $25.11 | $11.82 | $25.41 | $12.05 |
| Labor | | | | | | |
| 2.25 hours @ $13.00 | $29.25 | | $29.25 | | $29.25 | |
| Land | | | | | | |
| Cash rent equivalent | $225.00 | | $266.00 | | $296.00 | |
| Total fixed, variable | | | | | | |
| Per acre | $300.16 | $185.25 | $341.46 | $191.84 | $371.76 | $198.80 |
| Per bushel | $6.67 | $4.12 | $6.83 | $3.84 | $6.76 | $3.61 |
| Total cost per acre | $485.41 | | $533.30 | | $570.56 | |
| Total cost per bushel | $10.79 | | $10.67 | | $10.37 | |

*[1/] Chisel plow, tandem disk, field cultivate, plant, and two sprays. See the Estimated Machinery Costs table.*
*[2/] Estimates do not include any insecticide or fungicide costs.*

soils generally translate into higher yields, and those higher yields often result in higher land rental costs. Second, fertilization schemes shift higher with the yield target.

## 1.3 Marketing

Soybean, like many row crops, is a bulk commodity. This means that soybeans produced by one farmer/practice/variety are not differentially marketed or priced from soybeans produced by others (with some rare exceptions for food-grade non-genetically modified varieties or organic practices). Or to put it another way, a soybean is a soybean no matter who produces it or how it is produced. So soybean producers are transacting in a competitive market, where producers can enter and exit the market fairly easily and there is very little room for differentiation. But there is a similar situation for the entities that purchase soybeans as well. So the soybean market is made up of many sellers (producers) and buyers (country elevators, crushing facilities, river terminals, and exporters). In competitive markets like this, prices are generally equal to production costs. If prices exceed costs, the resulting profits will inspire additional production from existing

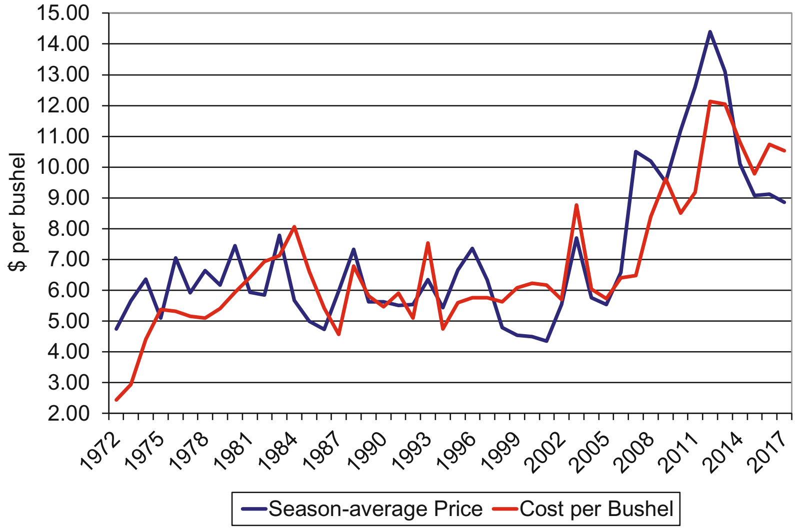


**Fig. 1.7** Iowa soybean prices and costs per bushel (*Source* USDA-NASS 2016a, b; Johanns and Plastina 2016)

and new producers, leading to larger supplies and lower prices. If costs exceed prices, the resulting losses will drive some producers out of the business, leading to supply reduction and price improvement.

Over the past 45 years, that balance between prices and costs has held on average. The cost line in Fig. 1.7 below shifts due to changes in the costs per acre, but also changes in soybean yield. In a weather-stressed year, such as 1993 (with floods affecting the central USA) or 2012 (with drought doing the same), costs jump to the reduced yields. But prices rise in those years as well. In general, the soybean market, like most agricultural markets, goes through multi-year profitability swings. Prices will exceed costs for a few years, followed up by a reversal where costs exceed prices for a few years. In the early to mid-1970s, soybean exports surged. Prices followed and the profitability of soybeans led to increased soybean plantings through the late 1970s and 1980s. Soybean supplies eventually caught up to demand, costs rose, and prices fell, leading to a soybean contraction in the late 1980s. Similar swings have continued since then. During the last 10 years, the soybean market has experienced record high prices, again spurred on by strong exports. Starting in 2007, cash soybean prices rose above $10 per bushel. And prices stayed above $10 per bushel over several years after that. The profitability generated by those high prices led farmers to increase soybean planting above 80 million acres. The recent combination of record acreage and yield has lowered prices significantly, putting pressure on the soybean industry to contract.

Given the numerous participants in the soybean market and the competitive nature of the market, soybean pricing has developed a pathway to link local market prices to global ones. Global soybean prices are mainly determined at large commodity exchanges (with the largest for soybeans being the Chicago Mercantile Exchange). The futures' prices established at these exchanges reflect the global supply and demand situation and provide all market participants current and future price signals for soybean sales and production decisions. Local soybean prices are derived from the futures prices, with the difference between the futures prices and local cash prices referred to as the basis. The basis reflects local supply and demand conditions and incorporates the transportation costs between the local market and the commodity exchange. The basis is often negative (i.e., the local cash price is below the futures price) in local markets, where soybean supplies are large, for example, in Iowa and Illinois. But the basis can become positive when soybeans are in short supply, for example, when a drought limits soybean production. This pricing formula means that local market prices are often quoted as two pieces of information, the futures price, capturing the global picture, and the basis, highlighting local changes.

## References

Gibson L, Benson G (2005) Origin, history, and uses of soybean (Glycine max). http://agron-www.agron.iastate.edu/Courses/agron212/Readings/Soy_history.htm. Accessed 25 June 2016

Johanns A, Plastina A (2016) Historical costs of crop production. http://www.extension.iastate.edu/agdm/crops/pdf/a1-21.pdf. Accessed 9 Sept 2016

Plastina A (2016) Estimated costs of crop production in Iowa—2016. http://www.extension.iastate.edu/agdm/crops/pdf/a1-20.pdf. Accessed 10 Sept 2016

United States Department of Agriculture-Foreign Agricultural Service (USDA-FAS) (2016) Production, supply, and distribution database. https://apps.fas.usda.gov/psdonline/. Accessed 27 June 2016

United States Department of Agriculture-National Agricultural Statistics Service (USDA-NASS) (2016a) Acreage. Various issues. http://usda.mannlib.cornell.edu/MannUsda/viewDocumentInfo.do?documentID=1000. Accessed 9 Sept 2016

United States Department of Agriculture-National Agricultural Statistics Service (USDA-NASS) (2016b) Quick stats database. https://www.nass.usda.gov/Quick_Stats/. Accessed 25 June 2016

# 2 Botany and Cytogenetics of Soybean

R.J. Singh

**Abstract**

Soybean [*Glycine max* (L.) Merr.], an economically important dicot legume, is a member of the family Fabaceae and belongs to the genus *Glycine* Willd. Based on classical and molecular taxonomy, the genus *Glycine* has been divided into two subgenera; the subgenus *Soja* (Moench) F.J. Hermann includes soybean and its wild annual progenitor *G. soja* Sieb. & Zucc. Both species contain $2n = 40$ chromosomes, are cross-compatible, produce fertile $F_1$ plants, and belong to the primary gene pool. The subgenus *Glycine* consists of 26 wild perennial species. Vegetative and reproductive morphology of soybean has been examined extensively. The cytogenetic knowledge of soybean lags far behind that of other model economically important crops (viz. rice, maize, wheat, tomato), because its somatic chromosomes are symmetrical and only one pair of satellite chromosomes can be identified. Molecular linkage maps have been associated with specific chromosomes, and soybean genome has been sequenced. The soybean breeders, worldwide, are confined to crossing within the primary gene pool; thus, genetic base of soybean is very narrow. Wild perennial *Glycine* species of the tertiary gene pool have been recently exploited to broaden the genetic base of modern soybean cultivars.

## 2.1 Introduction

The soybean [*Glycine max* (L.) Merr.; $2n = 40$] is an economically very important leguminous seed crop for feed and food products that is rich in seed protein (about 40%) and oil (about 20%). Taxonomy of the genus *Glycine* Willd. is well defined based on morphological features, cytogenetics,

R.J. Singh (✉)
USDA-Agricultural Research Service,
Soybean/Maize Germplasm, Pathology, and
Genetics Research Unit, Department of Crop
Sciences, University of Illinois, 1101 West Peabody
Drive, Urbana, IL 61801, USA
e-mail: ramsingh@illinois.edu



H.T. Nguyen and M.K. Bhattacharyya (eds.), *The Soybean Genome*,
Compendium of Plant Genomes, DOI 10.1007/978-3-319-64198-0_2

and molecular methods (Chung and Singh 2008). Vegetative (Lersten and Carlson 2004) and reproductive (Carlson and Lersten 2004) morphological features of soybean have been extensively described. However, soybean is not considered a model plant for cytogenetic studies because of the large number of chromosomes ($2n = 40$) (Karpechenko 1925; *Soja hispida*; syn. *G. max*), their small and similar chromosome size (1.42–2.84 μm) (Sen and Vidyabhusan 1960), and the lack of morphological distinguishing landmarks (Singh 2003). Using primarily restriction fragment length polymorphism (RFLP) and simple sequence repeat (SSR) loci, 20 molecular linkage groups (MLGs) have been developed (Song et al. 2004; Xia et al. 2007) but not all linkage groups have been associated with the respective chromosomes (Zou et al. 2003b). The cytogenetic knowledge of the soybean lags far behind many crops such as maize, barley, rice, wheat, tomato, brassicas, pea, and faba bean (Singh 2003; Singh et al. 2007a, b).

The genetic base (number of ancestral varieties that contributed to the development of modern commercial varieties) and diversity are narrow for public soybean cultivars being grown worldwide. Soybean breeders have been confined to improving soybeans using land races and occasionally used *G. soja* Sieb. & Zucc. Soybean breeders have not exploited wild perennial *Glycine* species for broadening the genetic base of soybean (Chung and Singh 2008). Despite the apparent limitation of having a narrow genetic base for world soybean production, soybean breeding has continued to make significant progress.

The objective of this chapter is to document brief information on vegetative and reproductive features (botany) of soybean and describes cytogenetics of the genus *Glycine*. Cytogenetics covers handling of soybean chromosomes, genomes of the *Glycine* species, origin of polyploid complexes, chromosomal aberrations and wide hybridization.

## 2.2 Botany

### 2.2.1 Taxonomy

The taxonomy of wild annual and cultivated soybean is as follows:

| | |
|---|---|
| Order | Fabales |
| Family | Fabaceae (Leguminosae) |
| Subfamily | Papilionoideae |
| Tribe | Phaseoleae |
| Subtribe | Glycininae |
| Genus | *Glycine* Willd. |
| Subgenus | *Soja* (Moench) F.J. Herm. |
| Species | *Glycine soja* Sieb. & Zucc. |
| Species | *Glycine max* (L.) Merr. |

The taxonomy of the genus *Glycine* to which soybean belongs has been revised many times. Hermann (1962) divided the genus *Glycine* into three subgenera (Tables 2.1 and 2.2): The subgenus *Leptocyamus* included six wild perennial species indigenous to Australia, South China, South Pacific Islands, Philippines, and Formosa (Taiwan). The subgenus *Glycine* contained two species (*G. petitiana* from Ethiopia and *G. javanica* from India and Malaya (Malaysia)). *Glycine javanica* included two subspecies: the subspecies *G. micrantha* with four varieties and subspecies *G. pseudojavanica* with one variety, and all were indigenous to Africa (Tables 2.1). He included cultigen soybean [*G. max* (L.) Merr.] and *G. ussuriensis* Regel and Maack. in the subgenus *Soja*.

Verdcourt (1966) questioned the validity of *G. javanica* L. since it has $2n = 22$ or 44 chromosomes and the chromosomes (morphology) are larger than those of other species of the genus *Glycine*. He kept the generic name and proposed *G. wightii* (R. Grah. Ex Wight and Arn.) Verdcourt as the species name. He changed the names of the genus *Glycine* L. assigned by Hermann (1962) to *Glycine* Willd., and the names of two of the subgenera of *Glycine*: subgenus *Leptocyamus* (Benth.) Hermann became