Genetics and Genomics of Soybean

Plant Genetics and Genomics: Crops and Models

Series Editor: Richard A. Jorgensen

Forthcoming and planned volumes

Vol. 1 Genomics of Tropical Crop Plants (eds: Paul Moore/Ray Ming)
Vol. 2 Genetics and Genomics of Soybean (ed: Gary Stacey)
Vol. 3 Genetics and Genomics of Cotton (ed: Andy Paterson)
Vol. 4 Plant Cytogenetics: Genome Structure and Chromosome Function (eds: Hank Bass/Jim Birchler)
Vol. 5 Plant Cytogenetics: Methods and Instruction (eds: Hank Bass/Jim Birchler)
Vol. 6 Genetics and Genomics of the Rosaceae (eds: Kevin Folta/Sue Gardiner)
Vol. 7 Genetics and Genomics of the Triticeae (ed: Catherine Feuillet/Gary Muehlbauer)
Vol. 8 Genomics of Poplar (ed: Stefan Janssen et al.)

Gary Stacey
Editor

Genetics and Genomics of Soybean

Foreword by Bob Goldberg

Springer

Editor
Gary Stacey
University of Missouri
Columbia, MO, USA
staceyg@missouri.edu

Series Editor
Richard A. Jorgensen
University of Arizona
Tucson, AZ, USA
raj@ag.arizona.edu

ISBN: 978-0-387-72298-6 e-ISBN: 978-0-387-72299-3

Library of Congress Control Number: 2008925775

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Foreword

Genetics and Genomics of Soybean, Edited by Professor Gary Stacey, is a remarkable collection of articles by internationally-recognized experts in the field of soybean genomics – many of whom helped to develop the tools and resources necessary to establish soybean as a powerful crop to investigate important basic and applied questions of plant biology. This collection of articles provides a comprehensive up-to-date review of the field of soybean genomics, and documents how far this field has advanced in the last few years. From the vantage point of someone like myself who first began investigating the organization and expression of the soybean genome thirty years ago, the insights provided by the authors in this book indicate that soybean has indeed "come of age," and that decades-old mysteries of the soybean genome are now being illuminated. *Genetics and Genomics of Soybean* is divided into four sections: (1) soybean genome natural history and diversity – which includes chapters on the genetic variation of the soybean genome and its relationship to other legume genomes; (2) tools, resources, and approaches – which includes reviews of technological advances that are being used to study the soybean genome – including the first glimpse of how the soybean genome is being sequenced and assembled; (3) investigations of soybean biology – which contains chapters that review how genomics tools have been used to study important questions – such as seed development, host-pathogen interactions, abiotic stress, and metabolic pathways; and (4) how Roundup Ready soybeans, generated by genetic engineering, have made an impact on global soybean agriculture. The chapters in this book are essential reading for students and investigators interested in basic and applied aspects of soybean biology. They provide a timely, comprehensive review of the field of soybean genomics, document the status of where the field is today, and, most importantly, raise many exciting questions about soybean evolution and biology that can now be answered using the genomics tools and resources outlined in this important book.

Los Angeles, CA 90095-1606 Bob Goldberg

Preface

Plant genomics is revolutionizing our understanding of basic plant biology and, yet, the impact on major crop plant species is still limited. Until recently, emphasis has been placed on ‘model’ plant species (e.g., Arabidopsis, and for legumes, *Lotus japonicus* or *Medicago truncatula*, see Chapters 3 and 4). However, if these are models, then what are they models of? Where will we apply the knowledge obtained from the ‘models’? Clearly, the targets must be crop plants, which ultimately provide the benefit to mankind. However, why work with models and then test these discoveries in crop plants, when the resources are available to make the original discoveries in the crop? In this scenario, application is direct and immediate.

The Fabaceae (leguminosae) comprise the second largest family of flowering plants with 650 genera and 18000 species. The soybean is a member of the tribe Phaseoleae, the most economically important of the legume tribes (Chapter 2). The soybean, *Glycine max* (L.) Merr. is the major source of vegetable oil and protein on earth (see Chapter 1). As described in detail in this volume, knowledge of soybean genomics and genetics has advanced rapidly to the point that many of the resources previously only available for ‘model’ species are now ready for exploitation in this crop. Soybean has a very detailed genetic map (Chapter 5), a recently completed physical map (Chapter 6) and developing resources for reverse genetics to study gene function (Chapter 9). As this volume goes to press, it is anticipated that the full sequence of the soybean genome is nearing public release through the efforts of the US Department of Energy-Joint Genome Institute (see Chapter 7 for a preview). This represents a major milestone in ***Genetics and Genomics of Soybean*** and will enable practical applications for soybean improvement.

Knowledge of the soybean genome is already enhancing soybean breeding through the application of molecular assisted selection (Chapter 8). In addition, this information is being applied to both basic and applied research in priority areas. For example, the soybean seed is the major product of the plant and detailed studies, using a full repertoire of functional genomic methods, are well underway (Chapter 11). These studies include the analysis of biochemical pathways involved in both oil and protein synthesis (Chapter 12). The recent resurgence of interest in soybean as a biodiesel source makes these studies particular relevant. Soybean is also a ‘heart health food’, as designated by the US Food and Drug Association. This is in large part due to the production of a wide variety of bioactive secondary products (Chapter 13). Genetics and genomic information also have an important role to play in improving soybean production. For example, efforts are well underway to apply this information to improve stress (both biotic and abiotic) resistance (Chapters 14, 15, 16, and 17).

The world's expanding population, coupled with growing concerns about the environment and climate change, present tremendous challenges for agriculture (Chapter 1). How will we feed the future expanded population of our planet, with decreasing land in the face of rising environmental challenges? Clearly, legumes, especially soybean, can make significant contributions due to the benefits of crop rotation and influences on soil fertility. It is also clear that biotechnology (for example, in the form of transgenic crop plants) will play an ever increasing role in agriculture. However, this remains a controversial area in many parts of the world. The experience of herbicide resistant soybeans, one of the first transgenic crops to be grown on a large scale, may provide insight into the benefits and future use of biotechnology in agriculture (Chapter 19).

This volume represents a compilation of timely topics pertinent to modern genetics and genomics of soybean. The chapters are written by recognized experts and provide an excellent primer for the no-doubt astounding developments that will come in the future from the full knowledge of the soybean genome sequence. I thank all of the authors for their wonderful and timely contributions. I also thank Jinnie Kim, Senior Editor, Springer Science and Business Media, for originally suggesting this idea and aiding in its development. Finally, special thanks to Jillian Slaight, Editorial Assistant, for moving the volume into production.

Columbia, MO, USA Gary Stacey

Contents

Part I Natural History and Genetic Diversity

Part II Tools, Resources and Approaches

Part III Investigations of Soybean Biology

Part IV Early Messages

Contributors

John All
Department of Entomology, University of Georgia, Athens, GA 30602-6810, USA, e-mail: jall@uga.edu

Thomas J. Baum
Department of Plant Pathology, Iowa State University, Ames, IA 50011, USA, e-mail: tbaum@iastate.edu

Kristin D. Bilyeu
USDA-ARS, Plant Genetics Research Unit, University of Missouri-Columbia, MO 65211, USA, e-mail: bilyeuk@missouri.edu

H. Roger Boerma
Department of Crop and Soil Sciences, University of Georgia, Athens, GA 30602-6810, USA, e-mail: rboerma@uga.edu

Steven Cannon
USDA-ARS, Department of Agronomy, Iowa State University, Ames, IA 50011, USA, e-mail: scannon@iastate.edu

Jarrod Chapman
Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA e-mail: JChapman@lbl.gov

Perry B. Cregan
Soybean Genomics and Improvement Laboratory, U.S. Department of Agriculture, Agricultural Research Service, BARC-West, Beltsville, Maryland, MD 20705, USA, e-mail: creganp@ba.ars.usda.gov

Paul Gepts
Department of Agronomy and Range Science, University of California, Davis, Davis, CA 95616, USA, e-mail: plgepts@ucdavis.edu

Delkin Orlando Gonzalez
Department of Crop Sciences, University of Illinois, Urbana, Illinois, IL 61801, USA, e-mail: dogonzal@uiuc.edu

Madge Graham
Department of Plant Pathology, Ohio State University, Columbus, OH 43210, USA, e-mail: graham.19@osu.edu

Michelle A. Graham
USDA-ARS CICGRU, Department of Agronomy, Iowa State University, Ames, IA 50011, USA, e-mail: magraham@iastate.edu

Terry Graham
Department of Plant Pathology, Ohio State University, Columbus, OH 43210, USA, e-mail: graham.1@osu.edu

David Grant
USDA-ARS CICGRU, Department of Agronomy, Iowa State University, Ames, IA 50011, USA, e-mail: dgrant@iastate.edu

Tomoko Hatanaka
Kobe University Kobe, 657–8501, Japan, e-mail: thata@kobe-u.ac.jb

Uffe Hellsten
Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA, e-mail: UHellsten@lbl.gov

David F. Hildebrand
University of Kentucky, Lexington, KY 40546-0312, USA,
e-mail: dhild@uky.edu

Scott A. Jackson
Department of Agronomy, Purdue University, 915 W. State St., West Lafayette, IN 47907, USA, e-mail: sjackson@purdue.edu

Sarah Jones
Department of Crop Sciences, University of Illinois, Urbana, Illinois, IL 61801, USA, e-mail: sijones@uiuc.edu

Nicholas Kalaitzandonakes
The Economics and Management of Agrobiotechnology Center, Department of Agricultural Economics, University of Missouri-Columbia, MO 65211-6200, USA, e-mail: KalaitzandonakesN@missouri.edu

Srinivasa Konduru
The Economics and Management of Agrobiotechnology Center, Department of Agricultural Economics, University of Missouri-Columbia, MO 65211-6200, USA, e-mail: spkdkc@mizzou.edu

John Kruse
The Economics and Management of Agrobiotechnology Center, Department of Agricultural Economics, University of Missouri-Columbia, MO 65211-6200, USA, e-mail: krusej@missouri.edu

Matt Lavin
Department of Plant Sciences, Montana State University, Bozeman, MT 59717, USA, e-mail: mlavin@montana.edu

Runzhi Li
University of Kentucky, Lexington, KY 40546-0312, USA,
e-mail: Runzhi.Li@uky.edu

M. A. Saghai Maroof
Department of Crop and Soil Environmental Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA, e-mail: smaroof@vt.edu

Phillip E. McClean
Department of Plant Sciences, North Dakota State University, Fargo, ND 58105, USA, e-mail: phhillip.mcclean@ndsu.edu

Melissa G. Mitchum
Christopher S. Bond Life Sciences Center, Division of Plant Sciences, University of Missouri-Columbia, MO 65211, USA, e-mail: goellnerm@missouri.edu

Rex T. Nelson
USDA-ARS CICGRU, Department of Agronomy, Iowa State University, Ames, IA 50011, USA, e-mail: nelsonrt@iastate.edu

Henry T. Nguyen
National Center for Soybean Biotechnology, Division of Plant Sciences, 1-31 Agriculture Building, University of Missouri-Columbia, MO 65211, USA,
e-mail: nguyenhenry@missouri.edu

Wayne Parrott
Department of Crop and Soil Sciences, University of Georgia, Athens, GA 30602-6810, USA, e-mail: wparrott@uga.edu

M. S. Pathan
National Center for Soybean Biotechnology, Division of Plant Sciences, University of Missouri-Columbia, MO 65211, USA, e-mail: pathanm@missouri.edu

Daniel Rokhsar
Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA; Department of Molecular and Cell Biology, Center for Integrative

Genomics, University of California at Berkeley, Berkeley, CA 94720, USA,
e-mail: DSRokhsar@lbl.gov

Jessica A. Schlueter
Department of Agronomy, Purdue University, 915 W. State St., West Lafayette, IN 47907, USA, e-mail: acissej@purdue.edu

Jeremy Schmutz
Department of Energy Joint Genome Institute, Stanford Human Genome Center, 975 California Avenue, Palo Alto, CA 94304, USA,
e-mail: jeremy@shgc.stanford.edu

Randy C. Shoemaker
USDA-ARS CICGRU, Department of Agronomy, Iowa State University, Ames, IA 50011, USA, e-mail: rcsshoe@iastate.edu

David A. Sleper
National Center for Soybean Biotechnology, Division of Plant Sciences, University of Missouri-Columbia, MO 65211, USA, e-mail: sleperd@missouri.edu

Françoise Thibaud-Nissen
Department of Crop Sciences, University of Illinois, Urbana, Illinois, IL 61801, USA, e-mail: fthibaud@tigr.org

Sue A. Tolin
Department of Plant Pathology, Physiology and Weed Science, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA, e-mail: stolin@vt.edu

Dominic M. Tucker
Department of Crop and Soil Environmental Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA, e-mail: dotucke2@vt.edu

Jigyasa Tuteja
Department of Crop Sciences, University of Illinois, Urbana, Illinois, IL 61801, USA, e-mail: tuteja@uiuc.edu

Brett M. Tyler
Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0477, USA, e-mail: bmtyler@vt.edu

Babu Valliyodan
National Center for Soybean Biotechnology, Division of Plant Sciences, 1-31 Agriculture Building, University of Missouri-Columbia, MO 65211, USA,
e-mail: valliyodanb@missouri.edu

Lila Vodkin
Department of Crop Sciences, University of Illinois, Urbana, Illinois, IL 61801, USA, e-mail: l-vodkin@uiuc.edu

David Walker
Department of Crop and Soil Sciences, University of Georgia, Athens, GA 30602-6810, USA; Current address: USDA-ARS & University of Illinois, National Soybean Research Center, USA, e-mail: walkerdr@uiuc.edu

Richard F. Wilson
National Program Leader, Oilseeds & Biosciences LLC, United States Department of Agriculture, Agricultural Research Service, National Program Staff, Beltsville, MD 20705-5139, USA, e-mail: richard.wilson@ars.usda.gov, rfwilson@mindspring.com

Oliver Yu
Donald Danforth Plant Sciences Center, St. Louis, MO 63132, USA, e-mail: OYu@danforthcenter.org

Gracia Zabala
Department of Crop Sciences, University of Illinois, Urbana, Illinois, IL 61801, USA, e-mail: g-zabala@uiuc.edu

Shuquan Zhu
Department of Crop and Soil Sciences, Center for Applied Genetic Technologies, University of Georgia, Athens, GA 30602-6810, USA, e-mail: szhu@uga.edu

Part I
Natural History and Genetic Diversity

Chapter 1
Soybean: Market Driven Research Needs

Richard F. Wilson

Introduction

Soybean (*Glycine max* L. Merrill) is the dominant oil-seed in world trade, accounting for about 56% of global oilseed production. The contribution this crop makes to the current global economy is estimated conservatively at $48.6 billion or about $18.7 billion in the U.S. alone. Demand for soybean remains strong and continues to grow because it is used as an ingredient in the formulation of a multitude of food, feed and industrial products. These applications include a wide range of soyfoods, shortening to biodiesel applications for soybean oil, and feed to vegetable protein substitutes for meat and dairy products for soybean meal/protein. In addition, soybean is a primary source of high-value secondary co-products such as lecithin, vitamins, nutraceuticals and anti-oxidants. The U.S., Brazil and Argentina are the predominant soybean producing countries, but global soybean production area has reached an apparent plateau. If this trend continues or worsens, extreme pressure will be placed on 'genetic-gain' in soybean yielding ability to ensure adequate supply to meet the escalating demand for soybean and soybean products. Failing to provide an ample supply of soybeans would be felt throughout the world, beginning with a decline in soybean exports. Currently, the U.S. and Brazil crush only about half of their annual production; whereas countries like Argentina, the People's Republic of China and the European Union-25 essentially crush their entire annual supply. Thus, the U.S. and Brazil are the only countries with the flexibility to export whole soybeans to major customers such as the People's Republic of China and the European Union. However, future levels of soybean exports likely will be eroded by the need to service greater domestic use. Already there are signs of a transition toward greater production and trade of refined vegetable oil and meal among soybean producing countries. Currently, the U.S. consumes 95% of its domestic production of soybean oil. Because of the emerging market for biodiesel,

---

R.F. Wilson
United States Department of Agriculture, Agricultural Research Service, National Program Staff, Oilseeds & Biosciences LLC
e-mails: richard.wilson@ars.usda.gov; rfwilson@mindspring.com

G. Stacey (ed.), *Genetics and Genomics of Soybean*,

a deficit in U.S. soybean oil production is projected by 2020. By the same token, U.S. end-stocks typically support only a 4–6 month supply of soybean meal. With anticipated growth in the livestock and aquaculture industries, a deficit in U.S. soybean meal production is predicted by 2020. If genetic-gains in the improvement of U.S. soybean production are not sufficient to ensure an adequate domestic supply of soybean meal, the U.S. may be in jeopardy of losing domestic livestock production to off-shore locations. The prospect for future deficits in U.S. production of soybeans, soybean oil and soybean meal gains credibility from the national emphasis on reduction of U.S. dependence on petroleum and fossil-fuels. It is estimated that about 30% of the U.S. corn crop may be converted to ethanol production by 2010, and the projected goal of 700 million gallons of biodiesel would consume 23–25% of U.S. annual production of soybean oil. Strong demand for ethanol production already has resulted in higher corn prices, which favors future increases in corn acreage at the expense of soybean. Thus, the impact of bioenergy alternatives on the ability to provide an adequate supply of soybean and soybean products is a serious challenge that must be addressed by the U.S. and global soybean industrial and research communities. It should be noted that significant progress is being made to enhance soybean yielding ability, largely through random exploitation of the wealth of genetic diversity that is harbored among accessions of soybean germplasm collections. However, to maintain the current rate of growth in U.S. soybean supply, assuming no change in U.S. soybean production area, it appears that U.S. soybean yields would have to increase to an average 4085 kg/ha (60.8 Bu/acre) by 2020. Optimistically, yielding ability can be enhanced, but the question that now faces the soybean genetics community is whether or not continued genetic-gains of the required magnitude may be attained through a traditional breeding approach alone. Better understanding of the genetic regulation of seed constituent composition also is needed to help ensure an adequate supply of high-quality protein and oil. In addition, effective strategies for protection against crop losses to diseases such as Asian Soybean Rust, pests such as soybean cyst nematode, and environmental stresses will require detailed analysis of the soybean genome. ‘Mining’ the soybean genome for this information will facilitate the development of useful DNA markers for genes of interest. Integration of those ‘genomic tools’ with modern breeding programs will lead to more effective utilization of the genetic diversity in *Glycine max*. The ‘tools’ and knowledge gained from soybean genomics will enable the ‘next generation’ advances in soybean breeding that are needed now to meet the needs of U.S. and global agriculture.

Origin and Development of Soybean as a Crop

Domestication of Soybean

Cultivated soybean [*Glycine max* (L.) Merr.] appears to draw its origin from a domestication event in the wild soybean (*Glycine soja* Seib. et Zucc.) that may have occurred in ancient central or southern China nearly 5000 years ago (Gai 1997; Gai and Guo 2001). This estimate is derived, in part, from references to soybean which appeared in Chinese literature during the Shang dynasty from 1700 to 1100 BC (Qiu et al. 1999). However, anecdotal evidence and oral traditions recorded during that time also suggest a much older association of soybean in the Chinese culture (Guo 1993).

The versatility of soybean in preparing various soyfoods is perhaps the major factor that favored its cultivation as an agricultural crop. Soyfoods, like *tofu* (thought to be invented during the Han Dynasty), *douchi* (a fermented salty garnish made from whole soybean) and *doujiang* (a thick sauce made from fermented soybean) were then, and remain today, staples of the Chinese diet. In addition, the beginnings of the soybean oil industry may be traced to China, at least 1000 years ago, when historical records report the common practice of frying *tofu* with soy oil (Gai and Guo 2001).

However, the cultivation of domesticated soybean beyond ancient China did not spread rapidly. For example, soybean may have been introduced to Japan from China or Korea only about 2000 years ago (Li and Nelson 2001). Documented reference to soybean cultivation in Japan does not appear until the early Yayoi culture (Kihara 1969; Sugiyama 1992). In any case, soybean has long been important in the Japanese diet, leading to the development of a unique food culture. Japanese innovations in soyfoods include: vegetable soybean (*edamame*), soybean sprouts (*moyashi*), soymilk (*tonyu*), frozen and baked soybean curd (*kori-dofu*, *yaki-dofu*). Small-seeded soybean may be used for fermented soybean (*natto*), and boiled or fermented medium-sized seeds are used for the production of soybean paste (*miso*). Yellow or green soybean meal (*kinako*) is used in confectionery products or can be fermented to produce soy sauce (*shoyu*) (Wilson 1995).

Soybean was first introduced into North America by Samuel Bowen in 1765, principally to manufacture soy sauce. In 1770, Benjamin Franklin also experimented with soybean in the U.S.; however, his interests were limited to its utility as a forage and ground cover (Hymowitz and Harlan 1983). It was not until early in the 20th century, when the impetus for modern U.S. soybean production was discovered. This occurred in 1915, when soybeans were first crushed for oil in Elizabeth City, North Carolina (Wilson 1987).

The discovery of soybean as an important source of vegetable oil permanently changed the perception of soybean from forage to a seed crop. This transition brought the need for more productive or agronomic types of soybean. By the early 1930s, the United States Department of Agriculture (USDA) and the State Agricultural Experiment Stations at land grant universities established soybean breeding programs in the northern and southern states (Bernard et al. 1988). These efforts were enabled and strengthened by the acquisition and identity preservation of over 4000 soybean landraces from China by the USDA. Today, the USDA soybean germplasm collection contains over 18,000 types of *Glycine max* and is actively used to ensure access to a broad range of genetic diversity for cultivated soybean (USDA, ARS 2007).

By 1950, nearly 100% of the U.S. soybean crop was grown for seed, and the U.S. became the world leader in soybean production. Again, it was the functional utility of soybean seed constituents in a wide array of products that provided the basis for development of the U.S. soybean industry. Even today, as a conservative estimate, food manufacturers in the U.S. routinely create over 400 new food products with soy as an ingredient each year (Liu 1997). Products from soybean oil include: margarine, shortenings, baking and frying fats. Soybean oil also is used in industrial products including soap, cosmetics, resins, plastics, inks, crayons, solvents, clothing, and biodiesel. Soybean meal provides the high-protein feed ingredient that sparked an American revolution in poultry and swine production, and more recently the aquaculture industry. Dietary uses for soy flour in the form of soy concentrate and soy protein isolate include formula for lactose-intolerant infants, and vegetable protein substitutes for meats and dairy products. Industrial uses for soy-protein include coatings, adhesives and building materials. In addition, soybean is the primary source of high-value co-products such as lecithin, vitamins, nutraceuticals and anti-oxidants.

World Soybean Production

Within the past 60 years, an infinitesimal period during its domestication, soybean emerged as the dominant oilseed in world trade. In 2005, the USDA (USDA, FAS 2007) estimated world soybean production at 218 million metric tons (MMT); about 56% of total global oilseed production which includes copra, cottonseed, palm kernel, peanut, rapeseed (canola) and sunflower-seed (Table 1.1). Soybean also is distinguished among these oilseed crops as the primary high-energy, high-protein ingredient for livestock feed. The trading standard set by the National Oilseed Processors Association (NOPA) for high-protein soybean meal is 48% crude protein. No other oilseed meal matches that level of protein or possesses a more desirable dietary complement of essential amino acids. Thus, soybean meal commands a dominant position with a 69% share of the world vegetable protein market. However,

Table 1.1 World Production of Major Oilseeds and Oilseed Product, 2005/06

Commodity	Oilseeds		Meal		Oil	
	MMT	%	MMT	%	MMT	%
Copra	5.8	1.5	1.8	0.9	NA	0.0
Coconut	NA	0.0	NA	0.0	3.5	3.0
Cottonseed	42.5	10.9	14.3	6.8	4.6	3.9
Olive	NA	0.0	NA	0.0	2.3	1.9
Palm	NA	0.0	NA	0.0	36.0	30.5
Palm Kernel	10.0	2.6	5.2	2.5	4.4	3.7
Peanut	33.7	8.7	6.0	2.9	5.2	4.4
Rapeseed	48.6	12.5	26.3	12.5	17.2	14.6
Soybean	218.0	56.1	144.7	69.0	34.3	29.1
Sunflowerseed	29.8	7.7	11.2	5.4	10.4	8.8
Total	388.4	100.0	209.6	100.0	117.8	100.0

United States Department of Agriculture, Foreign Agricultural Service, 2007

Table 1.2 World Production of Soybean and Soybean Products, 2005/06

Country of Origin	Seed		Meal		Oil	
	MMT	%	MMT	%	MMT	%
United States	83.4	38.2	37.4	25.9	9.3	27.0
Brazil	55.0	25.2	21.7	15.0	5.4	15.7
Argentina	40.5	18.6	25.0	17.3	6.0	17.5
China, PRC	16.4	7.5	27.3	18.9	6.1	17.9
India	6.3	2.9	4.3	3.0	1.0	2.8
Paraguay	4.0	1.8	na	0.0	na	0.0
Canada	3.2	1.4	na	0.0	na	0.0
Other	9.4	4.3	15.6	10.8	3.5	10.3
EU-25	na	0.0	10.4	7.2	2.4	6.9
Mexico	na	0.0	3.0	2.1	0.7	1.9
Total	218.0	100.0	144.7	100.0	34.3	100.0

United States Department of Agriculture, Foreign Agricultural Service, 2007

there is significantly more competition among sources of vegetable oil. Soybean and palm lead that market with equal shares accounting for about 70% of total vegetable oil production.

The predominant soybean producing countries at this time are the U.S., Brazil, and Argentina. Although the U.S. produces more soybeans than any other single country, the South American countries of Brazil, Argentina, Paraguay plus Mexico collectively have surpassed North American soybean production (Table 1.2). The demographics for world production of soybean meal and oil follow similar trends, where the U.S. leads Brazil, Argentina and the People's Republic of China (PRC). It follows that these four countries account for about 77% of the world's soybean crushing capacity.

The emergence of Brazil and Argentina as major soybean producing countries is attributed to a significant increase in total harvested area between 1996 and 2004

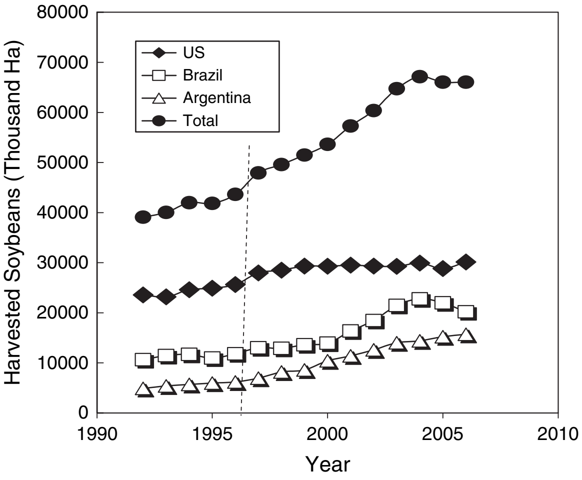

Fig. 1.1 Soybean harvested area in major producing countries

(Fig. 1.1). However, the up-surge in South American soybean production area may be short lived. Since 2004, the world total for harvested soybean area appears to have reached a plateau at about 66.4±0.6 Mha (164±1.5 million acres). This recent trend may be attributed in part to an apparent decline in total area for Brazilian soybean production, coupled with essentially no growth in U.S. area for soybean production. This situation obviously places more pressure on the translation of soybean genetic knowledge into more effective and efficient means to generate the elite yielding varieties that will help to ensure sustained future increases in global soybean production. As an example, global soybean production increased 83.6 MMT between 1996 and 2004. This achievement may be attributed to advances in soybean cultural practices and yielding ability, from a world average 2111.4 kg/ha to 2313.1 kg/ha (31.5 to 34.5 Bu/acre), plus an additional 30.7 Mha (75.8 million acres) in soybean production area. That expansion of world soybean production area exceeded the existing area for U.S. soybean production. If the world soybean production area during that period had not doubled, then it may be deduced that a global average yield of 3446.3 kg/ha (51.3 Bu/acre) would have been required to attain the level of 2004 output. Obviously, such a target is unrealistic given current technology. Hence, constant or declining global acreage is a major constraint.

Without the luxury of expanding production area, unpredictable or uncontrollable events, such as unfavorable weather or epidemics of severe diseases/pests, pose a more severe threat to global soybean production and necessitate significant and timely genetic measures to sustain the ability to keep pace with growing global market demand for soybean and soybean products.

Supply and Demand for Soybean Products

World Trends in Soybean Supply

As a result of the infusion of South American production area plus incremental gains in cultivar yielding ability, world soybean supply (production plus end-stocks) has more than doubled in the past 22 years, to a 2006 total of 282.5 MMT (Fig. 1.2). The rate of increase over that period was 8.5 MMT/yr (R^2, 0.94). At the same time, world use (crush plus exports) of soybean grew at 8.0 MMT/year (R^2, 0.92) from 1984 to 2006 (USDA, FAS 2007).

Thus, these data suggest no eminent limitation in global soybean supply in the foreseeable future, and by the same token, no relent in the growing demand for soybean products. However, there is cause for concern. Closer inspection of these data, 2006 for example, reveals a 37 MMT deficit between ‘world soybean use’ and ‘world soybean production’. Although this difference is covered by the level of end-stocks from 2005 (55.2 MMT), the carryover to 2007 (18.7 MMT, ca. 37 days supply) will be nearly 3-fold less than in 2006.

The economic equilibrium between determinants of ‘market price’ is maintained by the level of end-stocks, which acts as a necessary buffer to ensure relatively

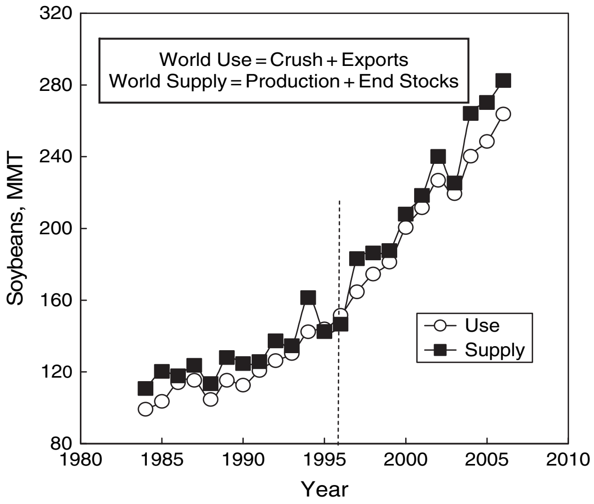

Fig. 1.2 World trends in soybean supply and demand

uninterrupted flow of produce through the marketing system. In fact, there is a very strong negative correlation between soybean end-stocks and the U.S. farm price for soybeans (Fig. 1.3). Hence, a decline in end-stocks relative to U.S. soybean supply typically is accompanied by a rise in the U.S. farm price per bushel. This statistic may be a good predictor of trends in this apparent cycle on a global scale. Although the data for 2006 are incomplete, preliminary estimates of world farm prices have tended higher in 2006–2007.

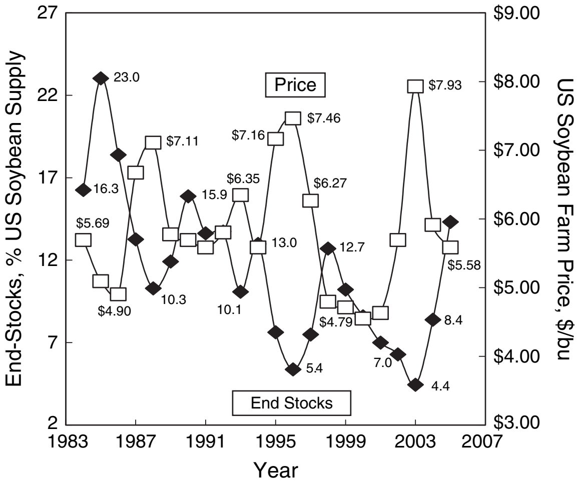

Fig. 1.3 Relation between end-stocks and price

World Trends in Soybean Use

Soybean use (crush plus exports) has increased nearly 3-fold in the past 22 years, to a total of 263.8 MMT. Although world soybean exports account for only about 27% of that total, exports also rose nearly 3-fold from 25.3 MMT in 1984 to 70.7 MMT in 2006. However, there has been a significant shift in demographics within this export market. In the past decade from 1995 to 2005, the U.S. share of the global soybean export market has declined from about 70% to 40% (Fig. 1.4). This change may be attributed to nearly a 5-fold increase (7.5 to 35.8 MMT) in soybean exports from South America, with about 71% coming from Brazil. The level of U.S. soybean exports during that period averaged 26.0 ± 3.0 MMT, which is not significantly different from the mean for the past 22 years. Therefore, Brazilian soybean exports were necessary to maintain the long-term (since 1984) growth rate of global soybean exports at 2.3 MMT/year (R^2, 0.88).

The PRC and the European Union (EU-25) are the recipients of about two-thirds of global soybean exports. Because of escalating demand for protein and oil, world soybean crushing capacity has expanded at a linear rate of 5.74 MMT/year (R^2, 0.96) from 1984 to 2006. The EU-25 crushes 95+% of their soybean imports for protein and oil; the PRC crushes about 78% of their imports plus domestic production. Among soybean exporting countries, Argentina crushes about 79% of their production, while the U.S. and Brazil crush only about half of their annual soybean harvest. Overall, the U.S., PRC, Argentina, Brazil and the EU-25 (in top to bottom order) account for 84% of the world production of soybean oil and meal. However, only the U.S. and Brazil have the apparent flexibility to provide or sustain adequate supply of whole soybeans to the Asian and European processing industries. If the U.S. and/or Brazil deploy greater soybean crushing-capacity in the near future, then the supply of soybeans to the PRC and EU-25 becomes less than certain.

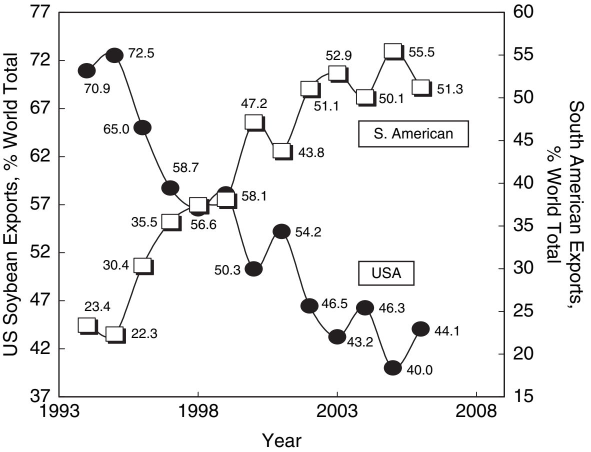

Fig. 1.4 Trends in relative share in the export soybean market

Trends in U.S. Consumption of Soybean Products

Historically, the U.S. crushes about 56% of its annual soybean production. In 2006, this resulted in about 38.5 MMT of meal and 9.2 MMT of oil (USDA, FAS 2007). Domestic consumption of soybean oil has increased at a rate of 0.185 MMT/year (R^2, 0.97) since 1984, to 8.7 MMT (Fig. 1.5), but that rate is expected to accelerate due to use of vegetable oils in the formulation of bio-diesel fuel (Conway et al. 2004). Currently, the U.S. consumes 95.4% of its annual production of soybean oil. Although U.S. end-stocks for soybean oil may exceed 0.5 MMT at this time, the long-term rate of change in domestic soybean oil production is 0.186 MMT/year (R^2, 0.93). Thus by 2020, a deficit of U.S. soybean oil is projected due to increased demand for bio-based alternatives to petroleum (United Soybean Board 2006; Westcott 2007). If genetic-gains in the improvement of U.S. soybean production are not sufficient to ensure an adequate domestic supply of soybean oil, the U.S. may become a substantial customer of Argentina and Brazil, the predominant soybean oil exporting countries.

Domestic consumption of U.S. soybean meal has increased at a rate of 0.67 MMT/year (R^2, 0.97) from 1984 to about 31 MMT in 2006. Approximately 80% of that annual domestic production is used in feeds and vegetable protein products, with the remainder in the international export market. As a result, the U.S. carries an extremely low surplus of soybean meal (Fig. 1.6). Since soybean is the preferred high-protein ingredient for livestock feed, demand for soybean meal in poultry and swine production alone is expected to grow to 29 MMT by 2020 (Westcott 2007). However, new feed markets are emerging. For example, demand for soybean meal/isolate in aquafeed is expected to reach 13 MMT by 2020 (United Soybean Board 2006). These estimates reaffirm forecasts of continued growth in U.S. demand for soybean meal. Yet, at the current rate of increase (0.75 MMT/year; R^2, 0.93), future U.S. soybean meal production, even with projected increases in

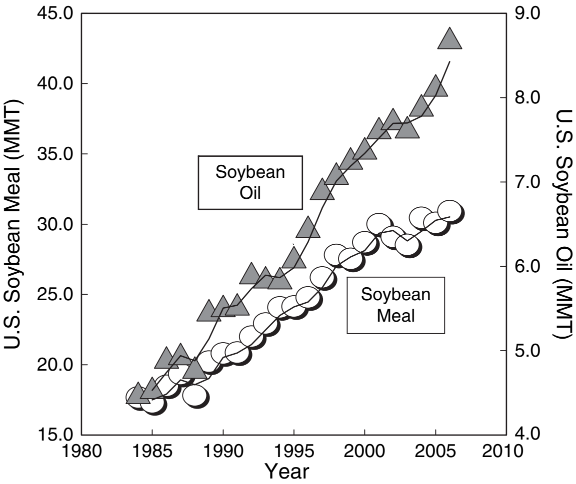

Fig. 1.5 Trend in U.S. consumption of soybean oil and meal

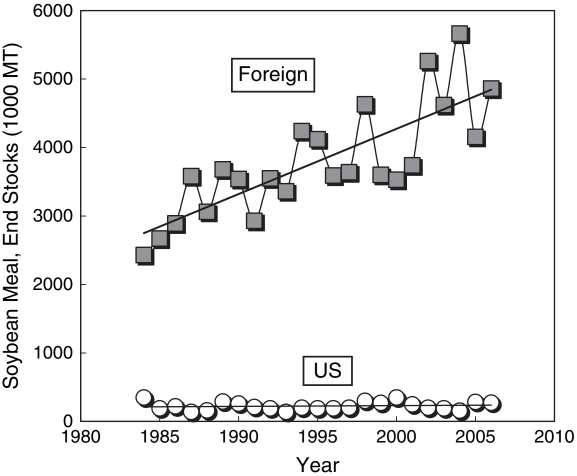

Fig. 1.6 Distribution of soybean meal end-stocks

crush-capacity, probably will not provide an adequate margin. Hence, a deficit of U.S. soybean meal is projected by 2020 due to increased demand for livestock production and aquaculture (United Soybean Board 2006). Once again, if genetic gains in the improvement of U.S. soybean production are not sufficient to ensure an adequate domestic supply of soybean meal, the U.S. may be in jeopardy of losing domestic livestock production to off-shore locations.

Further Constraints to Soybean Production

Soybean Production in an Energy Driven Environment

Although the U.S. produced the largest soybean crop on record in 2006, estimated to be 3.19 billon bushels at an average 2869 kg/ha (42.7 Bu/acre) on 30.2 Mha (74.6 million acres), the prospect for future deficits in U.S. production of soybean, soybean oil and soybean meal gain credibility in view of national energy policy to reduce U.S. dependence on petroleum and fossil-fuels. For example, the Energy Policy Act of 2005 mandates that renewable fuel use in gasoline and diesel reach 7.5 billion gallons by 2012 (Westcott 2007). In practice, higher petroleum costs combined with a variety of tax credits and import tariffs have provided economic incentives for expanded biofuel production capacity that may achieve outputs in excess of the original goal (Ash et al. 2006).

Most of the ongoing and projected biofuel expansion in the U.S. is focused on ethanol. With current technology, one bushel of corn should produce 2.8 gallons of ethanol. It is estimated that about 30% of the U.S. corn crop may be converted to ethanol production by 2010.

Biodiesel production capacity also has increased rapidly in the past five years. About one pound of refined soybean oil is required to formulate one pound of biodiesel. Based on that relation, the projected goal of 700 million gallons of biodiesel would consume 23–25% of U.S. annual production of soybean oil.

Strong demand for ethanol production already has resulted in higher corn prices, which favors future increases in corn acreage. In 2006, the U.S. produced 10.5 billion bushels of corn on 28.6 Mha (70.6 million acres), averaging 10,000 kg/ha (149.1 Bu/acre). With greater potential revenue, corn acreage could reach 36.5 Mha (90 million acres) by 2010. Much of that increase would come by adjusting crop rotations, causing a net decline in soybean acreage and soybean production.

Therefore, the impact of bioenergy alternatives on ability to provide an adequate supply of soybean and soybean products is a serious challenge that must be addressed by the U.S. and global soybean industrial and research communities. It is certain that crushing will continue to be driven by demand for livestock and aquafeeds, and government projections show no slowing of domestic demand for soybean oil, up to 2016, in food and fuel applications. Hence, even with incremental gains in the level of production, U.S. soybean exports may by necessity be significantly eroded in favor of greater crush volume.

Improving the Genetic Efficiency of Soybean Production

Given the significant challenge raised by competition within the U.S. for crop acreage, innovative research must be implemented to ensure there is continued growth in U.S. soybean production to meet the anticipated rise in demand for soybean and soybean products. Enhancement of soybean yielding ability through improved performance and reduced losses to disease and pests are obvious priorities. In that regard, the soybean breeding community already has made significant contributions though development of elite cultivars. Government statistics show fairly steady gains in average U.S. soybean yielding ability, from 2197 to 2896 kg/ha (32.7 to 43.1 Bu/acre), between 1993 and 2006. Foreign soybean production also demonstrated similar advances in yielding ability, although the yields may average 700 kg/ha (10 Bu/acre) less than in the U.S. However, to maintain the current rate of growth in U.S. soybean supply, assuming no change in U.S. soybean production area, it appears that soybean yielding ability in the U.S. would have to increase to an average 4085 kg/ha (60.8 Bu/acre) by 2020. Such a level in yielding ability may be achieved, but the question that now faces the soybean genetics community is whether or not continued genetic-gains of the required magnitude may be attained through a traditional breeding approach alone.

The great reservoir of genetic diversity that is harbored among the accessions of the world's soybean germplasm collections provides a foundation for future advances in genetic technology that are needed to provide elite soybean cultivars with adequate protection from pests and diseases, improved product quality and greater yielding ability. However, these putative genes reside in more than 156,849 accessions of *Glycine max* in about 40 different collections in 20 countries around the world (Carter et al. 2004). The PRC, Taiwan, U.S. and Japan account for about

74% of the world's repository of soybean germplasm (about half of that total is held in the PRC). Many of these accessions are not publicly available to the research community, but even so there has been little effort to characterize the material to improve its utility. Association of phenotypic traits with genotypic markers would be an extremely desirable step that is needed to help distinguish unique accessions, which in turn will facilitate the timely use of valuable genes in variety development.

Soybean genomics research, through analysis and comparison of genomic differences among unadapted and selected populations, will enable a better understanding of the genetic regulation of seed constituent composition. Such knowledge will augment efforts to ensure there is adequate supply of high-quality protein and oil. In addition, effective strategies for protection against crop losses to diseases such as Asian Soybean Rust, pests such as soybean cyst nematode, and environmental stresses will benefit from detailed analysis of the soybean genome. 'Mining' the soybean genome for this information will facilitate the development of useful DNA markers for genes of interest. Integration of those 'genomic tools' with modern breeding programs will lead to more effective utilization of the genetic diversity in *Glycine max*. The 'tools' and knowledge gained from soybean genomics will enable the 'next generation' advances in soybean breeding that are needed now to meet the needs of U.S. and global agriculture.

References

Ash, M., Livezey, J., and Dohlman, E. (2006) *Soybean Backgrounder*. U.S. Department of Agriculture, Economic Research Service, OCS-2006-01, Washington DC.

Bernard, R.L., Juvik, G.A., Hartwig, E.E., and Edwards, C.J., Jr. (1988) *Origins and Pedigrees of Public Soybean Varieties in the United States and Canada*. U.S. Department of Agriculture Technical Bulletin 1746. U.S. Gov. Print. Office, Washington, DC.

Carter, T.E., Jr., Nelson, R.L., Sneller, C.H., and Cui, Z. (2004) Genetic diversity in soybean. In H.R. Boerma and J.E. Specht (Eds.) *Soybeans: Improvement, Production and Uses*. 3rd Edition. American Society of Agronomy, Inc., Madison, Wisconsin, pp.303–416.

Conway, R.K., Erbach, D., Duncan, M., Gallagher, P., Shane, P.L., Smith, P.F., Tumbleson, M.E. (2004). *Bioenergy: Pointing to the Future*, 27A, Council for Agricultural Science and Technology, Ames, IA.

Gai, J. (1997) Soybean Breeding. In: J. Gai (Ed.), *Plant Breeding: Crop Species*. China Agriculture Press, Beijing, pp. 207–251.

Gai, J., and Guo, W. (2001) History of maodou production in China. In: T.A. Lumpkin and S. Shanmugasundaram (Eds.), *Proceedings Second International Vegetable Soybean Conference*, Washington State University, Pullman, WA., pp. 41–47.

Guo, W. (1993) *The History of Soybean Cultivation in China*. Hehai University Press, Nanjing, Jiangsu, China.

Hymowitz, T., and Harlan, J. R. (1983) Introduction of soybean to North America by Samuel Bowen in 1765. Economic Botany. 37, 371–379.

Kihara, H. (1969). History of biology and other sciences in Japan in retrospect. In: C. Oshima (Ed.), *Proceedings XII International Congress of Genetics*, The Science Council of Japan, pp. 49–70.

Li, Z., and Nelson, R. L. (2001) Genetic diversity among soybean accessions from three countries measured by RAPDs, Crop Sci. 41, 1337–1347.

Liu, K. (1997). *Soybeans: Chemistry, Technology, and Utilization*, Springer, New York.

Qiu, L., Chang, R., Sun, J., Li, X., Cui, Z., and Li, Z. (1999) The history and use of primitive varieties in chinese soybean breeding. In: H.E. Kauffman (Ed.), *Proceedings World Soybean Research Conference VI*, Superior Printing, Champaign, IL, pp.165–172.

Sugiyama, S. (1992) On the origin of soybean. *Glycine max* Merrill. J. Brewing Society of Japan 87, 890–899.

United Soybean Board. (2006) Soybean Meal Evaluation to 2020, St. Louis, MO.

U.S. Department of Agriculture, Agricultural Research Service, National Genetic Resources Program. Germplasm Resources Information Network-(GRIN). [Online Database] National Germplasm Resources Laboratory, Beltsville, MD. Available: http://www.ars-grin.gov/var/apache/cgi-bin/npgs/html/ (01 February 2007).

U.S. Department of Agriculture, Foreign Agricultural Service, (2007) Oilseeds: World Markets and Trade, Washington, D.C.

Westcott, P. (2007) Agricultural Projections to 2016, OCE 2007-01. United States Department of Agriculture, Office Chief Economist, Washington, DC.

Wilson, L.A. (1995) Soy Foods. In: D.R. Erickson (Ed.), *Practical Handbook of Soybean Processing and Utilization*, AOCS Press, Champaign, IL, pp. 428–459.

Wilson, R.F. Seed metabolism. p.643–686. In: Wilcox, J.R. (ed.). Soybeans: Improvement, Production and Uses. Second Edition. Am. Soc. Agron., Madison, WI, 888 pp. 1987.

Chapter 2
Soybean Molecular Genetic Diversity

Perry B. Cregan

Introduction

The cultivated soybean [*Glycine max* (L.) Merr.] and the wild soybean (*Glycine soja* Seib. et Zucc.) are annuals and the two members of the *Glycine* subgenus. *G. soja* grows wild in China, Japan, Korea, Russia and Taiwan (Hymowitz 2004). It is generally accepted that cultivated soybean was domesticated 3000–5000 years ago on the Chinese mainland from the wild soybean (Hymowitz and Newell 1981). Cultivated soybean exhibits wide phenotypic variability in terms of seed shape, size, color, and chemical composition; plant morphology and maturity, as well as resistance to a broad range of biotic and abiotic stresses. This genetic diversity and the underlying genetic control of numerous specific traits were described in works such as the recent Third Edition of Soybeans: Improvement, Production and Uses (Boerma and Specht 2004). In particular, Carter et al. (2004) thoroughly documented genetic diversity in terms of the formation, collection, evaluation and utilization of diversity by soybean geneticists and breeders in North American and Asia over 70 years and the impacts of their work on genetic diversity. It is the intent of this review to specifically focus on molecular genetic diversity of the nuclear genome and the multitude of research that was directed at the assessment of molecular diversity of cultivated and wild soybean. This research employed a number of different molecular genetic tools beginning with the analysis of isozyme variation followed by a range of DNA marker types and ultimately variation in DNA sequence. The literature relating to the assessment of isozyme variability in *G. max* and *G. soja* recently received a thorough review by Palmer et al. (2004) and will not be considered here.

The first reports of the assessment of genome-wide molecular genetic diversity of the soybean nuclear genome began in the 1980s with the application of restriction fragment length polymorphism technology (RFLP) (Roth and Lark 1984; Apuya et al. 1988). Subsequent analyses employed RFLP, random amplified polymorphic

---

P.B. Cregan
Soybean Genomics and Improvement Laboratory, U.S. Department of Agriculture, Agricultural Research Service, Beltsville, Maryland
e-mail: creganp@ba.ars.usda.gov

G. Stacey (ed.), *Genetics and Genomics of Soybean*,

DNA (RAPD) or arbitrary primer PCR, amplified fragment length polymorphism (AFLP), microsatellite or simple sequence repeat (SSR), and DNA sequence analysis for the quantification of genetic diversity in both cultivated and wild soybean. This research had a number of different objectives including (1) the assessment of particular DNA marker systems for appropriately distinguishing and grouping cultivated and wild genotypes, (2) the quantification and comparison of diversity within and among various groups of cultivated and/or wild soybean genotypes (3) the use of genetic diversity estimates as tools in soybean breeding for increasing useful genetic variation, (4) the development of unique DNA fingerprints for genotype and cultivar identification and (5) the assessment of linkage disequilibrium.

Applicability of DNA Marker Types in Soybean

Restriction Fragment Length Polymorphism (RFLP)

Apuya et al. (1988) analyzed 300 RFLP probes selected as low-copy clones in Southern hybridizations to genomic DNA of the genetically distinct soybean cultivars Minsoy and Noir 1. Genomic DNAs were digested with a number of different restriction endonucleases in order to detect RFLP. Of the 300 probes examined only one in five was polymorphic. Despite the low level of polymorphism, 27 loci were analyzed in a population of F_2 plants derived from Minsoy × Noir 1. All loci segregated in a Mendelian fashion and 11 of the 27 loci were contained in four linkage groups. Keim et al. (1989) conducted a survey of RFLP via the analysis of 48 cultivated, eight wild and two *G. gracilis* genotypes using 17 probes to assess the allelic structure of RFLP markers and to identify diverse genotypes that would maximize variability in a resulting mapping population. The *G. gracilis* genotypes were previously joined with *G. max* (Hermann 1962) but were included to maximize morphological diversity in the sampling of genotypes. Extremely low levels of RFLP were recorded despite the diversity of the germplasm analyzed. Two of the 17 probes detected three alleles per locus while the remaining 15 detected only two. The *G. max* genotype A81-356022 and *G. soja* PI 468916 were identified as being particularly diverse with a high level of RFLP that was approximately two-fold higher than that of the Minsoy × Noir 1 cross identified by Apuya et al. (1988). Based upon these data, as well as previous analysis of these two genotypes, a mapping population was created from the interspecific cross of A81-356022 × PI 468916. In a subsequent report, Keim et al. (1992) analyzed 132 RFLP probes in 18 ancestors of U.S. cultivars (ancestral cultivars) as well as 20 adapted cultivars. One objective was to estimate the usefulness of the probes in revealing variation in adapted germplasm. Only one in five markers were informative in any pair of adapted soybean genotypes, again suggesting the relatively low level of RFLP particularly among adapted soybean genotypes.

Skorupska et al. (1993) assessed the feasibility of using the markers from the A81-356022 × PI 468916 RFLP map in the distinct subpopulation of soybean