Compendium of Plant Genomes
Series Editor: Chittaranjan Kole

Marcelino Pérez de la Vega
Marta Santalla
Frédéric Marsolais *Editors*

The Common Bean Genome

Springer

Compendium of Plant Genomes

Series editor

Chittaranjan Kole
Raja Ramanna Fellow, Department of Atomic Energy, Government of India,
Kalyani, India

Whole-genome sequencing is at the cutting edge of life sciences in the new millennium. Since the first genome sequencing of the model plant Arabidopsis thaliana in 2000, whole genomes of about 70 plant species have been sequenced and genome sequences of several other plants are in the pipeline. Research publications on these genome initiatives are scattered on dedicated web sites and in journals with all too brief descriptions. The individual volumes elucidate the background history of the national and international genome initiatives; public and private partners involved; strategies and genomic resources and tools utilized; enumeration on the sequences and their assembly; repetitive sequences; gene annotation and genome duplication. In addition, synteny with other sequences, comparison of gene families and most importantly potential of the genome sequence information for gene pool characterization and genetic improvement of crop plants are described.

Interested in editing a volume on a crop or model plant? Please contact Dr. Kole, Series Editor, at ckole2012@gmail.com

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

Marcelino Pérez de la Vega
Marta Santalla · Frédéric Marsolais
Editors

The Common Bean Genome

Springer

Editors
Marcelino Pérez de la Vega
Facultad de Biología
Universidad de León
León
Spain

Marta Santalla
Misión Biológica de Galicia, Grupo de Biología de Agrosistemas
Consejo Superior de Investigaciones Científicas
Pontevedra
Spain

Frédéric Marsolais
Genomics and Biotechnology
London Research and Development Centre, Agriculture and Agri-Food Canada
London, ON
Canada

ISSN 2199-4781 ISSN 2199-479X (electronic)
Compendium of Plant Genomes
ISBN 978-3-319-63524-8 ISBN 978-3-319-63526-2 (eBook)
https://doi.org/10.1007/978-3-319-63526-2

Library of Congress Control Number: 2017956330

© Springer International Publishing AG 2017

Chapter 10 is published with kind permission of the Her Majesty the Queen in Right of Canada.

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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, and Devleena

Chittaranjan Kole

Preface to the Series

Genome sequencing has emerged as the leading discipline in the plant sciences coinciding with the start of the new century. For much of the twentieth century, plant geneticists were only successful in delineating putative chromosomal location, function, and changes in genes indirectly through the use of a number of ‘markers’ physically linked to them. These included visible or morphological, cytological, protein, and molecular or DNA markers. Among them, the first DNA marker, the RFLPs, introduced a revolutionary change in plant genetics and breeding in the mid-1980s, mainly because of their infinite number and thus potential to cover maximum chromosomal regions, phenotypic neutrality, absence of epistasis, and codominant nature. An array of other hybridization-based markers PCR-based markers, and markers based on both facilitated construction of genetic linkage maps, mapping of genes controlling simply inherited traits and even gene clusters (QTLs) controlling polygenic traits in a large number of model and crop plants. During this period a number of new mapping populations beyond F2 were utilized and a number of computer programs were developed for map construction, mapping of genes, and for mapping of polygenic clusters or QTLs. Molecular markers were also used in studies of evolution and phylogenetic relationship, genetic diversity, DNA-fingerprinting and map-based cloning. Markers tightly linked to the genes were used in crop improvement employing the so-called marker-assisted selection. These strategies of molecular genetic mapping and molecular breeding made a spectacular impact during the last one and a half decades of the twentieth century. But still they remained ‘indirect’ approaches for elucidation and utilization of plant genomes since much of the chromosomes remained unknown and the complete chemical depiction of them was yet to be unraveled.

Physical mapping of genomes was the obvious consequence that facilitated development of the ‘genomic resources’ including BAC and YAC libraries to develop physical maps in some plant genomes. Subsequently, integrated genetic-physical maps were also developed in many plants. This led to the concept of structural genomics. Later on, emphasis was laid on EST and transcriptome analysis to decipher the function of the active gene sequences leading to another concept defined as functional genomics. The advent of techniques of bacteriophage gene and DNA sequencing in the 1970s was extended to facilitate sequencing of these genomic resources in the last decade of the twentieth century.

As expected, sequencing of chromosomal regions would have led to too much data to store, characterize, and utilize with the-then available computer software could handle. But development of information technology made the life of biologists easier by leading to a swift and sweet marriage of biology and informatics and a new subject was born—bioinformatics.

Thus, evolution of the concepts, strategies and tools of sequencing and bioinformatics reinforced the subject of genomics—structural and functional. Today, genome sequencing has traveled much beyond biology and involves biophysics, biochemistry and bioinformatics!

Thanks to the efforts of both public and private agencies, genome sequencing strategies are evolving very fast, leading to cheaper, quicker and automated techniques right from clone-by-clone and whole-genome shotgun approaches to a succession of second generation sequencing methods. Development of software of different generations facilitated this genome sequencing. At the same time, newer concepts and strategies were emerging to handle sequencing of the complex genomes, particularly the polyploids.

It became a reality to chemically—and so directly—define plant genomes, popularly called whole-genome sequencing or simply genome sequencing.

The history of plant genome sequencing will always cite the sequencing of the genome of the model plant *Arabidopsis thaliana* in 2000 that was followed by sequencing the genome of the crop and model plant rice in 2002. Since then, the number of sequenced genomes of higher plants has been increasing exponentially, mainly due to the development of cheaper and quicker genomic techniques and, most importantly, development of collaborative platforms such as national and international consortia involving partners from public and/or private agencies.

As I write this preface for the first volume of the new series “Compendium of Plant Genomes”, a net search tells me that complete or nearly-complete whole-genome sequencing of 45 crop plants, eight crop and model plants, eight model plants, 15 crop progenitors and relatives, and three basal plants are accomplished, the majority of which are in the public domain. This means that we nowadays know many of our model and crop plants chemically, i.e., directly, and we may depict them and utilize them precisely better than ever. Genome sequencing has covered all groups of crop plants. Hence, information on the precise depiction of plant genomes and the scope of their utilization is growing rapidly every day. However, the information is scattered in research articles and review papers in journals and dedicated web pages of the consortia and databases. There is no compilation of plant genomes and the opportunity of using the information in sequence-assisted breeding or further genomic studies. This is the underlying rationale for starting this book series, with each volume dedicated to a particular plant.

Plant genome science has emerged as an important subject in academia, and the present compendium of plant genomes will be highly useful both to students and teaching faculties. Most importantly, research scientists involved in genomics research will have access to systematic deliberations on the plant genomes of their interest. Elucidation of plant genomes is not onlyof interest for the geneticists and breeders, but also for practitioners of an array of plant science disciplines, such as taxonomy, evolution, cytology, physiology, pathology, entomology, nematology, crop production, bio-chemistry, and obviously bioinformatics. It must be mentioned that information regarding each plant genome is ever-growing. The contents of the volumes of this compendium are therefore focusing on the basic aspects of the genomes and their utility. They include information on the academic and/ or economic importance of the plants, description of their genomes from a molecular genetic and cytogenetic point of view, and the genomic resources developed. Detailed deliberations focus on the background history of the national and international genome initiatives, public and private partners involved, strategies and genomic resources and tools utilized, enumeration on the sequences and their assembly, repetitive sequences, gene annotation, and genome duplication. In addition, synteny with other sequences, comparison of gene families, and, most importantly, potential of the genome sequence information for gene pool characterization through genotyping by sequencing (GBS) and genetic improvement of crop plants have been described. As expected, there is a lot of variation of these topics in the volumes based on the information available on the crop, model, or reference plants.

I must confess that as the series editor it has been a daunting task for me to work on such a huge and broad knowledge base that spans so many diverse plant species. However, pioneering scientists with life-time experience and expertise on the particular crops did excellent jobs editing the respective volumes. I myself have been a small science worker on plant genomes since the mid-1980s and that provided me the opportunity to personally know several stalwarts of plant genomics from all over the globe. Most, if not all, of the volume editors are my longtime friends and colleagues. It has been highly comfortable and enriching for me to work with them on this book series. To be honest, while working on this series I have been and will remain a student first, a science worker second, and a series editor last. And I must express my gratitude to the volume editors and the chapter authors for providing me the opportunity to work with them on this compendium.

I also wish to mention here my thanks and gratitude to the Springer staff, Dr. Christina Eckey and Dr. Jutta Lindenborn in particular, for all their constant and cordial support right from the inception of the idea.

I always had to set aside additional hours to edit books besides my professional and personal commitments—hours I could and should have given to my wife, Phullara, and our kids, Sourav, and Devleena. I must mention that they not only allowed me the freedom to take away those hours from them but also offered their support in the editing job itself. I am really not sure whether my dedication of this compendium to them will suffice to do justice to their sacrifices for the interest of science and the science community.

Kalyani, India Chittaranjan Kole

Preface

The relevance of common bean (*Phaseolus vulgaris* L.) is determined by the fact that it is the most important grain legume for food consumption worldwide and has a role in sustainable agriculture owing to its ability to fix atmospheric nitrogen. It is a staple crop with a major societal importance as main source of proteins and nutrients in developing regions such as Africa and Latin America. Due to its domestication, it can be grown in a wide range of environments, from near sea level up to 3000 m above sea level, in arid, semiarid and tropical environments, and under high or moderate temperatures. Its main consumption is for its edible dry seed (dry beans) or unripe fruit (green beans). This double use determines that modern cultivars have been bred for different fruit characteristics such as those that influence the seed (shape, color, cooking ability, etc.) and the pod (length, tenderness, stringless, etc.) apart from common features as disease resistance, yield, precocity and plant architecture.

Common bean was domesticated in two geographically isolated and differentiated areas, according to multiple lines of evidence, generating two distinguishable gene pools, Mesoamerican and Andean, within a single species. The domestication of two independent lineages implicated a parallel evolution and introgression events with important implications for modern common bean crop improvement. In fact, common bean was likely domesticated concurrently with maize as part of the 'milpa' cropping system (featuring common bean along with maize and squash), which was adopted throughout the Americas. Common bean was introduced in Europe soon in the XVI century through Spanish and Portuguese trade routes, and independently introduced into Africa. Large variation of common bean evolved in Europe as a result of adaptation to new ecological and manmade conditions. In consequence, common bean has four well-identified centers of diversification, two in America, where it was domesticated, one in Europe, and one in Africa, where it was introduced after the discovery of America.

Common bean is also an important species from the scientific point of view. It has been widely used in research and breeding. Both Darwin and Mendel used this species in their research, and it was widely used as material for genetic experimentation and breeding during the early years of the development of genetics, at the beginning of the twentieth century. For example, the demonstration by Johansen that the phenotype is the result of the interaction between the genotype and the environment; or the direct relationship between a quantitative and a qualitative character, seed size and seed coloration of the common bean seed, by Sax. A large corpus of genetic data on common bean has been generated throughout the twentieth century, now widely complemented and surpassed by the genomic information.

The recent release of the Andean and Mesoamerican common bean genomes is enabling a new wave of cutting-edge research, including epigenomics and translatome analyses, in a crop that has fed billions of people for more than 5000 years. Moreover, the comparison of the Andean and the recently released Mesoamerican genome has initially revealed interesting differences. For example, the size and the number of genes in the Mesoamerican genome is smaller than the Andean genome. The availability of the genome sequence has completely changed the paradigm of the species' genetics and genomics.

Genomic information on the other domesticated or wild *Phaseolu*s species range from scarce to null. However, the already sequenced genomes of common bean will help in sequencing and assembling the genomes of other species of the genus. In fact, common bean is one of the five domesticated species of the genus *Phaseolus*, a genus formed by a relatively large number of species with broader or narrower geographical distribution and some of them adapted to particular environments. The other four domesticated species are tepary bean (*P. acutifolius* A. Gray), runner bean (*P. coccineus* L.), lima bean (*P. lunatus* L.) and year-long bean (*P. dumosus* Macfad.), all four originated and domesticated in America. The genomic information from the other *Phaseolus* species will give more complete information on the evolutionary processes associated to domestication and on the adaptive processes to particular environments with special relevance for breeding strategies.

Genomic resources include not only data from the nuclear genome but also from organelle genomes. In particular, mitochondrial and chloroplast genomes gives complementary information on the evolutionary processes within common bean and between species within the genus *Phaseolus*. Genomics data allow knowing the whole set of genes implicated in a metabolic route, including the genes coding for the transcription factors implicated in their expression control, helping in the transcriptomic analyses of their expression under different environments or in different tissues. Likewise, they allow the analysis of the relationships between different metabolic routes and biological processes. A particular set of genes of interest in a crop mainly exploited by its seeds is the set of seed storage proteins, which determine a great part of the nutritive value of the main crop product. New tools will be likely developed from these data such as complete set of molecular markers, microarrays, etc., which will speed and facilitate assisted selection processes.

The present book summarizes data on the economic and scientific relevance of common bean, its relation with other species of the genus *Phaseolus*, and insists on the importance of the domestication events, parallel evolutionary history and further expansion of the cultivated forms. It also describes the numerous genetic and genomic resources available, the identified genes and quantitative trait loci (QTL) identified, as well examples of the study of functionally related genes and future prospects. This book shows that these are exciting times for common bean research in a field with the potential to reduce the threat of food insecurity by releasing crops tolerant to biotic and abiotic stresses, increasing yields and enhancing the nutritional quality of beans.

León, Spain — Marcelino Pérez de la Vega
Pontevedra, Spain — Marta Santalla
London, Canada — Frédéric Marsolais

Contents

Abbreviations

ADP	Andean diversity panel
AFLP	Amplified fragment length polymorphism
AM	Association mapping
BAC	Bacterial artificial chromosome
BBH	Best-bidirectional blast hit
CDS	Coding sequence (Coding DNA sequence)
CGIAR	Consultative Group for International Agricultural Research
CIAT	International Center for Tropical Agriculture (Centro Internacional para Agricultura Tropical)
CMA	Chromomycyn A3
cpSSR	Chloroplast simple sequences repeat
DAF	Days after flowering
DAPI	4′-6-diamidino-2-phenylindole
DDP	Durango Diversity Panel
DRK	Denaturation-renaturation kinetics
ER	Endoplasmic reticulum
EST	Expressed sequence tag
FISH	Fluorescent *in situ* hybridization
GBS	Genotyping by sequencing
GO	Gene ontology
GWAS	Genome wide association study
IGS	Intergenic spacer
IR	Inverted repeat (plastid genome)
ITS	Internal transcribed spacer
KEEG	Kyoto encyclopedia of genes and genomes
KOG	Eukaryotic orthologous groups
LD	Linkage disequilibrium
LG	Linkage group
lncRNA	Long non-coding RNA
LSD	Large single sequence (plastid genome)
MAS	Marker assisted selection
MDP	Mesoamerican Diversity Panel
NCF	Nested chromosome fusion
NGS	Next generation sequencing
P450	Cytochrome P450

PA	Phytic acid
PANTHER	Protein analysis through evolutionary relationships
PBPI	Potential biparental plastid inheritance
PCC	Pearson correlation coefficient
PCR	Polymerase chain reaction
Pfam	Protein families database
PGC	Protein coding genes
QTL	Quantitative trait locus/loci
RAPD	Random amplified polymorphic DNA
RFLP	Restriction fragment length polymorphism
RGA	Resistance gene analog
RIL	Recombinant inbred line
RNAseq	RNA sequencing
RPKM	Reads per kilobase per million mapped reads
SCAR	Sequence characterized amplified region
SNP	Single nucleotide polymorphism
SSC	Short single sequence (plastid genome)
SSR	Simple sequence repeat
TEC	Tissue expression complementarity
UniProt	Protein database
WGD	Whole genome duplication
WGS	Whole genome sequencing

Common Bean: Economic Importance and Relevance to Biological Science Research

1

James R. Myers and Ken Kmiecik

Abstract

Common bean (*Phaseolus vulgaris* L.) is broadly adapted to environments with moderate growing temperatures, about 400 mm of precipitation and a growing season of 60–120 days. The popularity of the crop originates from the fact that it is relatively easy to produce, it is flavorful and versatile, and it is a good source of nutrition. The two major types of common bean are dry edible beans and snap or garden beans. Precise economic valuation of the common bean crop is difficult to obtain on a global scale because other species are often included in the statistical data collected in different countries, but with production of 18.9 million T for all types, it is the most widely produced grain legume and ranked third after soybean and groundnut for oilseed and grain legumes combined. Common bean is produced in both developed and developing countries and is an important source of protein, carbohydrates, some vitamins, and micronutrients. Common bean first became known to the scientific world with the Columbian exchange beginning in 1493, but little is known about the genetic diversity of the early introductions to Europe. Systematic breeding of common bean began in the nineteenth century in Europe and the USA. Common bean was the subject of Mendel's genetics research, was used by Johannsen to investigate quantitative inheritance, and has the distinction of being the first plant species where a quantitative trait locus was identified. Contemporary research on common bean in the recent past has been conducted in about 21 academic disciplines with plant physiology, medicine, microbiology, and food science, resulting in the most publications. Plant breeding, genetics, plant pathology, and

---

J. R. Myers (✉)
Department of Horticulture, Oregon State University, Corvallis, OR 97331, USA
e-mail: james.myers@oregonstate.edu

K. Kmiecik
7202 Portage Rd, De Forest, WI 53532, USA
e-mail: kakmiecik@sbcglobal.net

M. Pérez de la Vega et al. (eds.), *The Common Bean Genome*,
Compendium of Plant Genomes, https://doi.org/10.1007/978-3-319-63526-2_1

genomics and bioinformatics are generally less well represented, but this may change as more genomics studies are conducted. The special traits of common bean that have driven most research are the seed storage proteins, the symbiotic relationship with rhizobium species, the history of plant domestication, and the architecture of genetic diversity within the species.

Keywords
Grain legume · Dry bean · Snap bean · Production
Consumption · Genetics · Citation analysis

1.1 Introduction

Originating in the New World and widely dispersed because of its broad adaptation, the common bean (*P. vulgaris* L.) has become an important grain and vegetable legume on a global scale. It performs best in moderate growing temperatures (>10 °C and <30 °C) with about 400 mm of precipitation during the growing season. Common bean is found throughout temperate growing regions where the season permits 60–120 days of frost-free growth as well as in the tropical highlands with growing temperatures <30 °C. The popularity of the crop originates from the fact that it is relatively easy to produce, it is flavorful and versatile in its preparation, and it is a good source of nutrition. Uncooked dry bean contains approximately 22% protein, several micronutrients (Ca, Fe, Mg, P, and K), complex carbohydrates (62%), soluble fiber (15%) and is a significant source of folate (USDA 2015). As a vegetable, the immature pods have high moisture content, with raw pods containing about 1.9% protein and 7% carbohydrate, and significant quantities of vitamin C, carotenoids, and vitamin K, which dry beans lack. Common bean is particularly important to the developing world in providing a source of protein, calories, and trace nutrients to individuals who cannot afford more expensive sources of nutrition.

1.1.1 Kinds and Uses of Common Bean

The term bean is widely applied to many different species of large seeded legumes (and some non-legumes), but those English terms that apply specifically to common bean include dry, kidney, French, navy, pinto, garden, snap, green, wax, and string. These names often refer to specific types or uses of common bean (Table 1.1 and Fig. 1.1). The two major types of contemporary agriculture are dry edible beans where the mature seed is harvested and prepared in main dishes and garden beans—the vegetable form where immature pods are cooked and eaten (Table 1.2).

Considerable evidence from several disciplines points toward the independent domestication of common bean occurring in at least two regions (Gepts et al. 1986; Koenig and Gepts 1989; Gepts and Debouck 1991), which are generally termed the Mesoamerican and Andean centers of domestication. Within each of these centers, distinctive races and market classes are found (Table 1.3). The races of common bean (Singh et al. 1991) are generally defined by certain morphological and biochemical characteristics as well as by traditional regions of use. Beans of the Mesoamerican center are classified into three races, and those of the Andean center of domestication are also subdivided into three races (Table 1.3). Within races, dry beans are

Table 1.1 Some of the terms used for different types of *Phaseolus vulgaris* in the scientific literature and statistical databases

Name	Usage/definition	FAO classification
Dry grain		
Common	Term refers generally to all forms of *Phaseolus vulgaris* and is a direct derivative of the Latin name	None
Dry	Mature grain where the seed is boiled to prepare for consumption	Dry bean; in some countries, FAO data may refer to other species as dry bean (including other *Phaseolus* beans such as *P. coccineus, P. lunatus* and *P. acutifolius* as well as *Vigna angularis, V. aureus, V. calcaratus, V. mungo, V. radiata, and V. aconitifolia*, but excluding cowpea or *V. unguiculata*)
Kidney	Name appears in the scientific literature as a general reference to common bean, but specifically refers to a market class of dry bean, usually large seeded, reniform in shape, and red or white in color	None
French	A term commonly applied to *P. vulgaris* beans in Europe and in the scientific literature. The term is used for both dry and snap beans	None
Navy	Market class of dry bean, small seeded and white in color	None
Popping or Nuña	Dry bean that is cooked using dry heat and where the cotyledons expand and soften	None
Vegetable		
Shell or Fresh	Seeds are consumed when mature but still high moisture	Green bean and may include *Vigna* species in some countries
Snap (green, wax) or garden	Green pods at various stages of maturity are cooked and consumed. Snap beans lack pod wall fiber and suture fiber or strings	String bean and may include *Vigna* species in some countries
String bean	Similar to snap beans but pod suture string is present and must be removed by hand	No distinction is made between types with or without pod suture fiber

further separated into market classes, which are generally defined by seed characteristics, but have superimposed upon them their genetic heritage and its influence in adaptation and various developmental characteristics. A classification of dry beans for North America is shown in Table 1.3. It should be noted that the relationships among center of domestication, race, and market class are based on traditional information, but contemporary plant breeding has blurred the boundaries among groups, and what have once been considered good indicators of centers of domestication (e.g., phaseolin seed storage protein) may no longer show absolute associations.

Dry beans (Fig. 1.1a and b) are usually prepared by soaking in water to imbibe the seed followed by cooking in a water-based broth either boiling or using a pressure cooker to shorten preparation time. They may be eaten directly in this fashion, but are more often cooked or combined with other more savory ingredients. They may be reprocessed (such as with refried beans) into new forms. Plant type for dry edible beans ranges from determinate bush to indeterminate upright types and to indeterminate vining non-climbers to indeterminate climbers.

Another type of dry bean is nuña or popping bean (Fig. 1.1e). The popping trait is limited to a

Fig. 1.1 Examples of various forms of common bean used for food around the world. **a** Kablanketi type of dry bean used in eastern and southern Africa; **b** large red bean typical of that consumed worldwide; **c** fresh snap bean pods consumed as a vegetable; **d** snap bean seeds showing distinctive cylindrical shape; **e** Nuña beans after being popped for consumption; **f** dry bean leaves picked for use as a vegetable in southern Tanzania

Table 1.2 Pod and seed characteristics that influence preparation and consumption of the major common bean types

Type	Pod wall fiber	Pod wall thickness	Pod suture fiber (strings)	Method of cooking	Seed shape
Dry	Present	Thin	Present	Moist heat	Variable
Popping	Present	Thin	Present	Dry heat	Round
Shell	Present	Thin	Present	Moist heat	Variable
Romano	Absent	Thin	Generally absent	Moist heat	Variable
String	Absent	Thick	Present	Moist heat	Variable
Snap	Absent	Thick	Absent	Moist heat	Cylindrical

Types are described in Table 1.1

Table 1.3 Races and selected market classes of common bean [Modified from (Singh et al. 1991)]

Race	Seed size	Phaseolin	Market classes
Mesoamerican domestication center			
Mesoamerica	Small	S	Navy, Black, Small Red
Durango	Medium	S	Pinto, Bayo, Red Mexican, Great Northern, Pink, Blue Lake type snap beans
Jalisco	Medium	S	Flor de Mayo, Apetito, Flor de Junio
Andean domestication center			
Nueva Granada	Medium–large	T	Light Red Kidney, Dark Red Kidney, White Kidney, Yellow-eye, Peruano, Tendergreen type snap beans
Chile	Medium–large	C, H	Cranberry, Romano type snap beans
Peru	Medium–large	T, C, H	Nuña, Poroto

Examples of representative market classes of North America, including Mexico

narrow group of dry edible beans originating in the Andean areas of South America. At high altitudes, cooking dry beans by boiling becomes problematic because of the lowered boiling point of water and the extra time required to cook the food. Popping beans are heated directly and require less time to expand or puff into a soft edible state, thereby requiring less cooking fuel.

Vegetable use of common beans includes immature pods (Fig. 1.1c), high moisture seed, and leaves (Fig. 1.1f) as greens. Use of immature pods is common in North America, Europe, Middle East, Africa, and throughout Asia including India, Indonesia, Thailand, Philippines, China, and Japan. Preparation includes boiling, steaming, and frying. Immature pods may have suture strings or be stringless; some wall fiber or none is present depending on use such as processing, market garden, or home garden. The plant type encompassed in garden beans includes indeterminate climbing (pole) types or half runner prostrate types, and fully determinate (bush) plant habits.

The snap bean (Fig. 1.1c and d) is mainly consumed in developed countries. Snap beans have been selected for reduced fiber in the green pod. In its more ancestral form (typified by Romano beans), fiber is reduced in the pod walls only; pod walls are thin, and suture strings are present (Table 1.2). String beans in addition to having low pod fiber have a second trait for fleshy pod walls that give the pod an oval to round cross-sectional shape. However, suture strings may still be present. In the modern forms of snap beans, the pod suture string is also lacking (Tables 1.1 and 1.2).

Use of high moisture seed shelled in the green stage is less common. It includes the French Flageolet types, large Horticultural or Cranberry types in the USA, and similar types in Italy such as Borlotto. Frequently, large seeded types are also shelled when used in high altitudes in the Andes, rather than used dry. Many large seeded dry edible types grown in parts of Asia and Africa may be used as immature pods or shelled, in addition to dry seed.

The development of beans as vegetable is unclear, but it is evident that Andean and Mesoamerican germplasm introduced into Europe is part of the basis of modern garden bean cultivars. The presence of S, T, and C phaseolin types (Gepts and Bliss 1986; Angioi et al. 2010) in pre-1950s' snap bean lines suggests a complex history.

1.2 Value, Production, and Consumption

Exact valuation of the common bean crop is difficult to obtain on a global level. This is because the main source of worldwide agricultural statistics (FAOSTAT 2015) does not distinguish among species of grain legumes when collecting statistical information and more than one species of bean may be aggregated (Table 1.1). This is especially true of India and China where *Vigna* spp. may account for 93 and 56%, respectively, of dry bean use in these regions (Akibode and Maredia 2011). Estimates for the amount of dry beans as classified by FAO that are actually *P. vulgaris* for Asia range from 17% for 1998 data (Singh 1999) to 9% for 2006–2008 data (Akibode and Maredia 2011). According to FAO, the estimates exclude cowpea (Table 1.1). Similarly, green or shell bean and string or snap bean categories in FAOSTAT may include various *Vigna* species (Table 1.1). For the latter categories, even less is known about the percentages that are actually common bean. Here, we report figures adjusted using the more recent 9% estimate of Akibode and Maredia (2011). We applied this adjustment to both dry and shell beans, but not snap beans. Our rationale was that many of the same species are used as dual purpose dry and as shell-outs and that proportional usage should be similar. With snap beans, the other main edible pod bean besides common bean is yard-long bean or asparagus bean (*Vigna unguiculata*), which is the vegetable form of cowpea and generally of minor importance in the countries included in this analysis. Additional caveats to the data are that these numbers do not account for non-commercial production in small holder enterprises and that in many parts of the tropics, common bean is intercropped, which may lead to an overestimation of area and an underestimation of yield (Akibode and Maredia 2011).

Worldwide, approximately 18.9 million T (all types combined) are produced (Table 1.4 and Fig. 1.2). At 13.9 million T, dry bean production is considerably greater than shell (3.5 million T) and snap (1.6 million T) production. Hectares for dry beans show an upward trend over time, while that for shell and snap remain essentially unchanged (Table 1.4 and Fig. 1.3). Yields for dry and snap bean have remained essentially unchanged from 1992 to 2012, but those for shell beans have increased sharply. The increase in area for dry beans and the increase in yield for shell beans have increased production in these types over the past two decades, whereas snap bean production has remained essentially unchanged. On a megaregional basis, Africa and the Americas are the largest producers of dry beans, accounting for 66% of production (Fig. 1.2).

1.3 Importance of Common Bean in Science

1.3.1 Common Bean Becomes Known to Science

Common bean became known to science following the Columbian exchange. What was most likely common bean (and not a related species) was described from the first Columbian expedition (Markham 2010) and were introduced into Europe soon thereafter. What we know of the

Table 1.4 World area (Ha), yield (T ha^{-1}), and production (T) of dry, shell, and snap beans from 1992 to 2012

Bean type	Year				
	1992	1997	2002	2007	2012
Area harvested (1000 ha)					
Dry bean (adjusted)	13,510	13,888	15,360	15,156	16,230
Shell (adjusted)	350	384	395	394	377
Snap	217	217	223	239	182
Yield (T ha^{-1})					
Dry bean	0.65	0.64	0.72	0.73	0.82
Shell	6.51	6.92	8.08	12.34	13.51
Snap	6.82	7.82	8.56	8.59	8.47
Production (1000 T)					
Dry bean (adjusted)	9413	10,070	12,342	12,135	13,904
Shell (adjusted)	2269	2546	2850	3472	3484
Snap	1478	1695	1914	2052	1575

Data source FAOSTAT (2015) with dry and shell bean area and production adjusted to remove non-common bean data from the figures

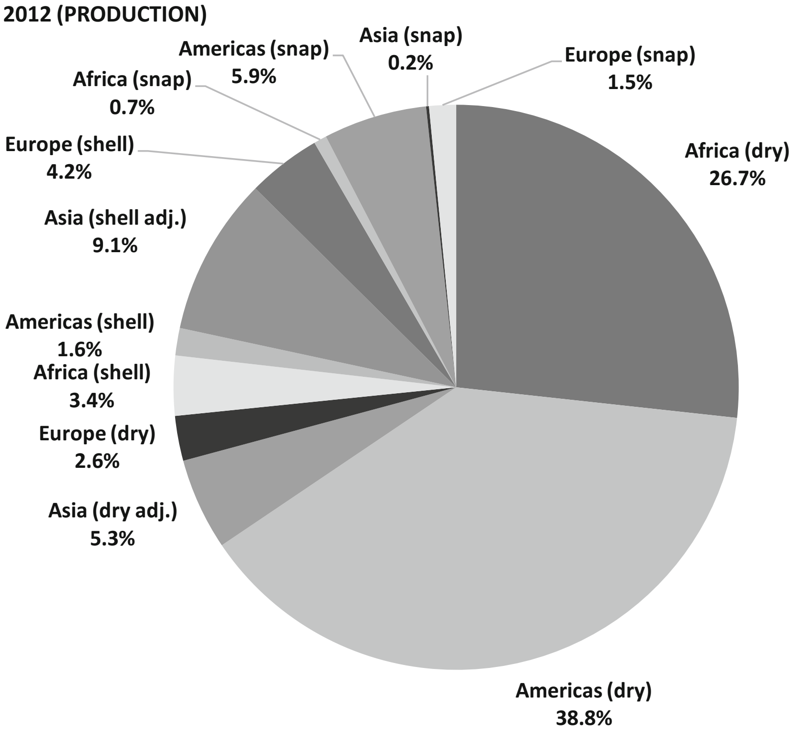

Fig. 1.2 Production of dry, shell, and snap beans in 2012 (FAOSTAT 2015)

early history of beans comes from paintings and illustrations in buildings, prayer books, and herbals. The oldest image of what is thought to be a bean plant is found in Livre d'Heures d'Anne de Bretagne which was illustrated between 1505 and 1508. Camus (1894) cataloged the plants in the prayer book and concluded that one represented *P. vulgaris*, but more recent examination of the image in question raises doubt as the image is not morphologically accurate

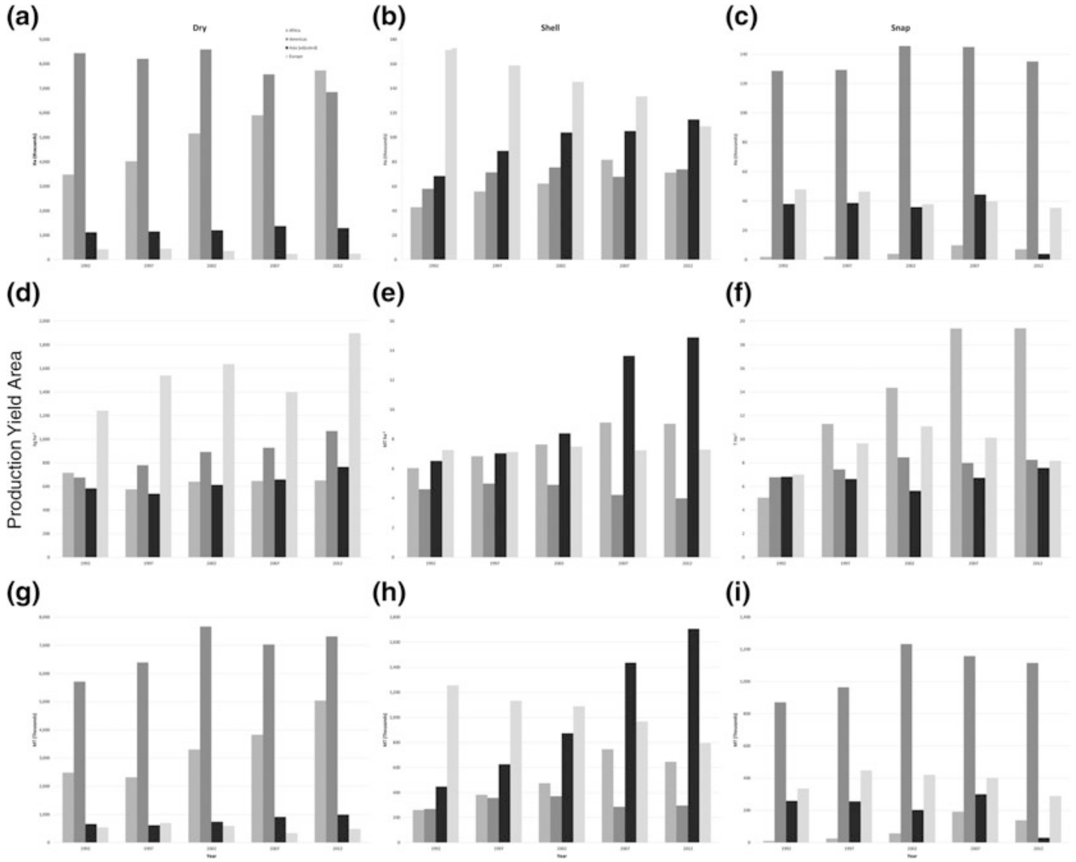

Fig. 1.3 Area (Ha), yield (Kg ha^{-1} or T ha^{-1}), and production (T) of common bean in four megaregions from 1992 to 2012. Data for Asia adjusted to remove non-common bean data from these figures. Data from FAOSTAT (Jan, 2015). **a** Dry bean area; **b** shell bean area; **c** snap bean area; **d** dry bean yield; **e** shell bean yield; **f** snap bean yield; **g** dry bean production; **h** shell bean production; and **i** snap bean production

(unpublished observations). In a loggia of the Villa Farnesina in Rome, festoons containing botanically accurate images of over 160 plant species were painted between 1515 and 1518 (Janick and Caneva 2005). These festoons have some of the earliest known images of New World crops including maize (Janick and Caneva 2005) and squash (Janick and Paris 2005). Bean pods are also illustrated and have been reported to be those of *P. vulgaris* (Caneva 1992). It is not until Fuchs' (1542) Di Historias Stirpium that we have unequivocal images of common bean. Fuchs herbal was followed by others (reviewed in Hedrick 1919) with similar images of Welsh (foreign) bean. These images provide a glimpse into what was being grown in Europe, but they do not show the diversity of germplasm present or any changes that may have been happening. The early science concerning common bean was focused on plant taxonomy, and little is known about the diversification of the crop in Europe during that time period. Various forms of evidence point toward the derivation of snap beans during the sixteenth–eighteenth centuries. Contemporary Native American varieties show little evidence that they were selected for snap bean traits, and the bulk of the evidence suggests that they were derived in Europe, before being disseminated back to the USA and elsewhere in the world (Myers and Baggett 1999).

1.3.2 Early Plant Breeding and Classical Genetics in Europe/USA

Efforts to systematically breed common bean are documented beginning in the nineteenth century in Europe and the Americas (Wade 1937), but scientific research did not begin in earnest until the beginning of the twentieth century. *Phaseolus* beans were the basis for several important studies of classical genetics. Mendel observed partial confirmation of his results in pea for three traits in intraspecific *P. vulgaris* crosses, but found complexities in the inheritance of flower color from an interspecific cross to *Phaseolus coccineus* (Olby 1985). Following the rediscovery of Mendel's work, researchers applied qualitative genetic analysis to the study of various traits in common bean. One of the earliest works was by Emerson (Emerson 1904) where he examined inheritance of traits ranging from plant habit to flower and seed color to pod characteristics. A good review of the bean genetics literature from the early twenteith century can be found in Kooiman (1931). A plant geneticist who contributed immensely to the qualitative genetics of common bean on many different traits was Herbert Lamprecht. Much of what we know about the genetics of seed coat colors comes from his work (Lamprecht 1932a, b, 1933, 1934a, b, 1935, 1936, 1939, 1940a, b, c, 1951, 1952, 1955, 1960, 1961, 1964) along with those of Dr. Prakken (1934, 1940, 1972a, b, 1974) and more recently by Bassett (2007). A list of genes is maintained by the Bean Improvement Cooperative Genetics Committee (BIC 2015) and includes a comprehensive set of references to qualitatively inherited genes.

Common bean has the distinction of being one of the species by which the principles of quantitative genetics were first elucidated. Johannsen (1903) investigated inheritance of several quantitative traits, including seed weight. He found that progeny of lines maintained by self-pollination over several generations tended to more closely resemble ancestral types of that lineage and were distinctly different from one another. He also argued that differences within a pure line were the result of environmental variation and not due to genetic causes (Carlson 2004). Based on these results, Johannsen formulated the concept of genotype versus phenotype.

The first report of linkage between qualitative and quantitative traits was that of Sax (1923) where he established the methodology for contemporary approaches to mapping quantitative trait loci (QTL). At the time of its discovery, it was regarded as a mysterious phenomenon whereby seed size (a quantitative trait) was found to be associated with seed color (a qualitative trait). Researchers were aware of the concept of linkage among qualitatively inherited genes and were constructing some of the first linkage maps, but until Fisher's landmark paper (Fisher 1918), the genetic control of quantitative traits was thought to be fundamentally different from qualitative traits. Modern genetic approaches have further elucidated the QTL that Sax discovered. The seed color gene in question is *P* (Emerson 1909), which controls expression of flavonoids in the seed testa and other plant parts. A gene underlying the QTL for seed size may be the phaseolin (*Phs*) seed storage protein (Johnson et al. 1996) where seeds with S phaseolin tend to be smaller than those with T or C phaseolin (Johnson et al. 1996). It is not known whether *Phs* is directly responsible for seed size differences, or whether other tightly linked loci might be involved. While *Phs* has been sequenced, the definitive experiment to re-engineer and transform back into the species is difficult to perform in common bean. *P* has been mapped to about 10 cM from *Phs* (Johnson et al. 1996). In preliminary work, a candidate gene for *P* has been identified (Mafi-Moghaddam et al. 2014).

1.3.3 Contemporary Research with Common Bean

To better understand where common bean has featured most prominently in contemporary research, a citation analysis was conducted in Google Scholar (GS) and Web of Science (WoS) databases using the search terms common, kidney, dry, French, green, and snap with bean and separately *P. vulgaris* where these words were found in the title or abstract of the paper. The two databases provide different aspects of the data. While WoS focuses on a set of accredited, refereed journals, GS counts citations from journals not accredited by WoS, non-refereed journals, and other gray literature. As a result, GS's citation counts are almost always higher than WoS, and GS can find important and highly cited works that were not published in refereed journals. Citations were reviewed and were classified by academic discipline and research topic. In addition to citation counts, we also calculated average citations per year as means of identifying those areas where current research is quite active, but may not yet have high total numbers of citations (particularly true for the field of genomics). WoS will also identify highly cited papers, and these were obtained for the search terms common bean and *P. vulgaris*. One caveat about using citation databases is as follows: citation counts are constantly increasing, and as such, the numbers reported in this chapter represent a snapshot in time as of March 2015.

Papers obtained from Google Scholar searches were ranked by number of citations per paper, and we chose an arbitrary cutoff of 50 citations or more in Google Scholar which resulted in 397 papers being retained. These were placed into 21 academic disciplines, some with many papers and others with only one or two (Table 1.5). Plant physiology was the largest category followed by plant genetics and plant pathology. Papers involving medicine (neurology, immunology, hematology, clinical nutrition, gastroenterology, and oncology), microbiology, and food science fell into the middle range, whereas plant breeding, molecular biology, and biotechnology comprised the low end. Considerable variation for most highly cited papers by academic discipline was observed, with over a thousand citations for a paper in neurology (Table 1.5). Other fields with relatively high citation rates included plant genetics, plant physiology, immunology, plant pathology, microbiology, and plant breeding.

The 397 papers were further reduced to a set of 46 (Table 1.6) by choosing the top three by citation count within an academic discipline and research area. From these data, it is apparent that seed storage proteins of common bean have played a significant role in research in several unrelated disciplines. These include the biotechnology (bruchid resistance), the medical sciences (neurology, oncology, immunology, and clinical nutrition), entomology (bruchid resistance), plant genetics (especially plant domestication and genetic diversity), and plant physiology. The most highly cited paper from the Annual Report of the Bean Improvement Cooperative (52 citations) concerns standardization of linkage groups and chromosomes for common bean (Table 1.7). The papers identified in WoS as those representing areas of increased activity are a diverse set, ranging from studies on isotopic diversity in Mexico, to genetic diversity studies, to analysis of bean growth exposed to radiation (Table 1.7). Interestingly, while genomics and bioinformatics of common bean is an extremely active area of research, WoS has not yet identified this as an active area of inquiry.

1.3.4 Important Programs and Centers Supporting Common Bean Research and Genetics Predating the Genomics Era

Several programs which have had a profound impact on common bean genetics and breeding are summarized below.

The International Center for Tropical Agriculture (CIAT) based in Cali, Colombia, has been a catalyst for both basic and applied research on common bean. The center has the most extensive germplasm collection of *Phaseolus* species, and researchers at the institution have helped establish much of what we know of species relationships within the genus, as well as providing the foundation for understanding the domestication and dissemination of common bean. Researchers at CIAT helped establish some

Table 1.5 Disciplines with papers focusing on common bean research, total number of papers by discipline with more than 50 citations per paper, and number of citations for the most highly cited paper, by discipline

Research area	No. papers	No. citations for most highly cited paper
Biotechnology	9	318
Clinical nutrition	3	121
Crop ecology	1	290
Entomology	5	181
Food science	20	293
Gastroenterology	3	91
Genomics	4	147
Geography	1	227
Hematology	1	111
Immunology	4	502
Microbiology	25	488
Molecular biology	10	179
Neurology	38	1093
Oncology	2	185
Plant anatomy	5	174
Plant breeding	11	445
Plant genetics	61	688
Plant pathology	57	489
Plant physiology	134	575
Soil science	2	108
Weed science	1	103
Total no. papers	397	–
Mean no. citations	–	324

of the first molecular genetics efforts in bean and developed an extensive library of simple sequence repeat (SSR) markers that provided the backbone for efforts to understand genetic diversity, create linkage maps, and map QTL. CIAT researchers have examined host–pathogen relationships for several diseases of common bean that are especially problematic in the tropics. A program to develop germplasm with enhanced micronutrient content had led to a better understanding of genetic control. CIAT researchers have developed and released germplasm and cultivars that are used globally, especially in the tropics either directly by farmers or by national and regional breeding programs.

United Stated Agency for International Development (USAID) has funded common bean research for three decades, originally through the Bean/Cowpea Collaborative Research Support Program (Bean/Cowpea CRSP) and subsequently via Feed the Future Legume Innovation Lab (FtF LIL). The basic model has been pairing of US universities and research institutions with national programs and universities in developing countries in South and Central America, Caribbean, and sub-Saharan Africa to provide a multidisciplinary approach to increasing bean productivity in the developing world. Funds have not only supported training, infrastructure development, and crop improvement activities in developing countries but have had a profound impact on breeding and genetics programs in the USA. Many researchers from developing countries have acquired graduate degrees from US universities while working in bean breeding programs in the USA. This has facilitated research on a number of problems of significance to bean production in developing

Table 1.6 Most frequently cited papers on common bean listed by research area and ranked by Google Scholar

Article title	No. of citations[a]		Research topic[b]	Reference
	GS[c]	WoS		
Biotechnology				
Transgenic pea seeds expressing the α-amylase inhibitor of the common bean are resistant to bruchid beetles	318	215	Transformation (bruchids)	(Shade et al. 1994)
Regeneration in *Phaseolus vulgaris* L.: High-frequency induction of direct shoot formation in intact seedlings by N6-benzylaminopurine and thidiazuron	216	152	Tissue culture (regeneration)	(Malik and Saxena 1992)
Inheritance of foreign genes in transgenic bean (*Phaseolus vulgaris* L.) co-transformed via particle bombardment	177	97	Transformation (virus)	(Aragão et al. 1996)
Clinical nutrition				
A dietary supplement containing standardized *Phaseolus vulgaris* extract influences body composition of overweight men and women	121	–	Phytohemagglutinin (obesity)	(Celleno et al. 2007)
Crop ecology				
Studies on mixtures of maize and beans (*Phaseolus vulgaris*) with particular reference to plant population	290	98	Intercropping	(Willey and Osiru 1972)
Entomology				
Fumigant toxic activity and reproductive inhibition induced by monoterpenes on *Acanthoscelides obtectus* (Say) (Coleoptera), a bruchid of kidney bean (*Phaseolus vulgaris* L.)	181	125	Bruchid resistance (monoterpenes)	(Regnault-et al. 1995)
Growth Inhibitory effects of an α-amylase inhibitor from the kidney bean, *Phaseolus vulgaris* (L.), on three species of bruchids (Coleoptera: Bruchidae)	142	108	Bruchid resistance (Seed storage protein)	(Ishimoto and Kitamura 1989)
Resistance to the bean weevil and the Mexican bean weevil (Coleoptera: Bruchidae) in non-cultivated common bean accessions	124	58	Bruchid resistance (Seed storage protein)	(Schoonhoven et al. 1983)
Food science				
Functional properties of the great northern bean (*Phaseolus vulgaris* L.) proteins: emulsion, foaming, viscosity, and gelation properties	293	205	Functional properties (starch)	(Sathe and Salunkhe 1981)
Chemical composition, dietary fiber and resistant starch contents of raw and cooked pea, common bean, chickpea, and lentil legumes	200	94	Functional properties (starch)	(de Almeida Costa et al. 2006)
Hard-to-cook phenomenon in common beans—a review	184	120	Functional properties (hard-to-cook)	(Reyes-Moreno and Paredes-López 1993)
Genomics				
Sequencing and analysis of common bean ESTs. Building a foundation for functional genomics	132	98	Biochemical markers	(Ramírez et al. 2005)
Legumes as a model plant family. Genomics for food and feed report of the cross-legume advances through genomics conference	147	90	Review	(Gepts et al. 2005)

(continued)

Table 1.6 (continued)

Article title	No. of citations[a]		Research topic[b]	Reference
	GS[c]	WoS		
Geography				
Atlas of common bean (*Phaseolus vulgaris* L.) production in Africa	227	–	Production (statistics)	(Kirkby et al. 1998)
Hematology				
A method for the rapid separation of leukocytes and nucleated erythrocytes from blood or marrow with a phytohemagglutinin from red beans (*Phaseolus vulgaris*)	111	144	Phytohemagglutinin	(Li and Osgood 1949)
Immunology				
Characterization of the structural determinants required for the high-affinity interaction of asparagine-linked oligosaccharides with immobilized Phaseolus vulgaris leukoagglutinating and erythroagglutinating lectins	502	467	Phytohemagglutinin	(Cummings and Kornfeld 1982)
A mouse lymphoma cell line resistant to the leukoagglutinating lectin from *Phaseolus vulgaris* is deficient in UDP-GlcNAc: alpha-D-mannoside beta 1, 6 N-acetylglucosaminyltransferase	166	167	Phytohemagglutinin	(Cummings et al. 1982)
Mitogenic leukoagglutinin from *Phaseolus vulgaris* binds to a pentasaccharide unit in N-acetyllactosamine-type glycoprotein glycans	122	146	Phytohemagglutinin	(Hammarström et al. 1982)
Microbiology				
Rhizobium tropici, a novel species, nodulating *Phaseolus vulgaris* L. beans and *Leucaena* sp. trees	488	400	Host-microbe interactions (BNF)	(Martínez-Romero et al. 1991)
Biological nitrogen fixation by common beans (*Phaseolus vulgaris* L.) increases with bio-char additions	367	202	Host-microbe interactions (BNF)	(Rondon et al. 2007)
Classification of rhizobia based on nodC and nifH gene analysis reveals a close phylogenetic relationship among *Phaseolus vulgaris* symbionts	323	263	Host-microbe interactions (BNF)	(Laguerre et al. 2001)
Molecular biology				
Regulation of beta-glucuronidase expression in transgenic tobacco plants by an A/T-rich, cis-acting sequence found upstream of a French bean beta-phaseolin gene	259	179	Seed storage protein (phaseolin)	(Bustos et al. 1989)
The glycosylated seed storage proteins of *Glycine max* and *Phaseolus vulgaris*. Structural homologies of genes and proteins	224	160	Seed storage protein (phaseolin)	(Doyle et al. 1986)
Complete nucleotide sequence of a French bean storage protein gene: Phaseolin	210	170	Seed storage protein (phaseolin)	(Slightom et al. 1983)
Neurology				
Topographical organization of the efferent projections of the medial prefrontal cortex in the rat: an anterograde tract-tracing study with *Phaseolus vulgaris* leucoagglutinin	1093	990	Anterograde tracing	(Sesack et al. 1989)

(continued)

Table 1.6 (continued)

Article title	No. of citations[a]		Research topic[b]	Reference
	GS[c]	WoS		
An anterograde neuroanatomical tracing method that shows the detailed morphology of neurons, their axons, and terminals—immunohistochemical localization of an axonally transported plant lectin, *Phaseolus vulgaris* leukoagglutinin (pha-l)	939	1015	Anterograde tracing	(Gerfen and Sawchenko 1984)
Organization of the projections from the subiculum to the ventral striatum in the rat. A study using anterograde transport of *Phaseolus vulgaris* leucoagglutinin	630	587	Anterograde tracing	(Groenewegen et al. 1987)
Oncology				
The binding of kidney bean phytohemagglutinin by Ehrlich ascites carcinoma	185	228	Phytohemagglutinin	(Steck and Wallach 1965)
Consumption of black beans and navy beans (*Phaseolus vulgaris*) reduced azoxymethane-induced colon cancer in rats	148	64	Diet (cancer)	(Bennink 2002)
Plant anatomy				
Chromium VI induced structural and ultrastructural changes in bush bean plants (*Phaseolus vulgaris* L.)	174	99	Abiotic stress (Cr)	(Vazquez, Poschenrieder, and Barcelo 1987)
Structural and ultrastructural disorders in cadmium-treated bush bean plants (*Phaseolus vulgaris* L.)	169	118	Abiotic stress (Cd)	(Barcelo et al. 1988)
Leaf surface and histological perturbations of leaves of *Phaseolus vulgaris* and *Helianthus annuus* after exposure to simulated acid rain	122	106	Abiotic stress (pH)	(Evans et al. 1977)
Plant breeding				
Basis of yield component compensation in crop plants with special reference to the field bean, *Phaseolus vulgaris*	445	324	Yield (yield components)	(Adams 1967)
Broadening the genetic base of common bean cultivars	266	143	Genetic diversity	(Singh 2001)
Common bean breeding for resistance against biotic and abiotic stresses: from classical to MAS breeding	257	149	Abiotic/biotic stress (biochemical markers)	(Miklas et al. 2006)
Plant genetics				
The association of size differences with seed coat pattern and pigmentation in *Phaseolus vulgaris*	688	–	Quantitative genetics (QTL)	(Sax 1923)
Races of common bean (*Phaseolus vulgaris*, Fabaceae)	574	375	Domestication (genetic diversity)	(Singh et al. 1991)
Phaseolin-protein variability in wild forms and landraces of the common bean (*Phaseolus vulgaris*): evidence for multiple centers of domestication	410	303	Domestication (seed storage protein)	(Gepts et al. 1986)
Plant pathology				
Volatile products of the lipoxygenase pathway evolved from *Phaseolus vulgaris* (L.) leaves inoculated with *Pseudomonas syringae* pv *phaseolicola*	489	376	Plant defense (*Pseudomonas*)	(Croft et al. 1993)

(continued)