UNIVERSITEIT GENT

Faculteit Bio-ingenieurswetenschappen

Academiejaar 2011 – 2012

Breeding of food grade soybeans

Mélanie Van Schoote

Promotoren: Prof. dr. ir. Dirk Reheul
Ao.Univ.Prof. Dipl.-Ing. Dr.nat.techn. Johann Vollmann

Tutor: Prof. dr. ir. Dirk Reheul

Masterproef voorgedragen tot het behalen van de graad van
Master in de bio-ingenieurswetenschappen: cel- en genbiotechnologie

Preface

This thesis is the result of tight cooperation between Ghent University (Belgium) and the University of Natural Resources and Life Sciences, Vienna (Austria). This cooperation was made possible by the ERASMUS exchange program. All the practical work of the thesis was performed at the Department of Plant Breeding in Tulln an der Donau (Austria). The practical work as well as the writing of the thesis was performed with strong coordination of professor Johann Vollmann (University of Natural Resources and Life Sciences, Vienna). I would like to thank him specially as he gave me the possibility to perform my master thesis at his department and provided me of an uncountable help in order to achieve it. Furthermore, I would also like to give a special thanks to professor Dirk Reheul (Ghent University) for accepting to coordinate the thesis in Belgium and for his great help.

This thesis is an application of my favorite research field, namely plant breeding, in a context I particularly appreciate, which is the use of soybean in human nutrition. The current global consumption of animal products is one of the major cause of the environmental issues we face nowadays. The production of animal products is mainly based on soybean, as soybean meal is one of the main protein source in animal feed. Paradoxically, soybean can be used as protein source in human nutrition. Hence, the consumption of soybean directly as food, instead of using it as animal feed, could help to diminish the consumption of animal products. Breeding of food grade soybean cultivars is in this context essential. Therefore, I consider this thesis as a small participation in a greater cause.

Finally, I would like to thank: Roman Tumpold, Leo Vansteenkiste, Jan Van Deun, Geert Gerard, Hilde Vandecasteele, Dominique Langouche, Ulrike Piringer and the members of the jury: professor Monica Höfte and professor Heinrich Grausgruber.

Contents

List of abbreviations

11S/7S	The ratio of the 11S fraction to the 7S fraction
AGES	Österreichische Agentur für Gesundheit und Ernährungssicherheit
ANOVA	Analysis of variance
BOKU	University of Natural and Life Sciences, Vienna
DHA	Docosahexaenoic acid
EU	European Union
FAO	Food and Agriculture Organization
GHG	Green house gas
GM	Genetically modified
HVP	Hydrolyzed vegetable protein
IFS	Isoflavone Synthase
IU	International Unit
LS-mean	Least square mean
LSD	Least significant difference
NIRS	Near-infrared spectroscopy
RQn	Research question number *n*
USA	United States of America
WHO	World Health Organization

Summary

The legume crop soybean [*Glycine max* (L.) Merrill] is widely grown over the world. Its composition (approximately 40% protein and 20% oil) permits a broad range of applications. One major application of soybean is its consumption as food product. The growing interest in soybean based foods has resulted in the development of specialized markets in several regions of the world including in Austria. To respond to the demands of these markets, breeding for food grade soybeans is essential. The major breeding targets for food grade soybeans are high protein and sucrose content. The aim of the GPX_SATO experiment was to enhance the food-grade characteristics of Austrian adapted cultivars. Therefore, four crosses between an adapted mother-genotype with normal protein content and a high-protein donor were made. Selection based on protein content was performed per cross on 200 $F_{3:5}$-lines: the 24 lines with the lowest protein content and the 24 lines with the highest protein content were selected in each cross. Analysis of a field experiment of the F_6-generation was performed. It seemed that the two donors, GF4X-21-5-2 and Vinton 81, used for the crosses had good breeding values with respect to food-grade traits. The crosses performed with Vinton 81 resulted in populations where improved genotypes for both protein and sucrose content can possibly be found, due to an only moderately negative correlation between protein and sucrose content. Concretely, for one of the crosses with Vinton 81, one genotype with both improved protein and sucrose content compared to its adapted parent was found. For the other cross performed with Vinton 81, it appeared that selection based on enhanced protein content was not a recommendable strategy for finding improved genotypes, possibly, selection based on increased sucrose content could give better results. Furthermore it seemed that for the populations originating from crosses with Vinton 81, the inclusion of selection for enhanced thousand seed weight in the selection strategy could possibly increase the probability of finding improved genotypes. For the crosses performed with GF4X-21-5-2, improvement of both protein and sucrose content seemed unlikely due to highly negative correlation between those traits.

Samenvatting

Het gewas sojaboon [*Glycine max* (L.) Merrill], lid van de familie van de vlinderbloemigen, wordt over de hele wereld geteeld. Door zijn gunstige samenstelling (zijnde een eiwit- en oliegehalte van respectievelijk ongeveer 40% en 20%) zijn een brede variatie aan toepassingen mogelijk. Eén daarvan is het gebruik van soja in de menselijke voeding. Recent is de interesse naar zulke toepassing sterk toegenomen, hetgeen geleid heeft tot het ontstaan van een gespecialiseerde markt voor soja-gebaseerde voeding in verschillende regio's van de wereld en onder andere ook in Oostenrijk. Omwille van de vraag van deze markt, is de veredeling voor voedings-gerichte soja cultivars essentieel geworden. De belangrijkste veredelingsdoelen voor voedings-gerichte sojabonen zijn een hoog eiwit- en sucrosegehalte. Het GPX_SATO experiment heeft als doel het verhogen van de voedingskenmerken van Oostenrijkse geadapteerde soja cultivars. Daarom werden er vier gerichte kruisingen gemaakt tussen een geadapteerde moederplant met normale eiwitgehalte en een bestuiver met hoog eiwitgehalte. Selectie werd per kruising uitgevoerd op 200 $F_{3:5}$-lijnen volgens het eiwitgehalte: de 24 lijnen met het laagste eiwitgehalte en de 24 lijnen met het hoogste eiwitgehalte werden geselecteerd. De F6-generatie van de geselecteerde lijnen werd geanalyseerd. Uit de resultaten bleek dat de twee bestuivers die gebruikt werden, met name GF4X-21-5-2 en Vinton 81, goede kweekwaarde hadden voor voedingskenmerken in hun respectievelijk kruisingen. De twee kruisingen uitgevoerd met Vinton 81 leken populaties te produceren die het mogelijk maakt genotypes te vinden die betere eiwit- en sucrosegehalte bezitten dan de geadapteerde moederplant. Deze bevinding werd gestaafd door correlaties tussen eiwit- en sucrosegehalte die slechts middelmatig negatief waren. In één van de kruisingen met Vinton 81 werd er degelijk één verbeterde genotype gevonden. Voor de andere kruising die met Vinton 81 werd gemaakt, viel de selectie op eiwitgehalte tegen, een selectie die eerder gebaseerd is op sucrosegehalte zou mogelijk betere resultaten opleveren. Verder zou voor de twee kruisingen met bestuiver Vinton 81, een selectiestrategie waarbij er ook geselecteerd wordt voor verhoogde duizendkorrelgewicht, de kansen voor het vinden van betere genotypes verhogen. Voor de kruisingen uitgevoerd met GF4X-21-5-2, was het vinden van verbeterde genotypes onwaarschijnlijk omwille van een te negatieve correlatie tussen eiwit- en suikergehalte.

1. Introduction

1.1. Outline of the thesis

This thesis, consisting of six chapters, treats the breeding of soybeans destined for food applications.

In the first chapter an introduction is given on the cultivation of soybean worldwide, in Europe and particularly in Austria. The importance of soybeans at the global level will be emphasized. It will be seen that soybeans can be cultivated for various applications, the most important being animal feed. In contrast to that, a niche market for soybean destined for the human food consumption, so called food grade soybeans, has developed. In Europe and particularly in Austria, this market is present, leading into increasing interest for the breeding for food grade soybean cultivars.

The second chapter is devoted to the nutritional value of soybeans. There will be shown that the protein fraction of soybeans is appreciated for animal feed but also for human food. Moreover, soybeans possess other nutritional traits that suit humans. Especially the recent uncovering of functional components in soybeans leads into an increasing interest for soybean in human nutrition. Emphasize will be set on the fact that breeding work still needs to be done in order to enhance the beneficial nutritional traits of soybeans and to decrease the negative ones.

In the third chapter, the consumption of soybean based food and especially whole soybean products, so called soyfoods, will be discussed. Beyond the nutritional quality of soy products other opportunities given by the consumption of soy products will be discussed. In contrast to those opportunities, the constraints that can occur with the consumption of soybean based foods are given and it will be emphasized that breeding work is still needed in order to limit those constraints. Finally, the different applications of soyfoods will be outlined.

The fourth chapter treats the breeding of food grade soybean. The different breeding targets for improving food grade soybeans will first be treated. Obviously, a great share of those targets focuses on the enhancement of the nutritional value of soybeans outlined in chapter two. Following the targets, the specific food grade varieties and their food applications are listed, those applications are the same as the ones found in chapter three. Finally an overview of the Austrian soybean varieties and their characteristics will be given.

In the fifth chapter, the experiment made in the context of this master thesis is described. This experiment is an application of chapter three, where similar breeding targets are used in order to improve varieties adapted for Austria. In the experiment, focus is put on selection for enhanced protein content, importantly, the improvement of the sucrose content is also targeted, which is rather new in the breeding world. The objectives of this experiment are formulated in two hypothesis which will be accepted or refused by answering respective research questions.

In the sixth chapter a general conclusion about the experiment is given. Additionally, further possibilities for future investigations are proposed.

1.2. The origin and history of soybean

Soybean [*Glycine max* (L.) Merrill] is a member of the Fabaceae familiy, subfamily Papilionoideae, tribe Phaseoleae, genus *Glycine*. The genus *Glycine* is composed of two subgenera, *Glycine* (perennials) and *Soja* (annuals) which includes *Glycine max.*

The soybean that is currently cultivated originates from China and is the result of domestication that has probably taken place around 1500-1100 BC. By the first century AC it was probably grown in the Korean peninsula and central China. The introduction of soybean in Southeast Asia was finalized around the 15th to the 16th century and in Europe before 1713. The introduction in North America took place in 1765 [1].

It is assumed that the ancestor of the genus *Glycine* (x=10) has undergone tetraploidization approximately 59 and 13 million years ago [2]. However all described species of the genus *Glycine* exhibit normal diploid meiosis and are primarily inbreeders [1]. Therefore, soybean (2n=4x=40) can be considered as an ancient polyploid or paleopolyploid [2]. The further evolution of soybean started from a common wild perennial progenitor (2n=4x=40) that evolved to a wild annual (2n=4x=40) and finally to the domesticated soybean (2n=4x=40) [1].

1.3. The production of soybeans worldwide

From its domestication until now, soybean evolved to an important crop at the global level [3]. In 2010, soybean was globally grown on 102 million ha and the production reached 261 million metric tons. The United States of America (USA) that covers 35% of the world production is the major producer followed by Brazil, Argentina, China and India [4].

Soybean is the leading genetically modified (GM) crop produced worldwide. In 2007, over half (58.6%) of the soybean grown was GM. For this same year, 85% of the soybean production in the USA was GM, in Argentina 98% of the soybeans was GM and in Brazil 64% [4].

The typical composition of soybean of about 20% oil and 40% protein, has contributed in making soybean the world leading oilseed crop [1] and an appreciate protein source [5]. The use of soybean varies from feed, biodiesel, edible oils and other food products [6].

In the last 40 years, the world production of soybean has increased over 500%. The higher standard of living in country's like China results in an increasing demand for meat and thus for animal feed. Additionally, a higher demand for biodiesel feedstock is also occurring. It is therefore certain that the world production of soybean will continue to grow [6].

Furthermore, the unravelling of the health and nutritional benefits of soybean contributed to increased interest for soybeans destined for food production, so called food grade soybeans. Hence a niche market for food grade soybeans developed, where the breeding of cultivars with enhanced food grade traits is gaining great importance [7].

1.4. Processing and utilization of soybeans

Soybeans, typically possess protein and oil contents of approximately 40% and 20% respectively. This composition gives the possibility for the production of a broad variety of applications. The most important application of soybean is its use in animal feed. Soybeans can also be found in pharmaceutical, cosmeceutical and industrial applications. Finally, soybeans are also used in the food industry.

1.4.1. Application in animal feeding

Approximately 85% of the soybean is crushed into oil and meal. The greatest part of the meal is used as protein source in animal feed [6]. Soybean is the major protein source in livestock farming [8].

1.4.2. Pharmaceutical, cosmeceutical and industrial applications

Figure 1.1 gives an overview of the pharmaceutical, cosmeceutical and industrial applications in which soybeans can be found.

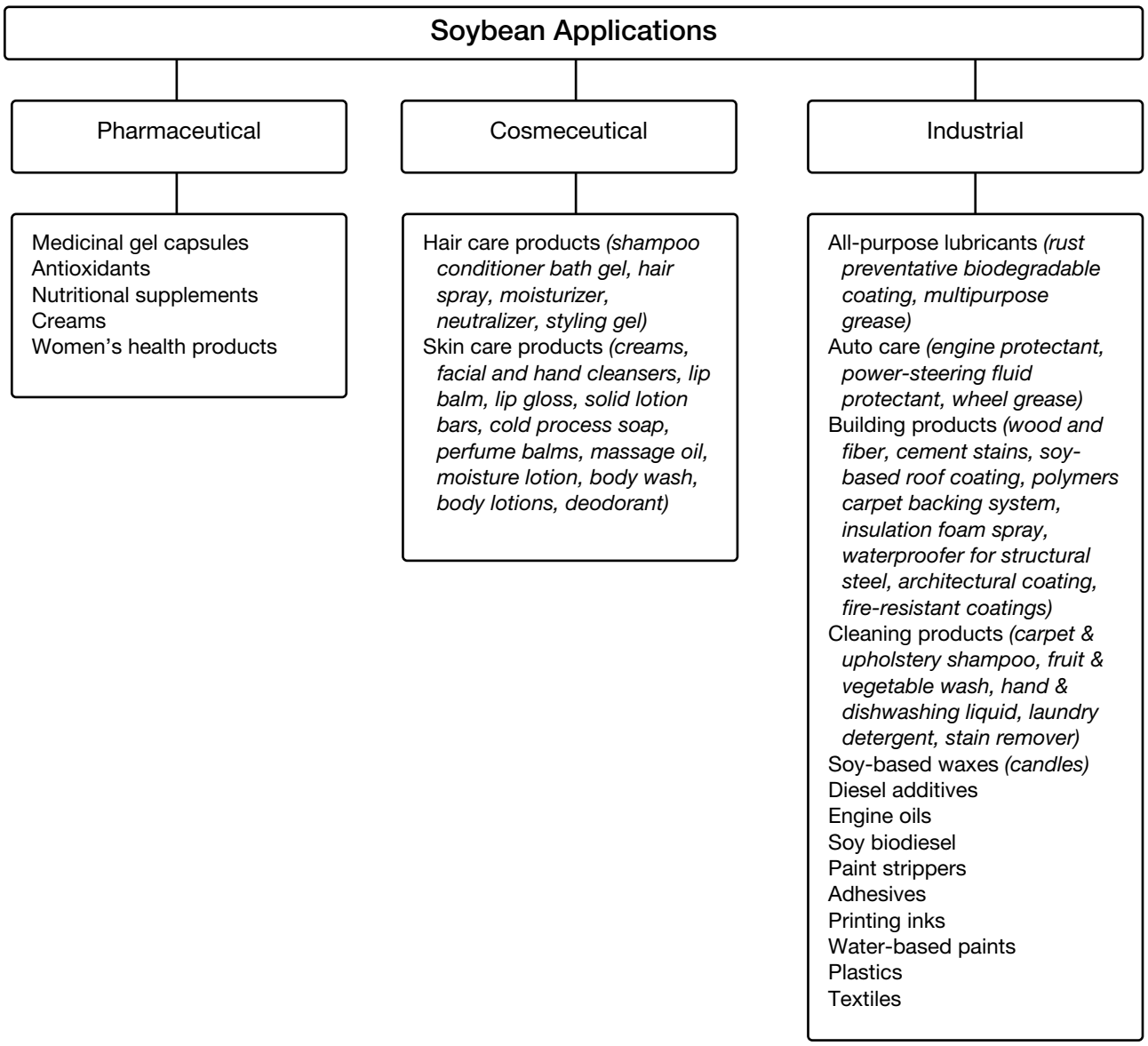


Figure 1.1 Pharmaceutical, cosmeceutical and industrial applications of soybeans [9]

1.4.3. Application in human nutrition

Figure 1.2 gives an overview of important food applications of soybeans, they can roughly be classified in whole bean products, protein products and oil products

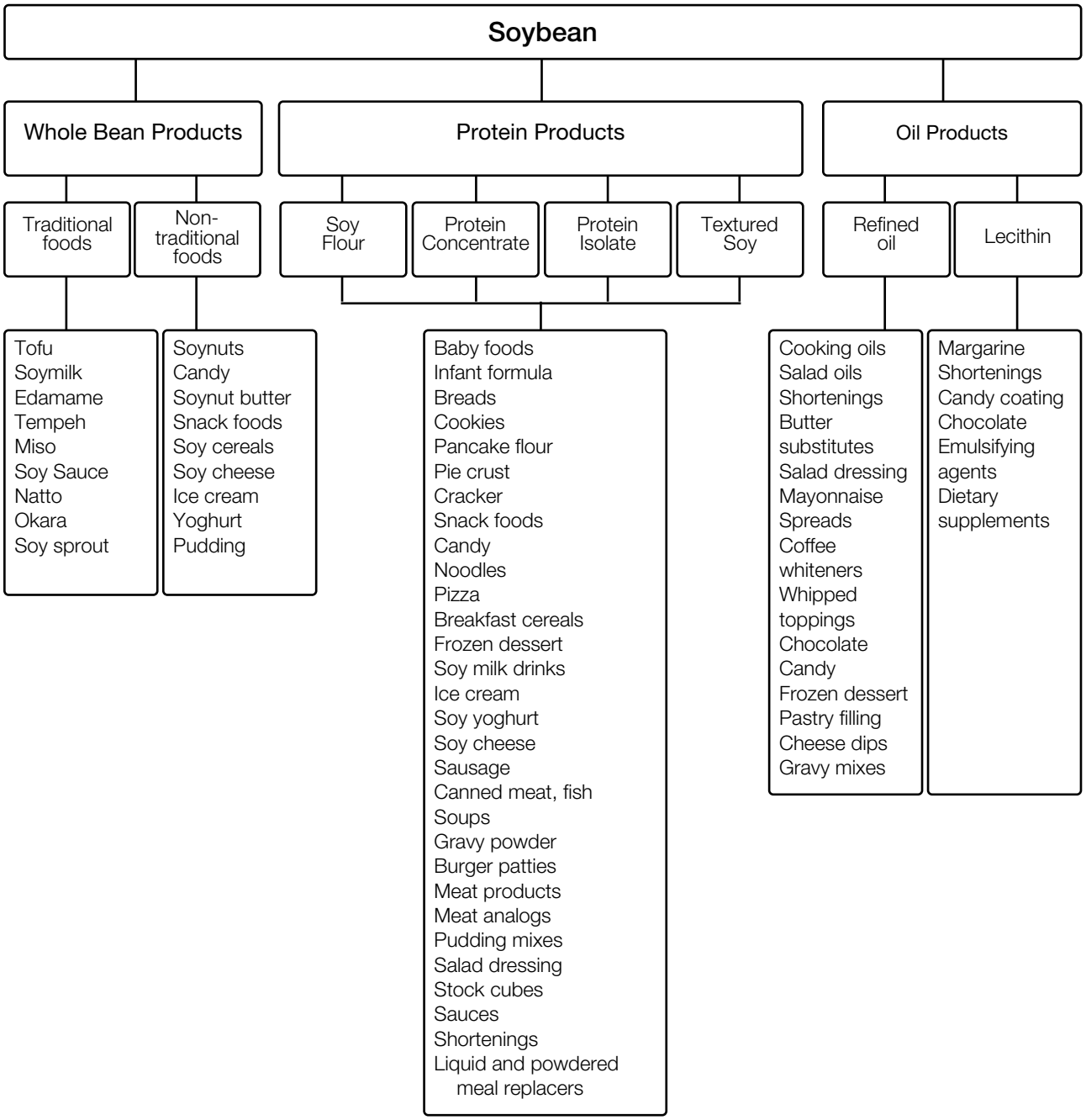


Figure 1.2 Important food applications of soybeans (adapted from [9]).

Oil products

Next to the soybean meal, soybean oil is also obtained after crushing. Soybean oil is refined and lecithin is obtained as main by-product. The refined oils are used as cooking oils, salad oils, they can also find their use in for instance mayonnaise, butter substitute. Lecithins are found in many products such as in margarine, chocolate and emulsifying agents [9].

Protein products
As already mentioned, most of the soybean meal obtained from crushing is used for animal feed. However approximately 2% of this meal is used for the production of soy flours and soybean proteins for the food-processing industry [6]. This is possible as next to its nutritional properties, soybeans also possess desirable functional properties such as water and fat absorption, emulsification, foaming, gelation, and binding [9]. Soybean protein products can for instance be added to wheat flour for the production of bread and cake. It can also be found in sausages where it improves water absorption and palatability [8].

Whole bean products
Approximately 6% of the produced soybeans are directly used for food purposes [6]. Soybean food products made from whole beans are also called soyfoods. Soyfoods can be divided in traditional and non-traditional soyfoods. Tofu, soymilk, edamame, tempeh, miso, soy sauce, natto, okara and soy sprouts are considered as traditional foods. They originate from Asia where soybean has been grown for centuries before its introduction to the rest of the world. Non-traditional foods are for instance soy yoghurt, soy cheese, pudding, snacks [9].

1.5. Soybean in Europe and Austria

The European Union (EU) imports each year approximately 40 million tons of raw soy products from primarily Brazil, the USA and Argentina [10]. Next to importing it, the EU also produces soybean, in 2010, it was harvested on more than 370 000 ha. Still in 2010, the EU produced over one million metric tons of soybeans covering 0.4% of the global production. The main EU countries for soybean production are Italy, Romania, France, Austria and Hungary [4].

In Central European countries such as Austria, soybean is mainly grown for livestock feeding whereas a smaller share is destined for food production. In those regions food grade soybeans are mostly grown for the protein fraction more than for the oil fraction, as vegetable oil mainly originates from oilseed rape [11].
The Austrian company “Mona Naturprodukte” is the leading food grade soybean producer in Central Europe and its products can be found in the whole European market. Mona processes only whole, non-GM soybeans which are produced in Europe. The Austrian food grade soybean production reflects the importance of this market in the EU, creating a context for the breeding of food grade soybean cultivars [12].

1.6. The cultivation of soybean in Austria

In 2010, Austria cultivated over 34 000 ha soybeans and produced over 94 000 metric tons soybeans, with an average yield of 2750 kg/ha [4]. The following figure gives an overview of the importance of several soybean-producing Austrian states in 2010.

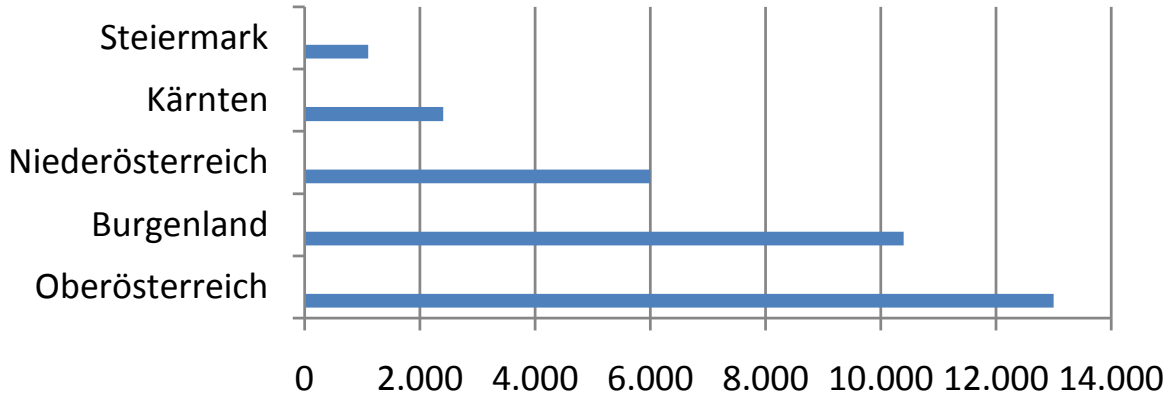


Figure 1.3 The area (in ha) of soybean production in different states of Austria in 2010 [13].

Climate
Soybeans favors warmer climates and therefore also profits from the climate change. The geographical laying of Austria as well as its good lands are appropriate for the cultivation of soybeans [13].

Maturity groups
Soybean are photoperiodic sensitive for flowering, which means that the day length determines the time when the plant will flower and reach maturity. Maturity in soybean is classified in 13 groups ranging from earliest (000) to latest (X) [14]. Soybean cultivars grown in Austria are early maturity cultivars of group 00 and 000 [13], meaning they need less days in order to reach maturity than later groups because day length during growth is long (May until September). As day length depends on latitude, soybean cultivars from the same group can be grown in regions in the same latitude such as the south of Germany, Switzerland, Hungary and some parts of France.

Inoculation
As soybean is a legume crop, it can provide itself with nitrogen by nitrogen fixation [13]. Fixation occurs by the symbiosis with root-nodulating rhizobia [*Bradyrhizobium japonicum* (Kirchner) Jordan] and an average of 111 kg N/ha can be generated [15]. Nitrogen fixation results in the fact that except in some cases fertilizing the soybeans with nitrogen is not necessary.
Soybean seedlings are inoculated with rhizobium and can be bought as such. After repeated cultivation of inoculated soybean on an area, permanent infection of the soil can occur making inoculation unnecessary [13].

Soil
Soybean needs neutral soil of pH between 6.5 to 7.5. In order to obtain efficient nodulation, a humid, well aerated soil is required [13].

Sowing
The best period for the sowing of soybean seeds is between mid-April and the beginning of May. Soybeans must be planted on depth of 3 to 4 cm in the soil. Plant densities of 40 to 60 plants/m^2 in the field with row distance between 12.5 to 50 cm are advised [13].

Weeds
Important weeds in soybean field are *Galium aparine*, *Chenopodium*, chamomile, amaranth, black nightshade and *Panicum*. Especially, root-weeds such as thistle, field bindweed and docks are difficult to control. Weed control is performed mechanically or chemically [13].

Diseases and pests
Until now, relatively few problems with diseases and pests occurred in Austrian soybean fields.
In 2009 a great incidence occurred with the caterpillars of the painted lady butterfly. Control is possible with insecticides.
In the juvenile phase of the soybean seedlings, pigeons, rabbits and deers can cause damage to the leaves.
Sclerotinia stem root can also pose problems and it can be minimized by 4-year crop rotation for *Sclerotinia* sensitive crops in general.
Other possible damage can be caused by downy mildew, bacterial and viral diseases. Due to the weather, root diseases such as *Fusarium* can eventually appear. [13]

Fertilization
Nitrogen fixation of the nodulating bacteria ensures generally enough nitrogen for the growth of the soybeans. However for poor soils or when excessive rainfalls occurred, a starting fertilization with nitrogen can be needed. Fertilization with phosphate and potassium are required in doses depending on the availability of the soil [13].

Irrigation
Especially between flowering time until grain formation soybean plants are drought sensitive. At this moment, supply of 70-80% the amount of water needed for maize within the same period is required [13].

Harvest
The harvest of soybeans occurs between mid-September until the beginning of October, depending on the maturity and the weather. Soybeans are ready for harvest when the leaves are drying and falling. At this moment, the soybean seeds are round, loose from their cotyledon and are difficult to carve with the fingernail [13].

2. The nutritional value of soybean

Figure 2.1 gives an overview of the most important nutrients contained in soybeans, the concentration are based on dry weight. Functional components and vitamins were not included as they are only found in trace concentrations.

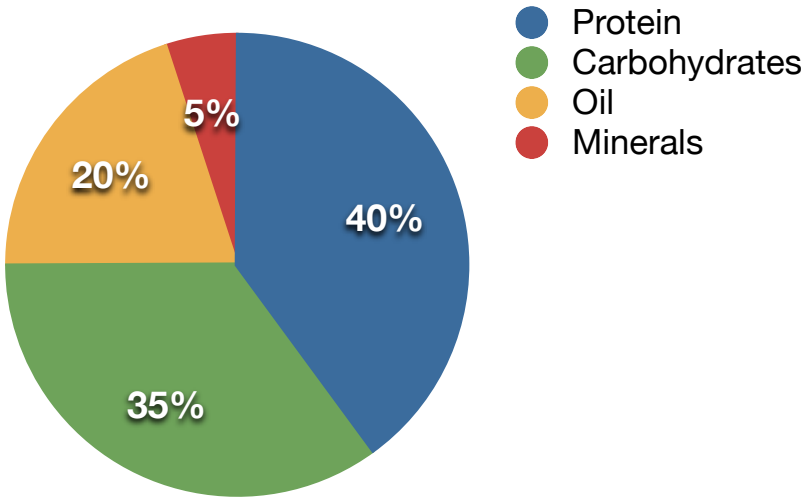


Figure 2.1 The typical composition of dry soybeans (based on [16][17]).

2.1. Soybean protein

Legume crops are well known for their high protein content thanks to the biological nitrogen fixation that occurs in association with nodule-forming bacteria. With a typical protein content of about 40%, soybean surpasses most of other legumes (except lupin, *Lupinus* species) [16] and even most commonly consumed food sources, such as meat, fish or cheese [5].

Beyond the content, also the quality of soybean protein is noteworthy as it is the most complete vegetable protein [18] and can be considered as being equivalent to animal protein [16]. Concretely, with exception of sulphur-containing amino acids such as methionine, the amino acid pattern of soybean resembles the pattern derived from high-quality animal protein sources [18]. Furthermore the lower amount of sulphur-containing amino acids is not limiting as it is not significantly less than required by the human body [16] . In fact, soybean protein can even enhance the nutritional quality of other vegetable protein. Protein sources which are deficient in some amino acids can be completed by soybean. As an example, the lysine deficiency of protein-containing grains, such as wheat or corn, can be corrected by soybeans that contain this amino acid in amounts that exceeds human requirement [18]. The advantages of the soybean amino acid pattern are also used for animals. Globally soybean meal accounts for 63% of all the protein sources in animal feed. It is rich in lysine, tryptophan, threonine, isoleucine and valine and therefore complements well with cereal grains that are deficient in those amino acids [16] .

By ultracentrifugation studies, four different fractions have been revealed, with approximate Svedberg coefficients of 2S, 7S, 11S and 15S [9].

2.1.1. 2S fraction

The 2S fraction contains from 8 to 22% of the extractable soybean protein. It consists of several enzymes, including the trypsin inhibitors, Bowman-Birk and Kunitz inhibitors [9]. Trypsin inhibitors inhibit the protein-cleavage effect of proteases (such as trypsin). It has a negative effect in animal feed, the digestibility is reduced leading into growth depression. Therefore, soybean meal needs first to be heated in order to inactivate the trypsin inhibitors. Trypsin inhibitors, however, can have an important role for humans as they have been found to be powerful anti-carcinogenic agents, therefore they can be considered as functional components of soybeans (see paragraph 2.6) [17].

2.1.2. 7S fraction

The 7S fraction covers 35% of the extractable soybean protein. The 7S fraction is mainly composed of 7S globulin (β-conglycinin) accounting for 85% of the fraction. Other proteins present in this fraction are lipoxygenase and lectins. Lypoxygenase will be introduced here, as it is of importance for the breeding section. Lectins are considered as a functional component of soybeans (see paragraph 2.6) [9].

Lipoxygenase
The lipoxygenase enzyme constitutes about 1-2% of the soybean protein. It poses problems in the taste of soybean based food, this issue will be handled in the limiting factors for soybean consumption (paragraph 3.2) [16].

2.1.3. 11S fraction

The 11S fraction comprises 31-52% of the extractable soybean proteins. A major part of this fraction (85%) consists of the 11S globulin (glycinin) [9]. The 11S fraction is responsible for the gelling character of tofu and hence the proportion of this fraction compared to 7S plays an important role in tofu firmness [19], this will be handled in (paragraph 4.1).

2.1.4. 15S fraction

This fraction comprises about 5% of the total extractable protein. It is only poorly characterized and is thought to be composed of polymers of the other soybean proteins [9].

2.2. Carbohydrates

Dry soybeans contain on average 35% of carbohydrates which can be divided in soluble and insoluble carbohydrates [17].

2.2.1. Soluble carbohydrates

Soybean seeds possess 15 to 20 different soluble carbohydrates that makes up approximately 15 to 25% of dry weight [20]. The most relevant soluble carbohydrates for breeding of food grade soybean are sucrose, raffinose and stachyose and they will therefore be treated in this paragraph.

Sucrose
Sucrose in dry soybean seeds are found in contents of typically 5.5% [17]. Sucrose is important for improving taste in soybean based products (see paragraph 4.1).

Oligosaccharides (raffinose and stachyose)
The oligosaccharides raffinose and stachyose typically constitute about 0.9% and 3.5% of dry soybean seeds respectively [17]. They can be found in soybean products but are also sold in Japan as oligosaccharides purified powder for human consumption [16].
Oligosaccharides have the disadvantage that they are difficult to digest for humans. This issue will be treated in the limiting factors for soybean consumption (paragraph 3.2). However they also have health benefits, for instance the stimulation of bifidobacteria in the colon. Therefore oligosaccharides are also considered as functional components (see paragraph 2.6) [16].

2.2.2. Insoluble carbohydrates

The seed coat of soybeans contains a major part of insoluble carbohydrates such as cellulose, hemicellulose, pectin and a trace amount of starch [17].

2.3. Soybean oil

Although the soybean seed oil content, about 20%, is not the highest among other crops such as peanut (*Arachis hypogaea*), mustard (*Brassica* species) and sunflower (*Helianthus annuus*), soybean oil is the second largest in terms of production and uses after palm oil (*Elaeis guineensis*) [16] .

Crude oil obtained after extraction with an organic solvent contains major and minor components [17].

2.3.1. Major components

The major components of crude soybean oil are triglycerides. After refinement of the oil, soybean oil is composed for 99% of triglycerides. Triglycerides are neutral lipids composed of one glycerol linking three fatty acids [17].
Fatty acid composition is an important trait for breeding of food grade soybeans (see paragraph 4.1), this will therefore be introduced here. Table 2.1 gives an overview of the fatty acid composition of soybean oil as well as of several other vegetable oils [21].

Table 2.1 Fatty acid composition of several vegetable oils [21].

	Fatty Acid (in %)				
Species	**Palmitic (16:0)**	**Stearic (18:0)**	**Oleic (18:1)**	**Linoleic (18:2)**	**Linolenic (18:3)**
Soybean (*Glycine max*)	11	4	22	53	8
Palm (*Elaeis guineensis, E. oleifera*)	40-69	4-5.5	20-44	4-12	<1
Canola (*Brassica napus*)	4	2	56	26	10
Sunflower (*Helianthus annuus*)	6	5	20	69	<1
Cottonseed (*Gossypium* spp.)	27	2	18	51	Trace
Peanut (*Arachis hypogaea*)	12.5	2.5	38	41	<1
Coconut (*Cocus nucifera*)	7	2.5	4.5	1.5	
Olive (*Olea europa*)	10-20	2-3	55-78	7-19	<1
Palm kernel	5-14	1-4	5-38	2-33	
Linseed (*Linum usitatissimum*)	6	3	17	14	60
Corn (*Zea mays*)	13	2.5	31	52	1

Saturates

The saturated fatty acids in soybean oil are palmitic acid (16:0) and stearic acid (18:0), they are found in average concentrations of about 11% and 4% (relative to the oil) respectively. Saturates are considered unhealthy but are useful in making low trans-fat margarines [1].

Monounsaturates

Soybean oil contains the monounsaturated fatty acid, oleic acid (18:1), which is on average present in about 22% of the oil. Monounsaturated fatty acids are healthy and have good oil stability [1].

Polyunsaturates

Soybean oil possesses the two polyunsaturated fatty acids, linoleic acid (18:2) an omega-6 fatty acid and linoleic acid (18:3) an omega-3 fatty acid [16]. They can be found in average concentration of 53% and 8% of the oil, for linoleic and linolenic acid respectively.
Linoleic and linolenic acid are known as being healthy and are furthermore essential fatty acids since humans lack the enzymes necessary to produce them [22]. Moreover, the ratio of omega-6:omega-3 fatty acids approximates the ideal ratio of 5:1, it has been suggested that an imbalance in this ratio could cause many chronic diseases such as diabetes, cardiovascular diseases and osteoporosis [16].

2.3.2. Minor components

Minor components found in crude soybean oil include phospholipids, unsaponifiable material, free fatty acids and metals. Unsaponifiable material consists of tocopherols, phytosterols and hydrocarbons [17]. Tocopherols and phytosterols are considered as functional components (see paragraph 2.6).

2.4. Minerals

Dry soybeans have an ash content[1]of about 5%. After processing, most of the minerals are found in the meal fraction rather than in the oil fraction [17].

2.4.1. Major minerals

Major minerals in soybean are potassium, which is present in the highest concentration followed by phosphorus, magnesium, sulfur, calcium, chloride and sodium. Those minerals are found in average concentrations ranging from 0.2 to 2.1% of the dry soybean seed [17].

2.4.2. Minor minerals

Minor minerals found in soybeans include silicon, iron, zinc, manganese, copper, molybdenum, fluorine, chromium, selenium, cobalt, cadmium, lead, arsenic, mercury, and iodine [17].

2.5. Vitamins

Soybeans contain water-soluble and oil-soluble vitamins.

2.5.1. Water-soluble vitamins

The water-soluble vitamins such as vitamin B_1 (thiamin), vitamin B_2 (riboflavin), vitamin B_5 (pantothenic acid) and vitamin B_6 (niacin) are not lost during soybean oil extraction. The newly discovered water-soluble vitamin pyroloquinoline, a new member of the vitamin B family, is present in some soyfoods such as tofu and natto. This vitamin plays a major role in the metabolism of lysine [16]. The amount of ascorbic acid (vitamin C) is essentially negligible in mature soybeans, although it is present in measurable amounts in both immature and germinated seeds [17].

2.5.2. Oil-soluble vitamins

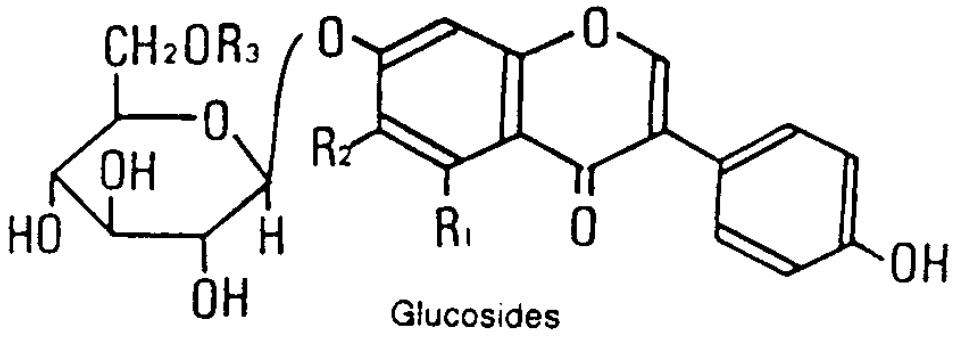
The main oil-soluble vitamins of soybeans are vitamin A and vitamin E (tocopherols). Vitamin A mainly exists in the form of β-carotene in immature and germinated seeds, whereas it is present in negligible amount in mature seeds [17]. Vitamin E or tocopherols are considered as functional components and will be handled in paragraph 2.6.

[1] The ash content is a measure for the concentration of minerals. Ash is the inorganic residue remaining after the water and organic matter have been removed by heating in the presence of oxidizing agents, which provides a measure of the total amount of minerals within a food [45].

2.6. Functional components

Although the health benefits of soybean were already known by the Chinese and the Koreans, it is only in the recent years that the special biological ingredients of soybean were highlighted in other countries. Therefore, there is more and more interest to consider soybean food products as functional foods, i.e. foods that contain biological components that deliver special health benefits to the consumer [16]. Example of beneficial health effects are anticancer, hypocholesteromic and antioxidative effects.
Functional components of soybeans include isoflavones, saponins, lecithin, trypsin inhibitors, lectins, oligosaccharides, tocopherols and phytosterols [17]. The functions of those components will not be detailed, as this is not the scope of this work. However, isoflavones and tocopherols will be introduced here due to the increasing interest for those components in the breeding world [19].

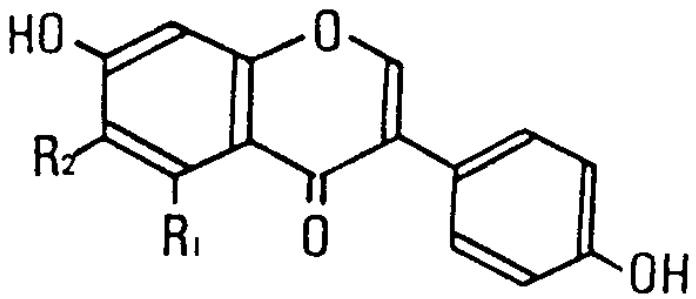
Isoflavones
Isoflavones are flavonoids found in soybean and other legume crops. In soybean seeds and non-fermented products isoflavones are predominantly found in their glycoside form (Figure 2.2). After consumption they undergo enzymatic metabolism in the small intestine which results in the release of their bio-active forms, daidzein, genistein or glycitein (Figure 2.3). Isoflavones are phytoestrogens and are known to have positive health effects such as the reduction of the risks for coronary heart disease, osteoporosis, certain types of cancer and the moderation post menopausal symptoms in women [23].



Glucosides

R1	R2	R3	Compounds
H	H	H	daidzin
OH	H	H	genistin
H	OCH3	H	glycitin
H	H	COCH3	6”-O-Acetyldaidzin
OH	H	COCH3	6”-O-Acetylgenistin
H	OCH3	COCH3	6”-O-Acetylglycitin
H	H	COCH2COOH	6”-O-Malonyldaidzin
OH	H	COCH2COOH	6”-O-Malonylgenistin
H	OCH3	COCH2COOH	6”-O-Malonylglycitin

Figure 2.2 Chemical structure of the glycoside forms of isoflavones [24].



Aglycones

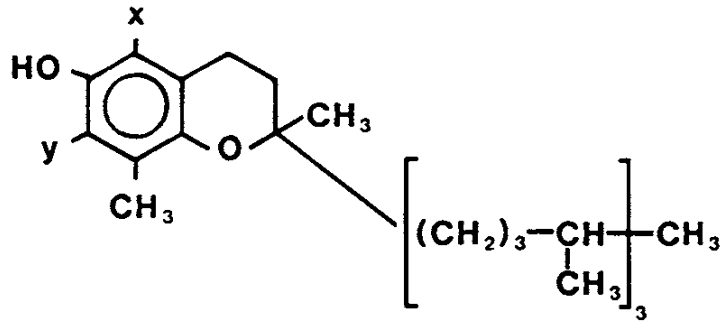
R1	R2	Compounds
H	H	daidzein
OH	H	genistein
H	OCH3	glycitein

Figure 2.3 Chemical structure of the aglycone form of isoflavones [24].

Soybeans possess 0.1-0.4% of isoflavones on a dry weight basis, hence soybean possesses the highest amount of isoflavones compared to all other crops [17]. The isoflavone concentration varies considerably depending upon the genotype and environmental conditions. It has also been shown that cultural practices such as increased doses of fertilizer and irrigation enhances their concentration. Furthermore, the concentration of isoflavones in soyfood is influenced by the different processing methods [16].
It is thought that isoflavones are mainly responsible for most of the health benefits from soybean based foods. Therefore they gained more and more attention from the scientific world [17] and research on breeding for enhanced isoflavone content is increasing (see paragraph 4.1)[19].

Tocopherols

Tocopherol is an important constituent of soybean oil [17]. Refined soybean oil possesses about 1000-2000 mg/kg. Tocopherol exist in four isomers, three of them, the α-, ϒ-, δ-isomers are present in soybean oil. α-tocopherol is known as natural vitamin E and soybean is the leading commercial source of this vitamin. Tocopherols protect the polyunsaturated fatty acids from oxidation, hence they are antioxidants [1].



α-tocopherol	x = y = CH3
β-tocopherol	x = CH3 , y = H
ϒ-tocopherol	x = H , y = CH3
δ-tocopherol	x = y= CH3

Figure 2.4 Chemical structure of the four isomers of tocopherol [17].

Tocopherols are used in pharmaceutical applications. Medical evidence has shown that the intake of 400 IU/day of tocopherols[2] results in a decreased risk for arteriosclerosis, cancers and degenerative disease such as Alzheimer and Parkinson and an improved immune system [16].
The tocopherol content decreases with the oil refinement process, therefore it requires hydrogenation to maintain oxidative stability [1]. Tocopherol content also varies significantly from one soybean variety to another [17]. In order to enhance oxidative stability and nutritional value of soybean oil, interest in modifying the tocopherol content by breeding is therefore increasing (see paragraph 4.1) [1, 19].

[2] 1 IU (International Unit) is defined as 1 milligram of an equal mix of the eight stereoisomers of α-tocopherol, which is a racemic mixture called all-rac α-tocopherolyl acetate. [46]

3. The consumption of soybean based food

As discussed in previous chapter, soybean is an ideal source of protein and possesses many health benefits. Soybeans are therefore favorable for human nutrition. The consumption of soybean based food or soy products will be discussed in this chapter. As mentioned in paragraph 1.4, soybean is used in human nutrition as protein products, soybean oil and whole soybean products. Protein products are not consumed directly as food but are used in the food industry. Soybean oil is essentially of importance in North-America but is almost not consumed in Austria. Most of the food grade cultivars grown in Austria are thus destined for the production of soyfoods [11]. This chapter, therefore, mostly focus on the consumption of whole soybean products. Moreover, in the last paragraph description of several types of soyfoods are given.

3.1. Other opportunities given by the consumption of soy products

Soy products can be consumed for their beneficial nutritional quality. Furthermore the consumption of soy products can undoubtedly give an answer to important issues.

3.1.1. Lactose intolerance and cow's milk protein allergy

People suffering from lactose intolerance can not properly digest milk sugar found in cow's milk as well as in many milk derivatives. Lactose intolerance origins from an insufficiency or deficiency of the enzyme lactase in the human body that is normally produced by the cells that line the small intestine. This results in an inability to break down lactose to simpler forms that can be absorbed into the blood stream. As a consequence lactose will be fermented by the intestinal bacteria leading to a number of disorders such as flatulence and diarrhea.
People with lactose intolerance must therefore limit or suppress the consumption of cow's milk products. Soy products give a good alternative for cow's milk products as they contain no lactose and can indisputably make part of a healthy diet thanks to their composition. Concretely soymilk enriched with calcium has a very beneficial nutritional profile high in vegetable protein and polyunsaturated fatty acids and is cholesterol-free. [25]

Allergy to cow's milk protein is the most common allergy in children worldwide, almost 5% of children below the age of 5 years suffer from this allergy. An allergy differs from an intolerance (e.g. lactose intolerance) by the fact that it affects the immune system and is not a reaction by the body to certain foods. Cow's milk allergy can manifest itself in different ways such as dermal, respiratory and/or gastrointestinal complaints.
The only way to treat cow's milk allergy is to exclude any form of cow's milk protein in the diet, this can be helped by the consumption of soy products as they are free of cow's milk protein. [26]

3.1.2. Soy against obesity

Worldwide obesity has nearly doubled between 1980 and 2008 becoming a global issue. This epidemic also concerns Europe, in 2008 over 50% of adults in the World Health Organization (WHO) European Region were overweight and roughly 23% of women and 20% of men were obese. Unfortunately obesity also particularly affects children, estimates of the number of overweight infants and children in the WHO European Region rose steadily from 1990 to 2008. Childhood obesity is strongly associated with risk factors for cardiovascular disease, type 2 diabetes, orthopedic problems, mental disorders, underachievement in

school and lower self-esteem [27]. Further, the rise of obesity also implicates economic consequences as it is estimated that obesity accounts for 7% of the total healthcare costs in Europe [28].
The consumption of soy products can be helpful to reduce obesity. It has been reported that soybean regulates the insulin level, which may impact obesity as a high concentration has been found to be a major cause of obesity. Furthermore studies have shown that a soy diet results in weight loss in women [16].

3.1.3. Sustainable development

Our current consumption of meat and animal protein has an important impact on the environment. The Food and Agriculture Organization (FAO) states that 18% of Green House Gas (GHG) emissions measured in CO_2 equivalent originates from livestock sector. This is higher than the emissions from the transport sector. Moreover, water and land use for livestock must be taken into account to estimate the environmental consequences of our current eating habits. It is estimated that 8% of the global water use is due to the livestock sector [29]. Because of those environmental issues we must reduce the consumption of animal products. This can be helped by the consumption of vegetal alternatives, like soy products. Soy products are from vegetal origin and therefore have a lower environmental impact than comparable meat and dairy products in terms of land use, water use and GHG emissions. Soy products can easily replace products of animal origins in the diet which can contribute in reducing our environmental impact [30].

3.2. Biochemical constraints that can limit the consumption of soy products

Despite of the many benefits of soy products, some factors limit the consumer to choose them. In this section the most important constraints for the consumption of soy products are pinpoint as well as the possibilities to rectify them.

3.2.1. Soy allergenicity

Soybean is included in the “big eight” foods that are believed to be responsible of 90% of all the food allergies. Currently the only treatment for soy allergy is to exclude soybean completely from the diet. This can be very challenging as soybean along with peanut was found in approximately 6000 foods and in most cases were not labelled on the commodity. Since avoidance of soybean is particularly difficult, research has been made by several groups for the hypoallergenization of soy products. The methods used for this purpose include thermal, enzymatic and chemical modification of the allergen, as well as breeding and genetic modification of soybean cultivars. However there are still issues related to hypoallergenization, to begin, not all the allergens of soybean have yet been identified. Further, it is possible to reduce the concentration of allergens but it is not always possible to remove them totally without impairing the structural composition and physiological development of the seeds which has negative consequences on the nutritional and functional properties of soybean. Another issue is that it is possible that by making cultivars with reduced allergens, alternative or novel proteins are expressed that might be allergenic. In other words, research is still needed to enhance the understanding of soy allergenicity and to find suitable techniques in order to remove it [9].

3.2.2. Digestibility

Consumers, especially in countries where fermented and vegetable soybean are not in vogue, may be skeptical towards the use of soy products because of flatulence and poor digestibility. These effects are caused by the oligosaccharides stachyose and raffinose (introduced in paragraph 2.2). As humans lack α1→6 glycosidase required for degradation of the α1→6 galactosidic linkage of stachyose and raffinose, those sugars remain undigested in the upper intestine and pass on the lower intestine. The metabolization by intestinal microflora in the lower intestine leads to the production of CO_2, hydrogen and methane. The gases cause abdominal discomfort. Although raffinose and stachyose can be reduced to an extent by soaking or boiling, genetic reduction or elimination is one of the prime plant breeding objectives (see paragraph 4.1) [16]. The reduction of raffinose and stachyose may also result in improved flavor from the increase in soluble sugars, additionally it could also lead to benefits gained from genetic traits that improve functional characteristics of soybean protein [19].

3.2.3. Undesirable flavor

The lipoxygenase enzyme (introduced in chapter 2.1), it generates a grassy-beany flavor when it oxidizes fats. Especially in countries where the consumers are not used to fermented and vegetable soybean, this flavor is not appreciated. It is possible to avoid the oxidation of the fats by heat inactivation of the lipoxygenase enzyme, however this is cost-ineffective and leads to insolubilization of proteins. Therefore, the genetic elimination of the lipoxygenase is preferred in order to reduce the beany flavor. Fortunately, genotypic variation and the influence of growing environment on lipoxygenase in soybean seed are well documented in the literature [16] [19].

Isoflavones and saponins may also be the cause of undesirable taste in soy products although this is not well documented yet. The breeding of cultivars with low isoflavones and saponins is theoretically possible but the positive marketing appeal of isoflavones overrides the negative taste aspects in some market segments [19].

3.3. Foods made from whole soybean seeds: Soyfoods

For over 5000 years soyfoods have been a part of the daily life in Asia in traditional foods such as tofu, soymilk, tempeh and natto [19, 31]. Some of these traditional soyfoods have also become popular in other parts of the world. In recent years new soyfoods have been developed that are familiar to consumers and that include the functional and nutritional properties of soybeans. They include soynuts, soymilk yoghurt, meat and cheese alternatives. Nowadays, soyfoods can be found in many parts of the world and in addition, in a broad range of applications. [31]

3.3.1. Traditional soyfoods

Tofu

Tofu is perhaps the most widely consumed soyfood in the world. It is a regular part of the diet in many Asian nations and is available in most Western nations. Tofu is appreciated for being versatile, having a mild flavor and having in addition a high nutritional value. Noteworthy is that since it is naturally processed it retains a good deal of important nutrients and phytochemicals such as the isoflavones [31]. Tofu typically contains 7.8% protein and 4.2%

lipid on a wet basis [32]. It has a relatively low carbohydrates and fiber content since the pulp is removed, making it easier to digest [31].

Tofu is a curd made by coagulation of the protein and oil in soymilk. There are two main types of tofu: silken, or soft, tofu and momen, or hard, tofu. The latter is made by soaking whole soybeans and grinding them into a slurry with water. The slurry is afterwards cooked to form soymilk. A coagulant is then added, this is most commonly magnesium chloride, calcium sulfate, or glucono-D-lactone; the coagulants can be used purely or in combinations to achieve different flavor or textural characteristics. Heating is also usually applied in order to facilitate the coagulation. The result of the coagulation is that after a few minutes the soymilk begins to curdle and large white clouds of tofu curd begin to form in a sea of yellow whey. The curds are then removed and placed in cloth-lined forming boxes where pressure is applied to the top in order to remove more liquid which results in a firm curd. Silken tofu in comparison to momen tofu is not pressed and is often coagulated in the container in which it is to be sold. [19, 31]

Soymilk

The popularity of soymilk has grown quickly in the USA and Europe since the 1980's. Traditionally it is made from whole beans in the same way as the first few steps of tofu manufacture. This soymilk contains nutrients, isoflavones, saponins, and other soluble components of the soybean from which the soymilk is made. However, beverage-quality soymilks available today are usually prepared from soy protein isolate, to which fats, sugars, and carbohydrates are added to improve flavor and generate a nutritional profile similar to that of cow's milk. Some manufacturers add isoflavones back into the soymilk in order to make health claims about the product. Additionally a number of soymilks available is fortified with vitamins and minerals, such as β-carotene and calcium or docosahexaenoic acid (DHA), an omega-3 fatty acid. [19, 31]

Vegetable Soybeans (Edamame, Mukimame)

Vegetable soybean consists of the whole soybean picked at the R6-R7 stage. At this stage the soybean is at its peak of green maturity, contains a high level of sucrose and chlorophyl, and has a firm texture. After harvest, the pod can be left entire or be shucked into individual beans. After being blanched and frozen the soybean can be sold as "edamame", referring to the entire pod or "mukimame", referring to the individual beans [19, 31].
Vegetable soybeans are sweet as they are picked at a stage of high sucrose content [31]. Vegetable soybeans contain 11-16% protein and 8-11% oil on a fresh weight basis [33]. They can be prepared in stir-fry dishes or as dips and other preparations [31].

Tempeh

Tempeh is a traditional fermented soyfood that originates from Indonesia. It has a typical texture, a distinctive nutty or mushroom-like flavor and is versatile. The production of tempeh goes as follows: whole soybeans are dehulled, cracked and cooked in water with vinegar so that the pH is low. After the cooking, the soybeans are mixed to spores of the fungus *Rhizopus oligosporus* and incubated at a temperature of approximately 31°C. After 24 hours incubation a compact, cake-like product, covered and penetrated with the mycelia of the fungus, is obtained. Variation in the taste and texture can be added by mixing with grains and seeds [31].
Tempeh possesses a high protein content (40-50% of dry matter) and can therefore be used as replacement for meat or fish in the diet [34].

Miso

Miso is a rich and flavorful paste extensively used as soup-base and flavoring ingredient in many Asian countries like Japan, Korea, Taiwan, Indonesia and China. It is produced by fermentation of aged whole soybeans or soybeans combined with koji nuggets (grains such as wheat, rice or barley which have been cultured with *Aspergillus oryzae*) [31]. The fermentation is performed at temperatures between 30-38°C for a period of six months, depending on the type of miso produced [32]. After fermentation and fully ripening, it is blended and packaged to be sold. The different types of miso that can be produced go from sweet white miso, that is quite mild, to dark savory miso, a more robust and salty type.
Above its taste, miso has also specific medicinal properties, it is believed to reduce the effects of environmental poisons in the body. It also contains enzymes and bacteria that facilitate the digestion. Miso has a high protein and sodium content; because of the latter, it should be consumed sparingly. [31]

Soy Sauce

Soy sauce is the most famous flavoring ingredient of the Asian cuisine. The traditional production method is similar to that of miso which means that it can either be produced exclusively with soybean, “tamari” soy sauce, or either with a fermented wheat starter, “shoyu” soy sauce. However most of the soy sauce currently produced is made of a combination of hydrolyzed vegetable protein (HVP), sugar, color and preservatives. HVP is the result of a chemically-induced fermentation of soy protein.
The protein content in soy sauce is the highest in tamari, followed by shoyu and finally HVP-based soy sauce. Similarly to miso, soy sauce must be used sparingly because of its high sodium level. Soy sauce with lower sodium level has therefore been made, as well as several flavored soy sauce products. [31]

Natto

Natto is a traditional soyfood that originates in Japan. It is produced by the fermentation of whole, small, cooked soybeans with *Bacillus natto* resulting in a sticky, viscous coating. It has a very typical taste and aroma appreciated by the Japanese consumers but it is less evident to people who are not used to it. Natto can be sold fresh or frozen. [31]

Okara

Okara is the fibrous residue obtained after processing of soybeans to soymilk. It contains the insoluble carbohydrates and dietary fibers of the soybean as well as some remaining protein and fat. Usually it is not sold in stores because it is very wet, heavy and perishable but it can be used to make meat alternatives or tempeh. It can also be cooked and blended to supplement breads and other baked goods. [31]

Soy sprouts

Soy sprouts are germinated soybean sprouts that have grown for five to seven days [31]. They are part of the traditional diet in Korea, where there are also known for their anti-hangover function. The main component of the soybean sprout extract is asparagine [35]. Soy sprouts also have a high protein and fiber content and are a good source of vitamin C. [31]

3.3.2. Second generation soyfoods

In response to the demands of the consumers for convenient and healthy soyfoods, new soyfoods or so-called second generation soyfoods were developed. Currently thousands of different soyfoods are available. The possibilities go from tofu-stuffed pasta, to pizza topped with soy cheese. Yoghurt, ice cream and cheese are produced based on soymilk. There exist flavored, marinated, baked or smoked versions of tofu and tempeh. Meat alternatives can be made by mixing gluten with soy concentrates. In summary, soyfoods exist in a wide variety of applications. Some examples are reviewed in this section [31].

Soynut

Soynuts are prepared by dry roasting or oil roasting of in water soaked, whole or split, soybeans. They can be coated with salt or other flavoring ingredients to be sold as crunchy nuts. Blended soynuts together with other nuts can be used in baking applications or in other preparations. Soynuts can also be used to make soynut butter, the production process is similar to peanut butter.
Soynuts have a high protein and fiber content, furthermore they contain isoflavones [31].

Meat alternatives

Meat alternatives include hundreds of applications from burgers, to hot dogs, sausages, meatballs and so on. Those products are made from tofu, tempeh, textured soy flour, textured soy concentrate, isolated soy protein and wheat gluten. A combination of different vegetable protein makes it possible to obtain the right texture and flavor for a specific use. These products are mostly low in fat or even completely fat-free [31].

Cheese alternatives

Cheese alternatives can be block, spliced, spreadable cheese flavored like American cheese, mozzarella, cheddar or Parmesan. They are made from soymilk, tofu or other vegetable protein ingredients. Frequently, casein (protein from cow's milk) is added, as this protein makes the cheese melt when it is heated. Cheese without casein soften but does not melt or stretch. The fat of cheese alternatives can be completely removed or replaced with vegetable oil [31].

Soymilk yoghurt

The production of soymilk yoghurt is similar to that of cow's milk yoghurt. Namely, pasteurized soymilk is incubated with a suitable culture such as *Acidophilus* until it has become yoghurt. The taste of soymilk yoghurt is very similar in comparison with that of cow's milk yoghurt, it is also available in all kinds of styles and flavors. Soymilk yoghurt can thus be used in the same manner as cow's milk yoghurt, also for instance in recipes that require yoghurt. In the USA soymilk yoghurt can not be labelled as "yoghurt" as it is not produced out of cow's milk. Other names such as "soygurt" or "cultured soymilk" are therefore mentioned on the packages.
Soymilk yoghurt has a high protein level and is rich in isoflavones. [31]

Nondairy frozen desserts

Nondairy frozen dessert can be produced from soymilk, soymilk yoghurt, tofu or isolated soy protein in approximately the same way as non-dairy frozen desserts. Although some niche markets exist, this kind of products is still trying to find a place in a broader market. [31]

4. Breeding of food grade soybean

Soybean is a self-pollinating crop, hence soybean cultivars are pure lines. To obtain a soybean cultivar, breeders first choose the parental and population structure from which the cultivar will be derived. The type of cross used for hybridization is mostly a biparental cross, but other crosses such as three way crosses or backcrosses are also used. Once the heterozygous plants are obtained from a successful cross, the breeding material must be returned to the near homozygous state. Mostly, for soybean breeding, this is achieved with the single-seed descent method. Pedigree method with or without early generation testing is also used but in a less extensive way [22].

4.1. Breeding targets for food grade soybeans

4.1.1. Agronomic traits

Agronomic traits for which breeders select are yield, maturity, plant height, standability, adaptability, disease and shattering resistance, and stress tolerance.

Yield
The breeding objectives for soybean cultivar development depend on many factors such as the geographical environment, the purpose of the soybean seeds, etc... However a high seed yield is the first objective for any breeding program [36].

Maturity
Breeding for maturity is indispensable to obtain cultivars adapted for the growth in a certain environment (maturity was introduced in paragraph 1.6).

Seed size
The desired seed size depends on the type of soyfood for which the soybeans are destined (this will be discussed in paragraph 4.2)
Another goal for breeders is to obtain lines that have soybeans with uniform seed size as they are more attractive and marketable. Therefore seeds can be screened with different-sized sieves. [35]

Seed appearance
Seed appearance is also important for marketing. Generally seeds with thin, strong seed coats and shining luster are desirable.
Seed coat mottling, purple stain, splits, insect damage, cracked seed coats and other physical defects are not desirable. Breeders mostly detect those traits by visual observation. It is also possible to screen for cracked seed coat by soaking seeds in water and checking the swell rate. [35]

4.1.2. Nutritional traits

4.1.2.1. Protein fraction

Protein content
High protein and low oil content add nutritional value to soyfoods. Germplasms that cover a wide range in protein content (33.1 to 55.9%) and oil content (13.6 to 23.6%), are available for breeders in order to modify the seed/oil ratio. The negative correlation between protein and oil facilitates the development of high-protein and low-oil lines. Protein and oil content are routinely measured by Near Infrared Reflectance Spectroscopy (NIRS).
High protein content is generally associated with low yield, which makes the development of lines that combine high protein and high yield difficult. However high yield is mostly achieved by selection for moderately high protein content (43 to 45%).[35]

11S/7S
Consumer preference for the degree of tofu firmness is a matter of personal taste. However, mostly high tofu firmness is preferred. The genotypic variation in this trait is partly due to the ratio of 11S to 7S protein fraction in the seed. The 11S fraction generally possesses greater gelling potential than 7S, hence high 11S-to-7S ratio is desirable as it results into harder than those with low ratio. The 11S-to-7S ratio is reported to range from 0.3 to 4.9. It is noteworthy that genotypes with same 11S-to-7S ratio do not always result in the same firmness because of different 11S subunit composition. The subunits possess differing gelling characteristics resulting in varying gelling potential of the 11S. In other words, a high 11S-to-7S ratio as well as suitable 11S composition is of importance for good tofu firmness [19].

Lipoxygenase
Normal soybean seeds contain three lipoxygenase isozymes that are responsible for the grassy beany flavor and bitter taste of soyfood (discussed in paragraph 3.2). Research is being conducted for the genetic elimination of lipoxygenase from soybean seeds to reduce undesirable flavors in soyfood products. Several combinations of lipoxygenase-null mutants have already been developed: 0-genotypes with one of the isozymes eliminated, 00-genotypes with two isozymes eliminated and 000-genotype with all isozymes eliminated. The 000-genotype yields no detectable level of the lipoxygenase proteins in mature soybean seeds, resulting in the absence of the grassy and beany flavor. The presence or absence of three lipoxygenase isozymes is determined by gel electrophoresis, spectrophotometer or by immunological or colorimetric methods. [35]

4.1.2.2. Carbohydrates

Sucrose content
Sucrose is the major source of energy for fermentation and contributes to the sweet taste of soyfoods especially for tofu, soymilk and edamame. In order to get a sweeter taste, breeders aim to increase the sucrose content in soybean seeds. The sucrose content in soybeans ranges from 1.5 to 10.2%, germplasm with even higher content, 13.6%, has been identified. [35]

Oligosaccharides content
Stachyose and raffinose are not readily digestible and cause flatulence when soyfoods are consumed (discussed in paragraph 3.2). Therefore, breeders aim to minimize the content of those oligosaccharides in soybean seeds. Stachyose and raffinose content among soybean germplasm range from 1.4 to 6.7%, and 0.1 to 2.1%, respectively. Breeding lines with less than 1% stachyose and raffinose have been developed and will have a positive impact on the soyfood and soymeal industries. [35]

4.1.2.3. Functional components

Isoflavones
Because of its beneficial health effect a good isoflavone content is desirable for soybean cultivars. Isoflavone content is influenced by genetic factors and environmental factors such as irrigation and temperature during seed maturation [35]. For instance, the total isoflavone content of soybean seeds appears to be negatively related to growth temperature [19].
Little is known about the genetic regulation of the synthesis of isoflavone in soybean. Interest has been put in the phenylpropanoid synthetic pathway which is catalyzed in its first step by Isoflavone Synthase (IFS). Two genes for IFS have been identified in soybean. Understanding the genetic regulation of this pathway may be necessary for obtaining cultivars with good isoflavone levels. Furthermore, negative correlation has been found between total isoflavone content and linolenic acid (18:3) concentration. Other data suggest negative correlation between isoflavone content and protein content. Therefore, obtaining an enhanced isoflavone content in an high-protein cultivar may pose problems in the future [19].

Tocopherols
Enriched amounts of α-tocopherol or natural vitamin E in oil should provide an additional beneficial aspect to soybean oil. Oils containing low contents of linolenic acid (18:3) have been shown to contain high amount of α-tocopherol. It is noteworthy that this trend results in lowered amount of ϒ-tocopherol [19].

4.1.2.4. Oil fraction

Seed oil concentration
Intentionally or unintentionally, increasing the seed oil concentration has been a breeding goal for centuries. The ancestor of the domesticated soybean used to have small, hard, black seeds with low oil content, high protein content and low yield. By the selection for yield, agronomic characteristics and seed quality, large yellow seeds with typical averages of 20% oil and 40% protein were obtained.
It is known that an increase in oil content is positively correlated with yield and negatively correlated with protein content. However, soybean is appreciated for its high protein meal and versatile vegetable oils; therefore, breeders mostly prefer to obtain modest gains in oil and yield without substantial loss in protein concentration. [1]

Reduced saturates
Saturated fatty acids are one of the major dietary components responsible for elevating cholesterol. Therefore, consumers favor a diet lower in saturates.
The saturated fatty acids present in soybean oil are palmitic acid, 16:0 and stearic acid, 18:0. Especially palmitic acid is a health concern as it is correlated to cardiovascular disease. It has been suggested that saturated fatty acids should be kept below 7-10% on a daily basis. Hence, breeders try to obtain high yielding soybeans that correspond to those limits. [1]

Increased saturates
Although reducing saturates is one of the key targets for oil improvement, oils with increased saturates are useful for the production of low trans-fat margarines. To avoid the health issues related to saturates, it is important to consider that, unlike palmitic acid, stearic acid either reduces or has no effect on cholesterol levels in humans. [1]

Increased monounsaturates
Soybean oil contains the monounsaturated fatty acid, oleic acid, 18:1. The oxidative stability of the oil is enhanced when the concentration of 18:1 is three times higher than the normal content which is about 22%. Therefore, breeders target a concentration of 18:1 of about 65-75% of total lipid in soybean. By the means of genetic engineering, 18:1 levels of about 80% total lipid have been achieved. [1]

Reduced polyunsaturates
For production of solid fats from oils partial hydrogenation is applied. However, when the oil is composed of polyunsaturated fatty acids, partial hydrogenation leads to the production of trans fatty acid which are unhealthy as they have been correlated with cardiovascular disease [22]. Therefore major processors favor oils that result in the production of reduced or zero trans-fat products. In response to this concern, soybean breeders target a reduced linolenic fatty acid, 18:3, concentration. [1]

Increased polyunsaturates
The majority of soybean oil is processed into food applications requiring reduced linolenic acid content. However linolenic acid is an omega-3 fatty acid essential for the human body and the consumption of it in edible oils has beneficial health effects. For this purpose, research has been made to increase the concentration of 18:3 by the means of genetic engineering. [1]

4.1.3. Soyfood product quality

Soyfood product quality aspects such as ease of cooking, taste, are particularly important in determining whether a variety is suitable and has potential for commercialization. Those aspects must therefore be tested, this happens usually in collaboration with marketing groups, distributors or manufacturers. [35]

4.2. Specific types of soybean varieties

The breeding of food grade soybeans can be classified in three major categories: the breeding of large-seeded soybeans, breeding of small-seeded soybeans and the breeding of soybean with unique seed composition [35].

4.2.1. Large seeded soybeans

Large-seeded soybeans are bred for tofu, soymilk, miso, edamame, and soynuts. [35]

Tofu and soymilk
Most of the breeding of large-seeded soybeans is destined for the production of tofu and soymilk.

An important factor for the breeding of tofu soybeans is the tofu yield, which is defined as the weight of fresh tofu produced from a unit of harvested soybean. By targeting specific traits, soybean breeders try to obtain soybeans with good yield and quality [19].

Tofu soybeans are preferably large and approaching spherical shape as those soybeans possess lower surface-to-volume ratio and thus reduced seed coat. The seed coat is insoluble and is not desirable for tofu because greater soluble dry matter leads to a greater tofu yield [19]. Tofu soybeans are larger than 20g/100-seeds [35].
Beyond seed size, seed appearance is also of importance for tofu soybeans. It is possible to produce good quality tofu with dark hilum beans but this requires prior dehulling of the beans and careful soymilk filtration [19]. In order to avoid these additional processing steps, soybeans with a yellow seed coat, yellow cotyledon and clear hilum are preferred. Moreover a thin but strong seed coat that is free from cracking and discoloration is desirable [35].

Nutritionally, high protein content exceeding 45% on dry matter basis is desirable for tofu soybeans as this enhances tofu yield. To enhance the gelling characteristics of tofu an improved ratio of 11S/7S is desired [19]. A high protein/oil ratio provide a higher tofu yield and firmer texture, therefore low oil content is preferred. Moreover tofu soybeans have lack of lipoxygenase, high water uptake, a low calcium content and a high germination rate. The composition and content of carbohydrates influences the taste of tofu and soymilk [35]. High total sugar content (above 8% on dry matter basis) [19], high sucrose, low raffinose and low stachyose are highly desirable. [35]

Examples of tofu and soymilk varieties: Black Kato, Toyopro, Grande, Proto (from Minnesota), Vinton-81, HP 204, IA1007, IA1008 (from Iowa). [35]

Edamame
Vegetable soybean seeds are very large (>30 g/100-seeds dry weight) [35]. As the pods are eaten directly, they preferably have sparse gray pubescence. Moreover edamame cultivars should possess as less as possible of one-seeded pods as they require greater effort to shell by consumers. The seed coat is preferably green [19] and thin [35], therefore cultivars with genetically “stay green” seed coat are favored. Those cultivars have delayed yellowing effect of maturity making it possible for growers to have extended harvest period closer to maturity [19].

Important nutritional traits are a high content of sugar (sucrose and maltose) and free amino acids to impart sweet and delicious taste. Sucrose is primarily responsible for the sweetness of vegetable soybeans, sucrose content is preferably higher than 10% on dry matter basis.

Certain free amino acids, such as glutamic acids, are major contributors to the taste of vegetable soybeans.
It is also important that vegetable soybeans have tender texture in order to have a better mouth feel. [35]

Examples of edamame varieties: Merrimax, Peterson Jade, Disoy, Magna, Prize, Grande, Verde, Emerrald, Saturn [35].

Miso and soynuts
The desirable seed quality for producing miso is similar to that for tofu.

Cooked or roasted mature soybeans are called soynuts. Similar features to those for tofu and edamame are desired. [35]

4.2.2. Small-seeded soybeans

Natto
For natto, small to ultra-small soybeans (smaller than 9g/100-seeds [19]) of maximum of 5.5 mm diameter are preferred for better fermentation. The seeds have preferably a near-spherical shape as this reduces the ratio of the tough seed coat to softer cotyledon [19]. Also clear hilum and thin seed coat are desirable traits for natto soybeans [35].

Natto soybeans are nutritionally characterized by a high carbohydrate content [35]. A high content of soluble sugars results in a softer natto product, an important requirement for natto. The total sugar content must be higher than 10% on a dry weight basis [19]. The composition of sugars is important for the effectiveness of fermentation [35]. Sucrose is consumed faster than raffinose and stachyose during fermentation. To obtain a steady and controlled fermentation low sucrose content with high stachyose and raffinose content is favored [19]. In order to provide amino acids for the fermentation process, a moderately high protein content is desirable. Oil content must be low, less than 18% based on dry matter is preferable, to enhance water absorption [35].

For a softer natto product, seeds must additionally possess high water-absorption capacity. Some natto producers require that the soybeans absorb a certain amount of water during soaking, the first step of natto manufacturing. Breeders use standard small-seeded lines, such as the cultivar Vance (known for having a medium ability for water uptake), to compare selected lines for water-absorption capacity [19].

Examples of natto varieties: Canatto, TNS, Nattosan, Nattawa (from Canada), Prato Chico, Minnatto, UM3 (from Minnesota), Vance, Camp (form Virginia) [35]

Soy sprouts
Soybeans with medium seed size (10 to 12g/100-seeds) and a high germination rate are preferred for bean sprouts. High-protein, high-isoflavone, high-sugar, and lipoxygenase-free soybeans are desirable for soybean sprouts. [35]

Examples of soy sprouts cultivars: N7484, Willcross 9640, SS516, MFS-553 and Chico. [35]

4.2.3. Food grade soybean varieties with unique compositional traits

Breeding for food grade soybeans with unique seed composition has focused on a specific nutritional trait of the soybean seed. Examples of such varieties are given according to the fraction from which the targeted trait origins.

4.2.3.1. Protein fraction

Examples of varieties that target a specific component of the protein fraction are varieties high in total protein content, high in β-conglycinin, low in lypoxygenase, high in specific amino acids such as lysine, methionine and threonine, low in allergenic proteins [35].

High-protein soybeans
High-protein soybeans possess protein levels of 43% or greater. They are used for tofu, soymilk, soy sauce, beverages, baked goods, pudding, cheese and meat analogs. High-protein soybeans are available as whole beans, full-fat flour, low-fat flour or soymilk powder. [35]

Examples of high-protein varieties are: IA3001, HP201, HP202, HP203, HP204, Prolina, AC Proteus, AC Hercule, AC Proteina, Toyopro and NE 3396. [35]

4.2.3.2. Carbohydrates

Examples of varieties that target a specific component of the carbohydrate fraction are varieties high in sucrose content and varieties low in oligosaccharides [35].

High-sucrose soybeans
High-sucrose soybeans provide high sucrose, full isoflavone content and lower indigestible carbohydrates. Compared to conventional soybeans, high-sucrose soybeans contain 40% more sucrose but 90% less stachyose and raffinose.
High-sucrose soybeans are used to produce tofu, soymilk, beverages, baked goods, puddings, cheese and meat analogs. They are available as whole beans, full-fat flour, low-fat flour or soymilk powder. [35]

4.2.3.3. Functional components

Examples of varieties that target a specific functional components are varieties high in tocopherol content [1], varieties with high isoflavone level and varities low in trypsin inhibitor [35].

High-isoflavone soybeans
High-isoflavone soybeans soybeans contain more than 0.4% isoflavones compared to levels of 0.15 to 0.25% for traditional soybean varieties. [35]

4.2.3.4. Oil fraction

Example of varieties with unique seed composition from the oil fraction are for instance varieties with high oleic acid content, high stearic acid content, low linolenic acid content or low palmitic acid content [35].

4.3. Food grade varieties grown in Austria

Figure 4.1 gives an overview of the characteristics of Austrian varieties in 2007, respectively for varieties of maturity group 000 and 00. Each data point represents a variety. Yellow data points correspond to varieties that produces soybeans with yellow hilum and red data points correspond to varieties produces soybeans with red hilum. The points are displayed according to their protein content and seed size. The varieties in the green square have relatively high seed size and high protein content, they are considered as food grade varieties. The list of the registered Austrian varieties can be find on the website of AGES (Österreichische Agentur für Gesundheit und Ernährungssicherheit) [37].

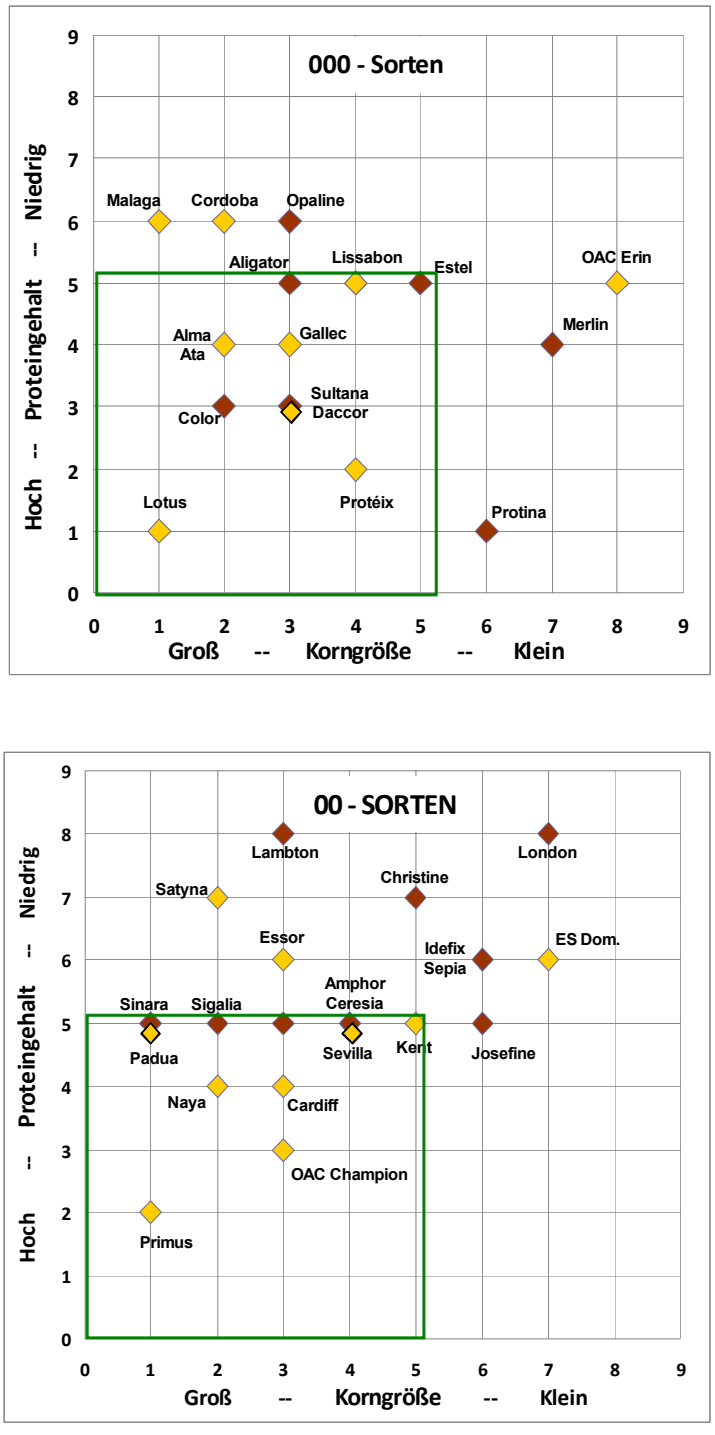


Figure 4.1 Varieties of maturity group 000 and maturity group 00 grown in Austria. [38]

5. Experiment: GPX_SATO

5.1. Objectives

In this experiment, 4 differently oriented crosses were performed between a cultivar with normal[3] protein content, adapted for the growth in Austria and a non-adapted, high-protein[4] pollinator. In order to obtain soybeans with food grade characteristics, one of the parental genotypes of each cross is also a food grade cultivar. A first selection was performed on maturity, while a second selection was performed based on protein level. The effect of the protein selection on other food grade traits was tightly investigated.

This experiment aims to enhance the food grade characteristics of Austrian adapted cultivars. To determine whether improved food grade genotypes are obtained and if differences are observed between crosses, 2 hypotheses are formulated. The hypotheses are accepted or refused on the basis of answers given on respective research questions.

Hypothesis 1:
By crossing two cultivars with contrasting protein content and by applying selection on protein content, it is possible to produce improved food grade soybean genotypes.
This hypothesis will be discussed individually for each cross made in this experiment.

- Research question 1 (RQ1): Is a genotype with higher protein content than its adapted standard parental genotype obtained in the F6 generation?
- Research question 2 (RQ2): Is a genotype with higher sucrose content than its adapted standard parental genotype obtained in the F6 generation?
- Research question 3 (RQ3): Is a genotype with higher protein content and higher sucrose content than its adapted standard parental genotype obtained in the F6 generation?
- Research question 4 (RQ4): Which traits are indirectly affected by selection for protein?
- Research question 5 (RQ5): Would an other selection-strategy be useful for the production of improved food grade soybeans?

Hypothesis 2:
Similarities are observed between some crosses: the similarities in those crosses are related to the fact that they possess the same high-protein pollinator.

- Research question 6 (RQ6): Are there similarities in traits between crosses with one common parent (different mother and same pollinator; same mother and different pollinator)?
- Research question 7 (RQ7): Are there similar patterns of indirect selection between crosses with one common parent (different mother and same pollinator; same mother and different pollinator)?

[3] "Normal" protein content means typical ranges between 38%-42%.

[4] "High" protein cultivars typically have protein contents that exceed 43%.

5.2. Materials and methods

5.2.1. Plant Material

5.2.1.1. Parental genotypes

Table 5.1 Overview of the parental genotypes, their respective maturity group, important traits and origin, used for the initial crosses of the experiment GPX_SATO

	Maturity Group	Protein content	Food Grade	Adapted	Origin
Gallec	000	normal	yes	yes	Switzerland/Austria
Essor	00	normal	yes	yes	Canada/Austria
GL601	00	normal	no	yes	BOKU
GF4X-21-5-2	0	high	no	no	BOKU
Vinton 81	I	high	yes	no	USA

Protein content: typical protein content range of the parental genotypes. “Normal” protein content means typical ranges between 38%-42%, While “high” protein genotypes typically have contents that exceed 43%.
Food grade: yes: the genotypes possess food grade characteristics such as e.g. relatively high protein and sucrose content, large seed, yellow hilum; no: the genotypes do not possess enough food grade characteristics to be considered as a food grade cultivar.
Adapted: yes: the genotype is adapted for the growth in Austria; no: the genotype is not adapted for the growth in Austria.

5.2.1.2. Initial crosses

The experiment started in 2005 with the performance of 4 initial crosses:

Table 5.2 Overview of the parents for each cross performed in the GPX_SATO experiment.

	Mother	Pollinator
Cross 1	Gallec	GF4X-21-5-2
Cross 2	Essor	GF4X-21-5-2
Cross 3	Essor	Vinton 81
Cross 4	GL601	Vinton 81

5.2.1.3. Selection scheme

2005: initial crosses

2006: **F_1**: for each cross the F_1 generation is grown in pots in greenhouse. The F_1 plants are of course uniform for each cross. A first selfing round results into the F_2-seeds, 800 seeds for each cross are further grown in the F_2. (Figure 5.1)

2007: **F_2**: about 800 plants per cross are grown in small plots. This is the first generation of segregation, hence the plants are not uniform and selection can take place. For each cross the seeds of 500 F_2-plants selected for early maturity are harvested. Plants with late maturity are discarded since they are not suitable to be grown in Austria. (Figure 5.1)

2008: **F_3**: 500 rows for each cross are grown (so in total about 2000 rows are grown). The selection goes as follows: the seeds (F_4-seeds) of one good plant from each row is again grown in a row in the next generation. (Figure 5.1)

2009: **F_4**: 500 rows for each cross are grown in the F_4-generation as $F_{3:4}$-families. All the seeds of each row are harvested. A sample of seeds of each row is analyzed for protein and oil content. For each cross, the 24 F_4-rows with the highest protein content are selected, as well as the 24 F_4-rows with the lowest protein content. (Figure 5.1)

2010: **F_5**: from this generation on 200 genotypes (lines) are grown, including:

For each cross: the 24 F_4-rows with the highest protein content and the 24 F_4-rows with the lowest protein content.

- Cross 1: 24 highest protein, 24 lowest protein : 48 $F_{3:5}$-lines
- Cross 2: 24 highest protein, 24 lowest protein : 48 $F_{3:5}$-lines
- Cross 3: 24 highest protein, 24 lowest protein : 48 $F_{3:5}$-lines
- Cross 4: 24 highest protein, 24 lowest protein : 48 $F_{3:5}$-lines

Standards: the parents except Vinton 81 and 4 test lines.

- GF4X-21-5-2 (one of the parents) : 1 line
- Gallec (one of the parents) : 1 line
- GL601 (one of the parents) : 1line
- Essor (one of the parents) : 1 line
- GH13X-4 (test line): 1 line
- Proto (test line): 1 line
- GK5X-3-8 (test line): 1 line
- GF2X-9-1-7 (test line): 1 line

200 small rows or plots, each consisting of a randomly chosen sample of approximately 100 seeds from each individual 200 lines, are grown in a lattice design.

The design is replicated in 2 blocks resulting in an experiment consisting of 400 experimental units. The F_6-seeds are harvested individually for each plot. Figure

5.2 gives an illustration of the design used in the F_5 and F_6-generation as well as the selection applied.

2011: **F_6**: A random sample of approximately 100 seeds is taken from the F_6-seeds of each line individually. The lines are grown in the same design as in previous generation (Figure 5.2). The F_7-seeds are harvested individually for each plot and analyzed. The results of the analyses are shown in the following section.

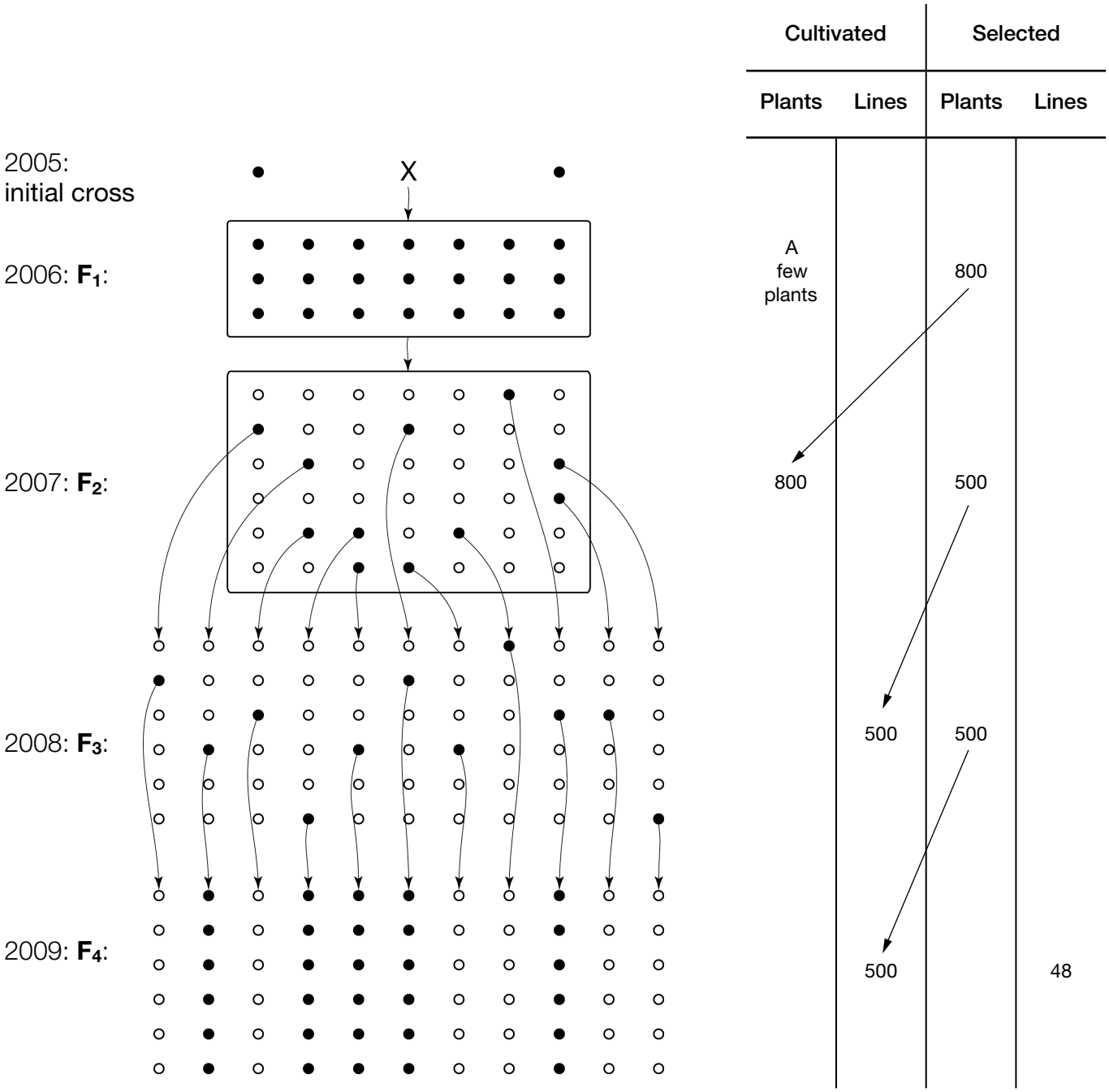


Figure 5.1 Selection scheme from the initial cross until the F4-generation, applied for each of the 4 crosses in the GPX_SATO experiment.

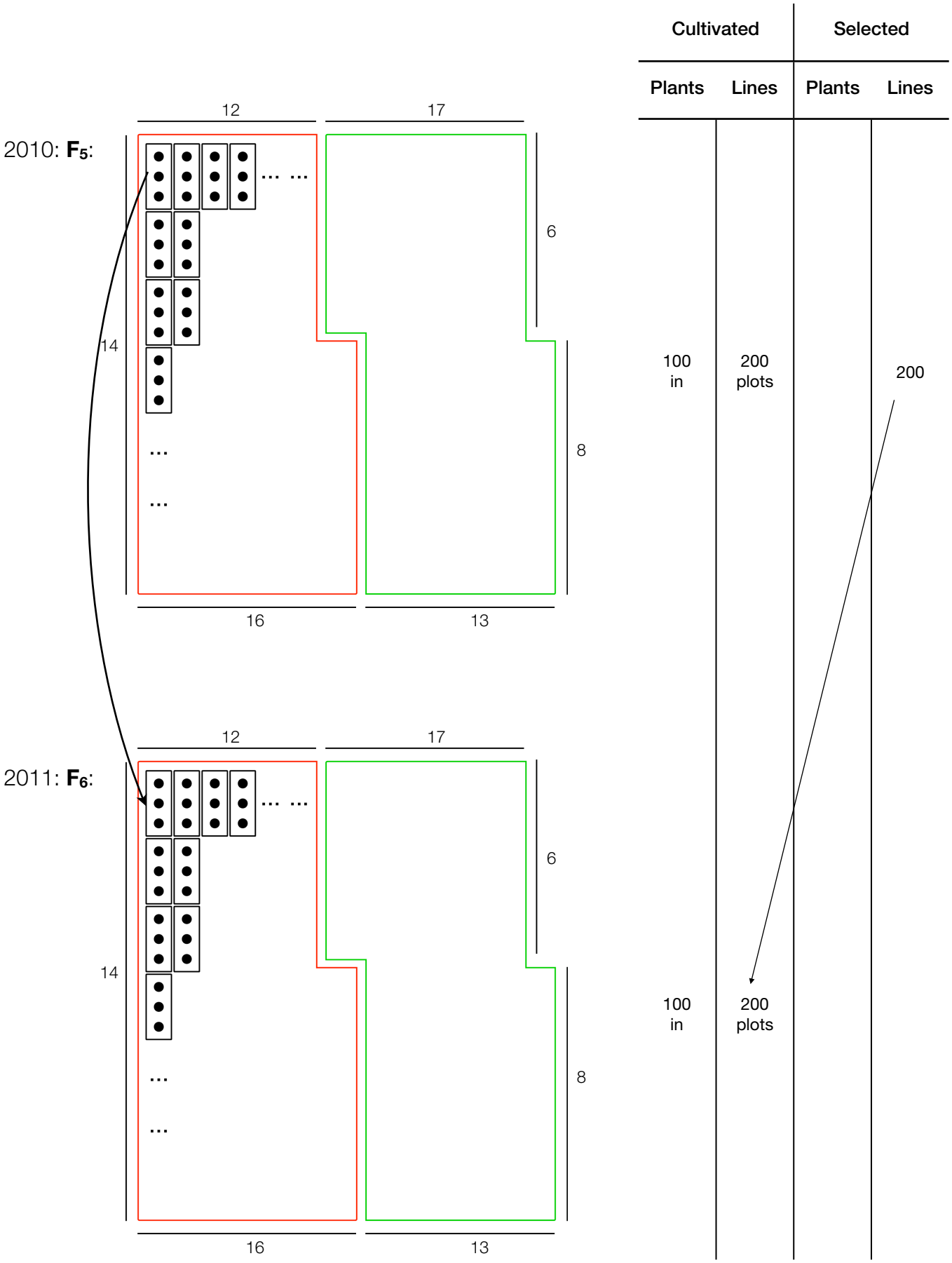


Figure 5.2 Scheme of the design and selection applied in the F5 and F6 generation of the GPX_SATO experiment. (The design consists of two blocks: the red block corresponds to the first replicate, the green block corresponds to the second replicate. Each block consists of 200 plots, the numbers display the number of plots placed on each length of the block).

5.2.2. Field experiments

All field experiments were performed at the BOKU university experimental station Gross Enzersdorf near Vienna, Austria. The soybean seeds of the F_6-generation were sown on the 3rd of May 2011 and the harvest took place the 4th of October 2011.
The experiment was performed according to a lattice design where single-row plots of 2.5 m row length and 50 cm row spacing with 50-70 plants per row were grown. Prior to sowing, soybean seeds were inoculated with a commercial preparation (Nodular G, Serbios, Badia Polesine, Italy) of soybean specific rhizobia (*Bradyrhizobium japonicum* (Kirchner) Jordan) to promote nodule formation and di-nitrogen fixation. No additional nitrogen fertilizer was applied.

5.2.3. Measurements

The measurements were performed on each experimental unit. The field measurements were performed on the plants of each plot indiidually in the field. The laboratory measurements were performed on the harvested seeds of each plot individually.

5.2.3.1. Field measurements

SPAD
The SPAD value gives an estimation of the chlorophyll content. 15-20 readings were taken from individual leaves from each plot by using a SPAD-502 chlorophyll meter (Konica Minolta Sensing, Osaka, Japan). The absorbance of the uppermost fully developed leaf of each plant on 19/07/2011 was measured at 650 nm wavelength (chlorophyll a and b both have an absorbance maximum at this wavelength).

Maturity
Maturity is counted as the number of days after 31/07/2011 until one reading of each plot reaches full maturity. The maturity was evaluated visually where full maturity means that all leaves from the plant felt down and about 90% of the pods are brown.

Plant height
Plant height expressed in cm, was measured on one reading per plot at harvest time.

Pod set score
Pod set score was estimated visually at harvest time, resulting to a score between 1 (= excellent) and 5 (= very weak). The pod set gives an estimation of the yield, the better the pod set (a low pod set score), the higher the yield should be. A high pod set score thus results probably in lower yields.

5.2.3.2. Laboratory measurements

Thousand seed weight
Thousand seed weight is used as an estimate of seed size. Thousand seed weight was determined by counting 100 seeds for each line with a CONTADOR counter (Pfeuffer, Kitzingen, Germany) and weighting those seeds with a Sartorius BASIC balance (Sartorius, Göttingen, Germany). The obtained value was then multiplied by 10, resulting to the thousand seed weight.

To determine protein, oil and sugar content, reflectance NIRS, Near Infrared Spectroscopy was performed. Therefore, near-infrared light is send on a sample (ground soybean seeds), part of this light is absorbed by the chemical functional groups of the sample, other parts are transmitted or reflected. In the case of reflectance NIRS, the reflected light is measured to obtain a spectrum. Based on the calibration for a specific analyte, the concentration can be determined from the resulting NIRS-spectrum. The calibration is performed by doing NIRS-analysis on a set of analyte samples that covers a wide concentration range. The NIRS-spectrum of each analyte is converted to a value for its concentration by the means of a calibration model (this can for instance be based on multiple linear regression). A calibration curve can then be set up with on the x-axis, the true concentration, based on a reference method, and on the y-axis, the NIRS prediction value of the concentration. [39]

To perform the NIRS-analysis for this experiment, about 10g of soybean seeds from each experimental unit was finely ground with a Cyclotec 1093 mill (Foss Tecator, Höganäs, Sweden). The obtained samples were analyzed by reflectance NIRS using a Bruker Matrix-I Fourier-Transform NIRS instrument and the OPUS software (Bruker, Ettlingen, Germany). This resulted in interferograms that were recorded as reflectance spectra in wave number range from 12492 to 3598 cm^{-1} (781 - 2779 nm wavelength). The protein, sucrose and oil content of the seeds were determined by making use of partial least square regression calibration models.

Protein content
Protein content was expressed in g/kg on a dry matter basis; the Kjeldahl method was used as reference for the calibration; the validation of the calibration model R^2 value was R^2 = 0.9737.

Sucrose content
Sucrose content was expressed in g/100g on a dry matter basis; the calibration was made on the basis of an enzymatic method; the validation R^2 value was R^2 = 0.9751.

Oil content
Oil content was expressed in g/kg on a dry matter basis; Soxhlet extraction was used as reference method for calibration; the validation R^2 value was R^2 = 0.925.

5.2.4. Data-analysis

5.2.4.1. Raw data

All the measurements of each experimental unit were put together in one data-file, so called raw data. Additionally, the protein+oil-content (the sum of the protein content and the oil content) was calculated since it is an important characteristic for food grade soybeans.

Table 5.3 Overview of the variables with respective units used in the raw data and adjusted data for the analysis of the F7 generation of the GPX_SATO experiment.

Variable	Unit
SPAD	SPAD-value
Maturity	number of days after 31/07/2011
Plant height	cm
Pod set score	score (1 = excellent; 5 = very weak)
Thousand seed weight	g
Protein content	g/kg
Sucrose content	g/kg
Oil content	g/kg
Protein+oil content	g/kg

5.2.4.2. Adjusted data

Spatial effects were removed from the raw data by lattice-analysis with the PLABSTAT software [40], in order to obtain the adjusted data (lattice adjusted plot values).

5.2.4.3. Means

The arithmetic mean for each trait was calculated among the two replicates of each genotype.

5.2.4.4. Analysis with PLABSTAT

Outgoing from the adjusted data, least significant differences and heritabilities of each trait, as well as the genotypic and phenotypic correlation for each pair-wise combination of traits were calculated with the PLABSTAT software [40], individually for each cross.

Least Significant Difference (LSD) was calculated on the 5% significance level for each trait within each cross. This makes comparison between values of a trait between genotypes within crosses possible, as two values of a trait significantly differ at the 5% level if their difference is greater than the respective LSD.

Narrow sense heritabilities[5] (h^2) of each trait were calculated within each cross by making use of the formula:

$$h^2 = \frac{\sigma^2 G}{\sigma^2 G + \frac{\sigma^2 GE}{e} + \frac{\sigma^2 E}{re}}$$

With σ_G^2 the genotypic variance, σ_E^2 the variance due to the environment and σ_{GE}^2 the genotypic by environment interaction. r is the number of replicates which is 2 and e is the number of environments which is 1.

Genotypic correlations[5] (r_{12G}) of each pair-wise combination of traits are calculated within each cross based on the genotypic variance and covariance:

$$r_{12G} = \frac{cov_{12G}}{\sqrt{\sigma_{1G}^2 + \sigma_{2G}^2}}$$

With cov_{12G} the genotypic covariance between the two traits, σ_{1G}^2 the genotypic variance of the first trait and σ_{2G}^2 the genotypic variance of the second trait.

Phenotypic correlations[5] (r_{12P}) of each pair-wise combination of traits are calculated within each cross based on the phenotypic variance and covariance:

$$r_{12P} = \frac{cov_{12}}{\sqrt{\sigma_1^2 + \sigma_2^2}}$$

With cov_{12} the phenotypic covariance between the two traits, σ_1^2 the phenotypic variance of the first trait and σ_2^2 the phenotypic variance of the second trait.

5.2.4.5. Analysis with SAS

SAS software [41] was used in order to perform F-tests between least square means (LS-means) of each measured trait of the lines selected for high protein content and those selected for low protein content. The F-tests were performed for each cross individually.

SAS software [41] was used in order to perform F-tests between LS-means of each measured trait between different crosses. The F-tests were performed:

Between cross 1 and cross 2 in order to assess significant differences between crosses with different mother but with the same pollinator.

Between cross 2 and cross 3 in order to assess significant differences between crosses with the same mother but with different pollinator.

Between cross 3 and cross 4 in order to assess significant differences between crosses with different mother but with the same pollinator.

---

[5] The variance components were determined by analysis of variance (ANOVA) within crosses. The ANOVA was performed based on a model which takes the replicates and the different genotypes and the interaction between replicates and genotypes into account. The **genotypic variance** is estimated by the component of variance due to the genotypes. The **variance of the environment** was estimated by the component of variance due to the replicates. The **phenotypic variance** is the sum of the genotypic variance and the variance of the environment.

5.3. Results

5.3.1. Hypothesis 1

By crossing two cultivars with contrasting protein content and by applying selection on protein content, it is possible to produce improved food grade soybean genotypes.

RQ1: Is a genotype with higher protein content than its adapted standard parental genotype obtained in the F6 generation?

RQ2: Is a genotype with higher sucrose content than its adapted standard parental genotype obtained in the F6 generation?

RQ3: Is a genotype with higher protein content and higher sucrose content than its adapted standard parental genotype obtained in the F6 generation?

RQ4: Which traits are indirectly affected by selection for protein?

RQ5: Would an other selection-strategy be useful for the production of improved food grade soybeans?

Table 5.4 gives the information in order to answer **RQ1**, **RQ2** and **RQ3** for cross 1, 2, 3 and 4. Table 5.4 summarizes the tables of Annex i.

Table 5.4 Heritabilities for protein and sucrose content for each cross, and the number of $F_{3:6}$-lines, within each selection class with a protein and/or sucrose content improved over the respective adapted parent.

Cross	Protein					Sucrose					Both	
	Heritability	Improved		Decreased		Heritability	Improved		Decreased		Improved	
		Low	High	Low	High		Low	High	Low	High	Low	High
1	0,86	20	24	0	0	0,7	0	0	9	14	0	0
2	0,77	11	22	0	0	0,78	0	0	15	22	0	0
3	0,25	0	0	0	0	0,61	0	0	3	3	0	0
4	0,63	15	14	0	0	0,58	0	3	1	0	0	1

Protein:
Heritability: Heritability for protein content for each cross.
Improved: Individually for each cross, the number of F3:6-genotypes, selected for either high or low protein content, with improved protein content compared to the respective adapted parent, based on the respective least significant difference.
Decreased: Individually for each cross, the number of F3:6-genotypes, selected for either high or low protein content, with decreased protein content compared to the respective adapted parent, based on the respective least significant difference.
Sucrose:
Heritability: Heritability for sucrose content for each cross.
Improved: Individually for each cross, the number of F3:6-genotypes, selected for either high or low protein content, with improved sucrose content compared to the respective adapted parent, based on the respective least significant difference.
Decreased: Individually for each cross, the number of F3:6-genotypes, selected for either high or low protein content, with decreased sucrose content compared to the respective adapted parent, based on the respective least significant difference.
Both:
Improved: Individually for each cross, the number of F3:6-genotypes, selected for either high or low protein content, with decreased protein content compared to the respective adapted parent, based on the respective least significant difference.

5.3.1.1. Cross 1

From table 5.4, it can be noticed that 44 genotypes have a significantly higher protein content, and thus show improvement, in comparison with the adapted parent Gallec (RQ1). It is noteworthy that from those 44 genotypes 20 were selected for low protein and still show improvement compared to Gallec. Moreover no genotypes have significant decrease of protein content compared to Gallec.
For the sucrose content, no genotypes show significant improvement (RQ2). This also results in the fact that no genotypes show significant improvement in both sucrose and protein content (RQ3). It is noteworthy that 23 genotypes had a significantly decreased sucrose content compared to Gallec.

Table 5.5 gives the information in order to answer RQ4 for cross 1:

Table 5.5 LS-means, F-test and correlations related to the selection on protein for each measured characteristics in the F6 generation of cross 1 in the GPX_SATO experiment:

		LS-means		F-test	Correlation with protein	
Parameter	**Unit**	**Low protein lines**	**High protein lines**	**Significance Level**	**Phenotypic correlation**	**Genotypic correlation**
SPAD	SPAD-value	40,9	42,5	0.0003	0.07	0.06
Maturity	days after 31/07/2011	41	37,7	0.0004	0.14	0.15
Plant height	cm	89	79	0.0004	0.16	0.20*
Pod set score	score (1 = excellent; 5 = very weak)	2,8	3	0.1598	-0.02	-0.12
Thousand seed weight	g	170	173	0.3737	-0.13	-0.12
Protein content	g/kg	448	463	<.0001	0,00	0.00
Oil content	g/kg	172	163	0.0128	-0.77**	-0.94**
Protein+Oil content	g/kg	619	626	0.0115	0.56**	0.69**
Sucrose content	g/kg	48	46	0.0505	-0.83**	-0.91**

LS-means: LS-means for each characteristic for the lines selected for low and high protein content respectively.
F-test: Significance level for the F-test, between the LS-means of the lines selected for low and high protein content, performed for each characteristic respectively.
Correlation with protein:
Phenotypic correlation between protein and each measured characteristic: ** : significant on the 1% level; *: significant on the 5% level.
Genotypic correlation between protein and each measured characteristic. Significance expressed as being greater than once (*) or twice (**) its standard error.

As expected the selection on protein had an effect on protein level as the LS-means for the low and high protein lines significantly differ at the 1% level.

From the correlations with protein it is clear that the selection for protein is negatively effecting oil content and sucrose content, as the respective phenotypic and genotypic correlations are negative and significant (respectively 1% and ** significance, for sucrose and oil content). For oil content this indirect effect of selection corresponds with the higher LS-mean for oil content of the low protein lines compared to the LS-means found for the high

protein lines. However, for sucrose content, no significant difference between the LS-means for sucrose content of the low and the high protein lines could be noticed.

It can be deduced that protein content is positively effecting protein+oil content (positive and significant phenotypic and genotypic correlations, respectively 1% and ** significance) and plant height (positive genotypic correlation at * significance). This is also confirmed by the respective comparison of the LS-means between the low and the high protein lines, the high protein lines display significantly higher LS-means for plant height and protein+oil content than the low protein lines.

The positive correlations between protein and protein+oil content, can easily be explained as protein is included to the calculation of protein+oil content; an increase in protein content automatically results in an increase in protein+oil content. Furthermore a new selection-strategy based on selection on protein+oil content can not be considered as more efficient as protein and oil content still would need to be measured.

The possibility of using the positive genotypic correlation between protein and plant height is investigated (RQ5 is investigated); phenotypic and genotypic correlation between plant height and each measured trait is displayed in following table, as well as the heritabilities for each trait.

Table 5.6 Heritability of each measured characteristic and phenotypic and genotypic correlations of plant height with each measured characteristic in the F6-generation of cross 1 in the GPX_SATO experiment.

		Correlation with plant height	
Parameter	**Heritabilities**	**Phenotypic correlation**	**Genotypic correlation**
SPAD	0,88	-0,46**	-0,51**
Maturity	0,89	0,73**	0,77**
Plant height	0,89	0	0
Pod set score	0,27	-0,38**	-0,67**
Thousand seed weight	0,87	-0,55**	-0,58**
Protein content	0,86	0,16	0,20*
Oil content	0,64	-0,09	-0,1
Protein+Oil content	0,49	0,16	0,28*
Sucrose content	0,70	-0,3	-0,36**

Heritabilities: Heritability of each measured trait for the F6-generation of cross 1

Correlation with plant height:
Phenotypic correlation between plant height and each measured characteristic: ** : significant on the 1% level; *: significant on the 5% level.
Genotypic correlation between plant height and each measured characteristic. Significance expressed as being greater than once (*) or twice (**) its standard error

Cross 1 possesses a high heritability for plant height, however the selection on increased plant height could result in a negative response of sucrose content, as the genotypic correlation is negative and significant.

5.3.2.Cross 2

Table 5.4 gives the information in order to answer RQ1, RQ2 and RQ3 for cross 2:

For the protein content, 33 genotypes have a significantly higher protein content in comparison with the adapted parent Essor, meaning, improvement in the protein content compared to the adapted parent can be noticed (RQ1). For 11 genotypes, significant improvement in protein content was obtained even if those genotypes were selected for low protein content. Moreover no genotypes have worse protein content than Essor.
For the sucrose content, no genotypes show significant improvement in comparison with the adapted parent Essor (RQ2). This also results in the fact that no genotypes show improvement in both sucrose and in protein content compared to the adapted parent (RQ3).

Table 5.7 gives the information in order to answer RQ4 for cross 2:

Table 5.7 LS-means, F-test and correlations related to the selection on protein for each measured characteristics in the F6 generation of cross 2 in the GPX_SATO experiment:

		LS-means		F-test	Correlation with protein	
Parameter	**Unit**	**Low protein lines**	**High protein lines**	**Significance Level**	**Phenotypic correlation**	**Genotypic correlation**
SPAD	SPAD-value	40,5	41,7	0,0017	0,03	0,03
Maturity	days after 31/07/2011	41,8	41,1	0,4490	0,21	0,25*
Plant height	cm	89	85	0,1408	0,11	0,19*
Pod set score	score (1 = excellent; 5 = very weak)	3	3	0,0227	0,11	-0,03
Thousand seed weight	g	169	169	0,9317	-0,36*	-0,40**
Protein content	g/kg	448	466	<0,0001	0,00	0,00
Oil content	g/kg	179	169	0,0016	-0,75**	-0,89**
Protein+Oil content	g/kg	627	635	0,0057	0,67**	0,78**
Sucrose content	g/kg	48	43	<0,0001	-0,89**	-0,95**

LS-means: LS-means for each characteristic for the lines selected for low and high protein content respectively.
F-test: Significance level for the F-test, between the LS-means of the lines selected for low and high protein content, performed for each measured characteristic respectively.
Correlation with protein:
Phenotypic correlation between protein and each measured characteristic: ** : significant on the 1% level; *: significant on the 5% level.
Genotypic correlation between protein and each measured characteristic. Significance expressed as being greater than once (*) or twice (**) its standard error.

As expected the selection on protein had an effect on protein level as the LS-means for the low and high protein lines significantly differ at the 1% level.

It seems that protein negatively effects thousand seed weight, oil content and sucrose content as the respective phenotypic and genotypic correlations are negative and significant (thousand seed weight: 5% and ** significance; oil and sucrose content: 1% and ** significance). For the thousand seed weight this is not confirmed by the comparison of LS-means as the high and low protein lines do not differ significantly. For oil and sucrose

content, the negative correlations are in accordance with the higher LS-means for oil and sucrose content found in the low protein lines compared to the high protein lines.

Apparently, protein could have a positive effect on maturity and plant height (genotypic correlation is positive at the * significance level), however this is not confirmed in the comparison between the low and high protein lines (no significant difference between the respective LS-means at the 5% level).

Protein content also positively effects protein+oil content. This is demonstrated by the positive and significant phenotypic and genotypic correlations (1% and ** significance). Moreover significant difference is observed between the LS-means of the low and the high protein lines (1% significance). As already mentioned in cross 1, this is due to the inclusion of protein content in the calculation of protein+oil content.

As higher maturity is not a desirable trait for soybeans destined for growth in Austria (early maturity is required). Hence, the possibility for producing improved genotypes by selecting on longer maturity will not be investigated. However the possibility of including the selection on an enhanced plant height is investigated in following table (RQ5 is investigated).

Table 5.8 Heritability of each measured characteristic and phenotypic and genotypic correlations of plant height with each measured characteristic in the F6-generation of cross 2 in the GPX_SATO experiment.

	Heritabilities	Correlation with plant height	
Parameter		**Phenotypic correlation**	**Genotypic correlation**
SPAD	0,90	-0,49**	-0,54**
Maturity	0,88	0,62**	0,70**
Plant height	0,83	0	0
Pod set score	0,62	-0,28	-0,35*
Thousand seed weight	0,92	-0,22	-0,24*
Protein content	0,77	0,11	0,19*
Oil content	0,56	0	-0,09
Protein+Oil content	0,60	0,19	0,24*
Sucrose content	0,78	-0,22	-0,25*

Heritabilities: Heritability of each measured trait for the F6-generation of cross 2

Correlation with plant height:
Phenotypic correlation between plant height and each measured characteristic: ** : significant on the 1% level; *: significant on the 5% level.
Genotypic correlation between plant height and each measured characteristic. Significance expressed as being greater than once (*) or twice (**) its standard error

Although cross 2 possesses a high heritability for plant height, selection for enhanced plant height could result into a decrease in sucrose content (negative genotypic correlation at the * significance level).

5.3.3.Cross 3

Table 5.4 gives the information in order to answer RQ1, RQ2 and RQ3 for cross 3:

For the protein content, no genotypes show significant improvement of protein content in comparison with the adapted parent Essor (RQ1). No genotypes show significant decrease in protein content.
For the sucrose content, no genotypes show significant improvement for sucrose content in comparison with the adapted parent Essor (RQ2). Moreover, six genotypes show significant decrease in sucrose content.
No genotypes show significant improvement in sucrose content and in protein content in comparison with the adapted parental genotype (RQ3).

Table 5.9 gives the information in order to answer RQ4 for cross 3:

Table 5.9 LS-means, F-test and correlations related to the selection on protein for each measured characteristics in the F6 generation of cross 3 in the GPX_SATO experiment:

		LS-means		F-test	Correlation with protein	
Parameter	**Unit**	**Low protein lines**	**High protein lines**	**Significance Level**	**Phenotypic correlation**	**Genotypic correlation**
SPAD	SPAD-value	40,3	41,1	0,0523	0,09	0,25
Maturity	days after 31/07/2011	43,7	42	0,0725	0,06	0,04
Plant height	cm	84	80	0,0641	0,13	0,43*
Pod set score	score (1 = excellent; 5 = very weak)	3,3	4	0,0059	0,50**	0,87**
Thousand seed weight	g	183	197	0,0003	0,23	0,50*
Protein content	g/kg	417	428	0,0014	0,00	0,00
Oil content	g/kg	193	188	0,0195	-0,51**	0,01
Protein+Oil content	g/kg	611	615	0,0974	0,73**	0,91**
Sucrose content	g/kg	54	53	0,4988	-0,49**	-0,52*

LS-means: LS-means of each characteristic for the lines selected for low and high protein content respectively.
F-test: Significance level for the F-test, between the LS-means of the lines selected for low and high protein content, performed for each measured characteristic respectively.
Correlation with protein:
Phenotypic correlation between protein and each measured characteristic: ** : significant on the 1% level; *: significant on the 5% level.
Genotypic correlation between protein and each measured characteristic. Significance expressed as being greater than once (*) or twice (**) its standard error.

As expected the selection on protein had an effect on protein level as the LS-means for the low and high protein lines significantly differ at the 1% level.

The selection on protein should negatively effect the selection on sucrose content as the phenotypic and genotypic correlations are negative and significant (respectively 1% and ** significance). However the low protein lines did not possess significantly higher LS-means for sucrose content compared to the high protein lines. In this cross, only the phenotypic correlation for oil content is negative and significant (1% significance). The comparison

between the LS-means for oil content resulted in higher LS-means for the low protein lines compared to the high protein lines.

The selection on protein apparently results in a higher pod set score (thus a lower pod set),. This is demonstrated by the positive phenotypic and genotypic correlation between protein content and pod set score, as well as, the significantly higher pod set score in the low protein lines compared to the high protein lines. Higher pod set score is not a desirable trait (as it could result in a lower yield), a selection-strategy based on higher pod set score is therefore not investigated.
The genotypic correlations between protein content and plant height and thousand seed weight respectively are both positive and significant (* significance). For plant height, this possible positive relationship is not reflected in the response on the selection on protein (no significant difference was observed between the LS-means). In the case of thousand seed weight, the response on the selection on protein confirms the possible positive relationship between protein and thousand seed weight; the high protein lines possess higher LS-mean for thousand seed compared to the low protein lines.

A selection-strategy based on the selection of plant height and/or thousand seed weight is investigated in the following table (RQ5 is investigated).

Table 5.10 Heritability of each measured characteristic and phenotypic and genotypic correlations of plant height and thousand seed weight with each measured characteristic respectively, in the F6-generation of cross 3 in the GPX_SATO experiment.

	Heritabilities	Correlation with plant height		Correlation with thousand seed weight	
Parameter		Phenotypic correlation	Genotypic correlation	Phenotypic correlation	Genotypic correlation
SPAD	0,82	-0,38**	-0,45**	0,35*	0,36**
Maturity	0,85	0,65**	0,73**	-0,36*	-0,44**
Plant height	0,87	0	0	-0,47**	-0,49**
Pod set score	0,65	-0,22	-0,28*	0,32*	0,45**
Thousand seed weight	0,92	-0,47**	-0,49**	0	0
Protein content	0,25	0,13	0,43*	0,23	0,50*
Oil content	0,23	-0,09	-0,23	-0.03	-0,15
Protein+Oil content	0,60	0,07	0,18	0,20	0,25*
Sucrose content	0,61	-0,46**	-0,60**	0,41**	0,47**

Heritabilities: Heritability of each measured trait for the F6-generation of cross 3

Correlation with plant height:
Phenotypic correlation between plant height and each measured characteristic: ** : significant on the 1% level; *: significant on the 5% level.
Genotypic correlation between plant height and each measured characteristic. Significance expressed as being greater than once (*) or twice (**) its standard error.

Correlation with thousand seed weight:
Phenotypic correlation between thousand seed weight and each measured characteristic: ** : significant on the 1% level; *: significant on the 5% level.
Genotypic correlation between thousand seed weight and each measured characteristic. Significance expressed as being greater than once (*) or twice (**) its standard error.

Including the selection on increased plant height in a selection-strategy would result in decreased sucrose content according to the negative phenotypic and genotypic correlations of plant height with sucrose content (respectively significant at the 1% level and ** significance). However applying selection on thousand seed weight could be interesting as cross 3 possesses high heritability for thousand seed weight and positive phenotypic and genotypic correlations between thousand seed weight and sucrose content, and protein content respectively. Hence, selection on thousand seed weight should result in a good response combined of thousand seed weight with a positive effect on protein and sucrose content.

5.3.4.Cross 4

Table 5.4 gives the information in order to answer **RQ1**, **RQ2** and **RQ3** for cross 4:

For the protein content, 29 genotypes show significant improvement in protein content in comparison with the adapted parent GL601 (**RQ1**). It is noteworthy that 15 genotypes that were selected for low protein, still show improvement compared to GL601. Moreover, no genotypes show significant decrease of protein content compared to GL601.
For the sucrose content, three genotypes show improvement as they have a significantly higher sucrose content in comparison with the adapted parent GL601 based on the least significant difference for sucrose (**RQ2**). It is noteworthy that those three lines were selected for high protein content.

One genotype, the genotype GP7X-1675, possesses an improved protein content and sucrose content in comparison to its adapted parental genotype (**RQ3**). In table 5.11, the displayed values of each measured traits are given for GP7X-1675 and compared with the adapted parent GL601 and with the LS-means of all $F_{3:6}$-lines from cross 4.
GP7X-165 possesses significantly higher SPAD and sucrose content compared to the LS-means of all $F_{3:6}$-lines. It also possesses lower maturity, this is desirable, as in Austria early mature soybeans are grown.
Considering the improvement of GP7X-165 compared to its adapted parent GL601, it can be noticed that GP7X-165, above improved protein and sucrose content, also possesses improved thousand seed weight compared to GL601. GP7X-165 seeds possess lower oil content than the seeds of GL601.Improved protein and sucrose content and lower oil content are desirable traits for food grade soybeans, large seed size is desirable for large-seeded soybeans to make tofu and soymilk, miso, edamame and soynuts. In other words GP7X-1675 seems to be a promising genotype.

Table 5.11 Comparison of the values of genotype GP7X-1675 with its adapted parent GL601, for each measured characteristic in the F6 generation of the GPX_SATO experiment.

		LS-means			Comparison towards GL601		Comparison towards cross 4	
Parameter	**Unit**	**GP7X-1675**	**GL601**	**Cross 4**	**>LSD (0=False; 1=True)**	**<LSD (0=False; 1=True)**	**>LSD (0=False; 1=True)**	**<LSD (0=False; 1=True)**
SPAD	SPAD-value	44,5	43,4	42	0	0	1	0
Maturity	days after 31/07/2011	34,3	37,9	41,9	0	0	0	1
Plant height	cm	74	75	83	0	0	0	0
Pod set score	score (1 = excellent; 5 = very weak)	3	2	3,1	1	0	0	0
Thousand seed weight	g	198	176	189	1	0	0	0
Protein content	g/kg	413	392	415	1	0	0	0
Oil content	g/kg	187	218	189	0	1	0	0
Protein+Oil content	g/kg	599	611	604	0	0	0	0
Sucrose content	g/kg	61	54	55	1	0	1	0

LS-means: LS-means of each characteristic for GP7X-1675, GL601 and all $F_{3:6}$-lines of cross 4, respectively.

Comparison towards GL601:

>LSD: Improvement of GP7X-1675 for each measured characteristic compared to GL601, based on the least significant difference at the 5% level for the corresponding characteristic. 1: the difference between GP7X-1675 and GL601 is greater than the corresponding LSD. 0: the difference between GP7X-1675 and GL601 is smaller or equal to the corresponding LSD.

<LSD: Decrease of GP7X-1675 for a trait compared to GL601, based on the least significant difference at the 5% level for the corresponding characteristic. 1: the difference between GL601 and GP7X-1675 is greater than the corresponding LSD. 0: the difference between GL601 and GP7X-1675 is smaller or equal than the corresponding LSD.

Comparison towards cross 4:

>LSD: Improvement of GP7X-1675 for each measured characteristic compared to the LS-means of all $F_{3:6}$-lines of cross 4, based on the least significant difference at the 5% level for the corresponding characteristic. 1: the difference between GP7X-1675 and cross 4 is greater than the corresponding LSD. 0: the difference between GP7X-1675 and cross 4 is smaller or equal to the corresponding LSD.

<LSD: Decrease of GP7X-1675 for a trait compared to the LS-means of all $F_{3:6}$-lines of cross 4, based on the least significant difference at the 5% level for the corresponding characteristic. 1: the difference between cross 4 and GP7X-1675 is greater than the corresponding LSD. 0: the difference between cross 4 and GP7X-1675 is smaller or equal than the corresponding LSD.

Table 5.12 gives the information in order to answer RQ4 for cross 4:

Table 5.12 LS-means, F-test and correlations related to the selection on protein for each measured characteristics in the F6 generation of cross 4 in the GPX_SATO experiment:

		LS-means		F-test	Correlation with protein	
Parameter	**Unit**	**Low protein lines**	**High protein lines**	**Significance Level**	**Phenotypic correlation**	**Genotypic correlation**
SPAD	SPAD-value	41,6	42,5	0,0056	-0,08	-0,15
Maturity	days after 31/07/2011	42,6	41,2	0,0631	-0,01	0,00
Plant height	cm	88	79	0,0013	0,00	0,01
Pod set score	score (1 = excellent; 5 = very weak)	3,2	3	0,1075	0,03	0,04
Thousand seed weight	g	185	192	0,0267	0,16	0,23*
Protein content	g/kg	415	415	0,859	0,00	0,00
Oil content	g/kg	191	186	0,1659	-0,64**	-0,67**
Protein+Oil content	g/kg	606	601	0,0516	0,36*	0,47**
Sucrose content	g/kg	55	56	0,0266	-0,58**	-0,66**

LS-means: LS-means for the lines selected for low and high protein content respectively.
F-test: Significance level for the F-test, between the LS-means of the lines selected for low and high protein content, performed for each measured characteristic respectively.
Correlation with protein:
Phenotypic correlation between protein and each measured characteristic: ** : significant on the 1% level; *: significant on the 5% level.
Genotypic correlation between protein and each measured characteristic. Significance expressed as being greater than once (*) or twice (**) its standard error.

It is noteworthy that selection on protein content seems not to have an effect on protein content in the F6-generation as the LS-means of the high and low protein lines do not differ significantly at the 5% level.

Cross 4 possesses negative and significant phenotypic and genotypic correlations between protein and oil content, and sucrose content, respectively (both 1% and ** significance for the phenotypic and genotypic correlations respectively). However no significant difference between the LS-means of the high and low protein lines for oil content was found (5% level). For sucrose content, the LS-mean of the high protein lines was higher than the LS-mean of the low protein lines (5% level), which is not expected considering the negative correlations. Possibly, the selection on protein content which was not resulting into a response of the protein content could be the cause of these phenomena.

The phenotypic and genotypic correlation of protein with the protein+oil content are positive and significant (respectively 5% and * significance). No significant difference was noticed between the LS-means for protein+oil content of the selected high and the low protein lines. This possible positive relationship is not further investigated for the same reasons as in cross 1, cross 2 and cross 3.

As the genotypic correlation between protein and thousand seed weight was positive and significant. It is interesting to investigate if selection for thousand seed weight could be included in a new selection-strategy in order to obtain improved genotypes with cross 4 (RQ5 is investigated). The genotypic and phenotypic correlation of thousand seed weight is displayed in table 5.13. The heritabilities of cross 4 for each trait is also included.

Table 5.13 Heritability of each measured characteristic and phenotypic and genotypic correlations of thousand seed weight with each measured characteristic in the F6-generation of cross 4 in the GPX_SATO experiment.

	Heritabilities	Correlation with thousand seed weight	
Parameter		Phenotypic correlation	Genotypic correlation
SPAD	0,76	0,07	0,03
Maturity	0,83	0,03	0,02
Plant height	0,84	-0,47**	-0,54**
Pod set score	0,78	0,57**	0,65**
Thousand seed weight	0,94	0	0
Protein content	0,63	0,16	0,23*
Oil content	0,43	-0,01	-0,05
Protein+Oil content	0,58	0,16	0,22*
Sucrose content	0,58	0,15	0,20*

Heritabilities: Heritability of each measured trait for the F6-generation of cross 4

Correlation with protein:
Phenotypic correlation between thousand seed weight and each measured characteristic: ** : significant on the 1% level; *: significant on the 5% level.
Genotypic correlation between thousand seed weight and each measured characteristic. Significance expressed as being greater than once (*) or twice (**) its standard error

The genotypic correlation between sucrose and thousand seed weight is positive and significant (* significance) as well as the genotypic correlation of thousand seed weight with protein. Moreover the heritability for thousand seed weight is high. Hence, selection on thousand seed weight should result in a good response combined of thousand seed weight with a positive effect on protein and sucrose content.

5.3.5. Hypothesis 2

Similarities are observed between some crosses: the similarities in those crosses are related to the fact that they possess the same high-protein pollinator.

RQ6: Are there similarities in traits between crosses with one common parent (different mother and same pollinator; same mother and different pollinator)?

RQ7: Are there similar patterns of indirect selection between crosses with one common parent (different mother and same pollinator; same mother and different pollinator)?

In order to answer RQ6 table 5.14 displays the LS-means of the 4 different crosses. Additionally table 5.15 displays the F-test between the LS-means of cross 1 and cross 2 which were performed with the same pollinator (GF4X-15-2), cross 2 and cross 3 performed with the same mother (Essor) and cross 3 and cross 4 performed with the same pollinator (Vinton 81).

Table 5.14 The LS-means of each measured characteristic in the F6 generation of cross 1, cross 2, cross 3 and cross 4 of the GPX_SATO experiment.

		LS-means			
Parameter	**Unit**	**Cross 1**	**Cross 2**	**Cross 3**	**Cross 4**
SPAD	SPAD-value	41,7	41,1	40,7	42
Maturity	days after 31/07/2011	39,3	41,5	42,8	41,9
Plant height	cm	84	87	82	83
Pod set score	score (1 = excellent; 5 = very weak)	2,9	3,1	3,5	3,1
Thousand seed weight	g	172	169	190	189
Protein content	g/kg	455	457	423	415
Oil content	g/kg	167	174	191	189
Protein+Oil content	g/kg	623	631	613	604
Sucrose content	g/kg	47	45	53	55

LS-means: LS-means of each measured characteristics for all the genotypes and their replicates in the F6-generation of respectively cross 1, cross 2, cross 3 and cross 4.

Table 5.15 Comparison of the LS-means of each measured characteristic in the F6 generation between cross 1 and cross 2, cross 2 and cross 3 and cross 3 and cross 4 from the GPX_SATO experiment.

	F-test		
Parameter	**Cross 1 - Cross 2**	**Cross 2 - Cross 3**	**Cross 3 - Cross 4**
SPAD	0,0483	0,1651	<0,0001
Maturity	0,0008	0,0389	0,1278
Plant height	0,1432	0,0034	0,3397
Pod set score	0,0018	0,0001	<0,0001
Thousand seed weight	0,3066	<0,0001	0,5875
Protein content	0,4534	<0,0001	0,0004
Oil content	0,0035	<0,0001	0,2714
Protein+Oil content	<0,0001	<0,0001	<0,0001
Sucrose content	0,0169	<0,0001	0,0005

F-test: Significance level for the F-test, between the LS-means of cross 1 and cross 2, the LS-means of cross 2 and cross 3 and the LS-means of cross 3 and cross 4, performed respectively for each measured characteristics.

Although cross 2 and cross 3 possess the same mother (Essor), their progenies differ significantly for all traits but SPAD. Considering the comparison between cross 1 and cross 2, they have common pollinator (GF4X-15-2), the progenies of those crosses are similar for, plant height, thousand seed weight and protein content. This trend can also be observed in cross 3 and cross 4 they have common pollinator (Vinton 81) and their progenies also possess similar traits that are maturity, plant height, thousand seed weight and oil content.

Furthermore, although significant differences are observed, it is clear that the progenies of cross 1 and cross 2 display a higher range of protein content (respective means 455 g/kg and 457 g/kg) compared to the progenies of cross 3 and 4 (respective means 423 g/kg and 415 g/kg). Additionally, the progenies of cross 1 and cross 2 possess lower range (respective means 47 g/kg and 45 g/kg) for sucrose content compared to cross 3 and cross 4 (respective means 53 g/kg and 55 g/kg).

Table 5.16 displays the information in order to answer question RQ7:

Table 5.16 The phenotypic and genotypic correlations between protein and each measured characteristic in the F6 generation of the 4 crosses made in the GPX_SATO experiment.

	Phenotypic correlation with protein				Genotypic correlation with protein			
Parameter	**Cross 1**	**Cross 2**	**Cross 3**	**Cross 4**	**Cross 1**	**Cross 2**	**Cross 3**	**Cross 4**
SPAD	0,07	0,03	0,09	-0,08	0,06	0,03	0,25	-0,15
Maturity	0,14	0,21	0,06	-0,01	0,15	0,25*	0,04	0,00
Plant height	0,16	0,11	0,13	0,00	0,20*	0,19*	0,43*	0,01
Pod set score	-0,02	0,11	0,50**	0,03	-0,12	-0,03	0,87**	0,04
Thousand seed weight	-0,13	-0,36*	0,23	0,16	-0,12	-0,40**	0,50*	0,23*
Protein content	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Oil content	-0,77**	-0,75**	-0,51**	-0,64**	-0,94**	-0,89**	0,01	-0,67**
Protein+Oil content	0,56**	0,67**	0,73**	0,36*	0,69**	0,78**	0,91**	0,47**
Sucrose content	-0,83**	-0,89**	-0,49**	-0,58**	-0,91**	-0,95**	-0,52*	-0,66**

Phenotypic correlation:
Phenotypic correlations between protein and each measured characteristic respectively for cross 3 and cross 4: ** : significant on the 1% level; *: significant on the 5% level.

Genotypic correlation:
Genotypic correlation between protein and each measured characteristic. Significance expressed as being greater than once (*) or twice (**) its standard error.

Similarities in indirect selection seem to occur between crosses with same pollinator:
Cross 1 and cross 2 both have strong negative phenotypic and genotypic correlations between protein and oil content and between protein and sucrose content.
Cross 3 and cross 4 seem to share similar phenotypic and genotypic correlation of protein with sucrose content and protein content with thousand seed weight. Notice, that the previously mentioned correlations are completely different between cross 2 and cross 3.

5.4. Discussion

5.4.1. Hypothesis 1

By crossing two cultivars with contrasting protein content and by applying selection on protein content, it is possible to produce improved food grade soybean genotypes.

5.4.1.1. Research question 1

Is a genotype with higher protein content than its adapted standard parental genotype obtained in the F6 generation?

Individually for cross 1, 2 and 4, a great number of genotypes are improved in protein content compared to their adapted parent in the F6-generation. This is in accordance with a study performed by Cober and Voldeng [42] in which a cross was performed between a high-protein cultivar[6] and an adapted high-yielding cultivar[7]. Selection was performed for high protein content in the F4-generation and the F5-generation and additionally selection for high protein content and high yield was performed in the F6-generation. They observed a general improvement of the protein content in comparison with the adapted parent in the F6-generation and the F7-generation.

The observed improvement obtained by the selection on protein can possibly be explained by the heritabilities for protein content. A large heritability for protein would mean that a large part of the phenotypic variability can be explained by genotypic variability, which makes improvement of protein content by selection for protein easier.
Broad sense heritabilities of 0.77 in the F6-generation and 0.74 in the F7-generation were found by Cober and Voldeng [42]. Jaureguy et al. [7] studied a population originating from a cross between a high-protein cultivar and a large-seeded, moderately high sugar food grade cultivar[8] that was randomly selected in the F4 generation. They reported broad sense heritabilities ranging from 0.46 in the F6-generation, 0.74 and 0.86 in the F7-generation in different locations.
The heritabilities calculated in the F6-generation of the GPX_SATO experiment were considered as narrow-sense as they were calculated in an advanced generation (F6-generation); at this point most loci are homozygous meaning dominance effects are negligible. Similarly, the broad-sense heritabilities of the studies of Cober and Voldeng [42] and Jaureguy et al. [7] were calculated in advanced generations. Following this reasoning it is valuable to compare those heritabilities with the heritabilities of the GPX_SATO experiment. The heritabilities of cross 1, 2 and 4 (cross 1: 0.86, cross 2: 0.77 and cross 4: 0.63) of the GPX_SATO experiment were in the same range as those found by Cober and Voldeng [42] and Jaureguy et al. [7].
In comparison cross 3, obviously has an extremely low heritability, 0.25. The fact that, for this cross, no improved genotypes were found can be explained by this low heritability, as only a small fraction of genetic variability influences the phenotypic variability. Furthermore, it is noteworthy that cross 3 was the only cross performed with two parents that possess food grade characteristics. Possibly, this resulted into a smaller genetic variability for some traits, such as protein content, leading to a difficult improvement.

---

[6] protein content of 521 g/kg and yield of 3341 kg/ha. (based on dry matter)

[7] protein content of 457 g/kg and yield 3610 kg/ha. (based on dry matter)

[8] No quantitative information on the parental material was mentioned.

5.4.1.2. Research question 2

Is a genotype with higher sucrose content than its adapted standard parental genotype obtained in the F6 generation?

Individually for cross 1,2 and 3, no improvement in sucrose content compared to the adapted parent could be noticed in the F6 generation.
For the case of cross 1 and 2, this can easily be explained by the strong negative phenotypic and genotypic correlations between protein and sucrose (cross 1: -0.83** and -0.91**; cross 2: -0.89** and -0.95**). For the sake of comparison, Jaureguy et al. [7] in their experiment, reported a correlation of -0.68**. Wilcox and Shibles [43] performed a study in which a cross was made between a cultivar with normal protein content[9] and a high-protein cultivar[10] and random selection was applied in the F4 generation. They reported correlation of -0.66**. The improvement in protein content in cross 1 and 2, with dramatically negative correlations between protein and sucrose resulted into the fact that no improvement of sucrose was obtained.
The phenotypic and genotypic correlations between protein and sucrose in cross 3 (-0.49** and -0,52**) are not that negative compared to cross 1 and 2. Therefore an improvement of sucrose together with protein improvement should theoretically be more probable. However, as the protein heritability of cross 3 is low, improvement of protein is difficult to obtain by protein selection. As protein improvement did not occur, the correlation between protein and sucrose did not affect the sucrose content.

In three genotypes of cross 4, improvement was achieved in sucrose content in comparison with the adapted parent. This could be the result of medium phenotypic and genotypic correlations between protein and sucrose (-0.58** and -0.66**). As protein improvement was achieved, sucrose improvement could therefore be moderately broken. Moreover, this improvement can partially be explained by genotypic effects as the sucrose heritability is 0.58 for in the F6-generation. In comparison with the heritabilities found by Jaureguy et al. [7] (0.33 in the F6-generation, 0.57 and 0.60 in F7-generation in different location), the sucrose heritability of cross 4 is similar.

5.4.1.3. Research question 3

Is a genotype with higher protein content and higher sucrose content than its adapted standard parental genotype obtained in the F6 generation?

For cross 1, 2 and 3 no genotype with both protein and sucrose improvement was found.

In cross 4, one genotype was found with protein and sucrose content enhanced, compared to the adapted parent GL601. This genotype, GP7X-1675, was found among 48 genotypes, the efficiency of the breeding was therefore 1/48, so around 2%. It is noteworthy that the three genotypes with improved sucrose content in cross 4, were only found among the 24 high protein lines. Moreover the high protein lines possess significantly higher LS-mean for sucrose content compared to the low protein lines. Hence, there can be supposed that producing genotypes with improved protein and sucrose content is possible by selecting only for high protein content, this should result in a breeding efficiency of 1/24, so around 4%.
Beyond improved protein and sucrose content, GP7X-1675 possesses improved thousand seed weight and lower oil content compared to GL601. Improved protein and sucrose content and lower oil content are desirable traits for food grade soybeans large seed size is

[9] protein content 397 g/kg; oil content 213 g/kg (based on dry matter)

[10] protein content 475 g/kg; oil content 184 g/kg (based on dry matter)

desirable for large-seeded soybeans to make tofu and soymilk, miso, edamame and soynuts. According to Chen [35] protein content for soybeans ranges from 33.1 to 55.9% and a moderately high protein content ranges between 43 and 45%. The protein content of GP7X-1675, 41.3%, is slightly lower than the medium range, however, it is far from the lowest concentration found 33.1%. Still according to Chen [35], sucrose content of soybeans ranges from 1.5 to 10.2%, the seeds of GP7X-1675 possess 6.1% sucrose, a medium high sucrose content. Chen [35] further reports that the oil content of soybeans ranges from 13.6 to 23.6%, the GP7X-1675 seeds with oil content of 18.7%, thus possess a medium high oil content. The seed size of GP7X-1675 is 19.8g/100-seeds would suit for the production of tofu and soymilk, miso, edamame and soynuts, as seed size of about 20g/100 seeds are required according to Chen [35].
Overall there can be stated that GP7X-1675 possesses improved desirable traits compared to GL601, furthermore those traits are in the required range in order to develop a food grade cultivar. GP7X-1675 is therefore a promising genotype for the development of improved food grade cultivars.

5.4.1.4. Research question 4

Which traits are indirectly affected by selection for protein? Would an other selection-strategy be useful for the production of improved food grade soybeans?

Oil content
The existence of a strong negative correlation between protein and oil is well known [1]. This was previously reported by Burton in Wilcox and Shibles [43] in populations with considerable variability among lines for protein and oil concentration. Wilcox and Shibles [43] reported a correlation between protein and oil of -0.88** and similarly Cober et al. [42] reported correlation of -0.84** in their experiment.

The correlations of cross 1 (phenotypic: -0.77**; genotypic: -0.94**) and cross 2 (phenotypic: -0.75**; genotypic: -0.89**) are in accordance with the negative relationship of protein and oil and are in the same range as those reported by Wilcox and Shibles [43] and Cober et al. [42]. This relationship in cross 1 and 2 is confirmed by the significant differences between the LS-means of the low and high protein lines.

In cross 3, significant difference between the LS-means is observed, however only the phenotypic correlation is significantly negative. Furthermore this cross possesses low protein heritability, thus deductions concerning indirect selection in this cross must be taken with caution.

In the case of cross 4, although the correlations between protein and oil are significantly negative, no significant difference between the oil content of the low and high-protein lines was found. This is possibly due to the fact that there was no response to the selection of protein (no significant difference in protein content was observed between the low and the high lines).

Sucrose content
The strong negative relationship between protein and sucrose content in cross 1 and 2 was already treated in RQ2. It is noteworthy that in cross 1 the significance level is 5.05% between the sucrose content LS-means of the low and high-protein lines, which means that strictly, there is no significant difference on the 5% level. However, there can still be quite safely stated that for cross 1 and 2, indirect selection against sucrose occurs, as no improvement of sucrose was obtained by protein selection.

As already stated in RQ2 the phenotypic and genotypic correlations between protein and sucrose in cross 3 (-0.49** and -0.52**) are not that negative compared to cross 1 and 2. Therefore an improvement of sucrose together with protein improvement should theoretically be possible. However, because of the low heritability of protein, deductions about indirect selection in this cross must be taken with caution.

For cross 4, the phenotypic and genotypic correlation are moderately negative. However, the high-protein lines possess higher LS-means for sucrose content than the low-protein lines. This is possibly due to the fact that there was no response to the selection of protein (no significant difference in protein content was observed between the low and the high lines).

Thousand seed weight
Geater and Fehr in Jaureguy et al. [7] reported a negative correlation between seed protein and seed size of -0.67, whereas Jaureguy et al. [7] found a smaller correlation of -0.32.

Similarly to the study of Jaureguy et al. [7] negative phenotypic and genotypic correlations were found for cross 2 (-0.36* and -0.40**). Probably, because these correlations are modest, LS-means of low and high-protein lines did not differ significantly for thousand seed weight. It can therefore be deduced that there is moderate negative selection of thousand seed weight by the selection on protein.

Cross 4 possesses a slight positive genotypic correlation between protein content and thousand seed weight. However, the phenotypic correlation was not significant and the significant difference between the LS-means of the low and high-protein lines can not be used to assess possible indirect selection, as no selection response for protein content was observed.

Cross 3 possesses moderate positive genotypic correlation between protein content and thousand seed weight. However, because of the low heritability of protein, deductions about indirect selection in this cross must be taken with caution.

For cross 1 no significant correlations could be found between protein and thousand seed weight.

5.4.1.5. Research question 5

Cross 3 and 4 have significantly positive genotypic correlations between thousand seed weight and protein content (see RQ4). For this reason the possibility of enhancing simultaneously protein and sucrose content was investigated. The genotypic correlations between thousand seed weight and sucrose content of cross 3 and cross 4 (phenotypic, cross 3: 0.41**; phenotypic cross 4: 0.15; genotypic, cross 3: 0.47**; genotypic, cross 4: 0.20*) were positive and significant. Jaureguy et al.[7] found a non-significant correlation of 0.20, hence the correlations found in cross 3 and cross 4 can be considered as moderately high. In other words, the selection on thousand seed weight combined with selection on protein and/or sucrose content could enhance the probability of finding genotypes improved

in protein and sucrose content.
Especially for cross 3, a new selection strategy should favor selection on sucrose content as the heritability for protein content was too low to obtain protein improvement by protein selection (see RQ1). The heritability for sucrose content is moderately high and sucrose content possesses a not too dramatically negative correlation with protein content (see RQ2 and RQ4). Selection for sucrose (eventually combined with selection for thousand seed weight) should therefore theoretically enable to obtain genotypes improved in sucrose content simultaneously with protein content, without testing a too big number of genotypes.
The selection on protein for cross 4 resulted into improved genotypes in protein and one genotype improved in protein and sucrose content. However, there was no response on the selection of protein (see RQ4). Therefore, selection based on sucrose content (eventually combined with selection for thousand seed weight) should be investigated.

5.4.1.6. Accepted or refused?

By crossing two cultivars with contrasting protein content and by applying selection on protein content, it is possible to produce improved food grade soybean genotypes.

For cross 1 and cross 2 no improved genotypes were found. Furthermore selection for protein resulted in strong negative response for sucrose, hence the probability of obtaining improved genotypes with those crosses and by selecting for protein content is low. Therefore hypothesis 1 is not accepted for cross 1 and cross 2.

For cross 3 no improved genotypes were found. Moreover it seems that obtaining an improved genotype by selection for protein is unlikely probably because of of small genetic variability. Therefore hypothesis 1 is not accepted for cross 3. However, it seems possible to obtain improved genotypes by selection on sucrose. Additionally, the selection for enhanced thousand seed weight could be helpful to increase probabilities of finding improved genotypes.

For cross 4, one improved genotype was found. Furthermore, this genotype had promising food grade traits and it should be further investigated, e.g. by performing yield testing and measuring isoflavone content. Obtaining improved genotypes by protein selection seems fairly possible with cross 4, a breeding efficiency of about 2% was obtained in this experiment. If selection is performed only on high protein level, the breeding efficiency could be enhanced up to about 4%. Therefore, hypothesis 1 for this cross is accepted. Additionally a selection strategy that includes selection for thousand seed weight should increase probabilities of finding improved genotypes. A combination of selection for thousand seed weight and selection on sucrose content, rather than on protein content, should be investigated.

5.4.2. Hypothesis 2

5.4.2.1. Research question 6

Are there similarities in traits between crosses with one common parent (different mother and same pollinator; same mother and different pollinator)?

In the GPX_SATO experiment, the crosses with pollinator GF4X-15-2 resulted in progenies with similar plant height, thousand seed weight and protein content. The crosses with pollinator Vinton 81 resulted in progenies similar for maturity, plant height, thousand seed weight and oil content. Moreover progenies from crosses with same pollinators displayed protein and sucrose content in same range order. In contrast to that crosses with same mother plant Essor showed no similarities among progenies.

The previous points seem to indicate that similarities in traits are observed between crosses with same pollinator. This is partially in accordance with a study of Scott and Kephart [44] in which crosses were made between unrelated adapted cultivars and plant introductions. Similarities in yield, protein, oil and protein+oil content of the progenies between the crosses were investigated.
Crosses made with same adapted mother plant and two different pollinator, (Norsoy x PI437.666) and (Norsoy x PI36.653), resulted in progenies that were similar for yield, protein and oil content. Note that Norsoy was also crossed with a third pollinator, (Norsoy x PI297.513), but no similarities were found. When progenies of the cross (Kato x PI297.513) and (Kato x PI36.653) were compared, no similarities were found.
When PI297.513 was used for crosses with two different mothers, (Norsoy x PI297.513) and (Kato x PI297.513), similar protein and oil contents were found among progenies. When an other pollinator was used for cross with two different mothers, (Norsoy x PI36.653) and (Kato x PI36.653), no similarities among the progenies were found.

5.4.2.2. Research question 7

Are there similar patterns of indirect selection between crosses with one common parent (different mother and same pollinator; same mother and different pollinator)?

The similarities in correlations found between cross 1 and cross 2 and between cross 3 and cross 4 seemed to be related with their respective pollinator. Similar trends in correlations were not found in the study of Scott and Kephart [44].

5.4.2.3. Accepted or refused?

Similarities are observed between some crosses: the similarities in those crosses are related to the fact that they possess the same high-protein pollinator.

There are similarities observed in the crosses that possess same pollinator. This means that the pollinators, GF4X-21-5-2 and Vinton 81, have good breeding value in the population in which there were crossed. In other words a great deal of the effect of their genes are transferred into their progenies. Hypothesis 2 is therefore accepted.

6. Conclusion

This experiment demonstrates that crosses with a food grade high protein pollinator such as Vinton 81 result in populations more suitable for selection of food grade soybeans than crosses with a non-food grade, high-protein donor such as GF4X-21-5-2. However the use of a food grade, high protein cultivar to pollinate a mother-plant that also possesses food grade characteristics can result into low heritability for the trait used for selection, due to low genetic variability between the parents. The determination of the most efficient selection strategy is therefore decisive in order to enhance the probability of finding improved genotypes. Rather than selecting on protein content as was already often done, direct selection on sucrose content can be efficient, thanks to the easiness of measurement with NIRS. The results from the present experiment in one environment need to be confirmed by the performance of lines across different environments. Moreover, in the GPX_SATO experiment pod set score was used as an estimation of yield, but a more precise idea of yield is essential. Hence, yield testing of favorable genotypes should be performed. Finally when a cultivar seems to possess desirable agronomic traits and improved food grade characteristics, a further step is the evaluation of the quality of the product obtained from this cultivar. This is important, as the success of a food grade soybean cultivar is determined by the preferences of the consumers.

References

1. Cober, E.R., Cianzion, S. R., Pantalone, V.R, Rajcan, I., *Soybean,* in *Oil Crops,* J. Vollmann and I. Rajcan, Editors. 2009. p: 57-90.
2. Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D. L., Song, Q., Thelen, J. J., Cheng, J., Xu, D., Hellsten, U., May, G. D., Yu, Y., Sakurai, T., Umezawa, T., Bhattacharyya, M. K., Sandhu, D., Valliyodan, B., Lindquist, E., Peto, M., Grant, D., Shu, S., Goodstein, D., Barry, K., Futrell-Griggs, M., Abernathy, B., Du, J., Tian, Z., Zhu, L., Gill, N., Joshi, T., Libault, M., Sethuraman, A., Zhang, X. C., Shinozaki, K., Nguyen, H. T., Wing, R. A., Cregan, P., Specht, J., Grimwood, J., Rokhsar, D., Stacey, G., Shoemaker, R. C., Jackson, S. A., *Genome sequence of the palaeopolyploid soybean*. Nature, 2010. **463:** 178-83.
3. Singh, G., *Preface*, in *The soybean: botany, production and uses*, G. Singh, Editor 2010, CABI. p. xi-xii.
4. FAOSTAT. 2010 [cited 20-01-2012]; Available from: http://faostat.fao.org/.
5. Krishnan, H.B., *Engineering soybean for enhanced sulfur amino acid content.* Crop Science., 2005. **45:** 454-461.
6. Soyatech. *Soy facts.* [cited 20-01-2012]; Available from: http://www.soyatech.com/soy_facts.htm.
7. Jaureguy, L.M., P. Chen, and A.M. Scaboo, *Heritability and correlations among food-grade traits in soybean.* Plant Breeding, 2011. **130:** 647-652.
8. Qiu, L.-J. and R.-Z. Chang, *Origin and History of Soybean*, in *The soybean: botany, production and uses*, G. Singh, Editor 2010, CABI. p. 1-23.
9. L'Hocine, L. and J.I. Boye, *Allergenicity of soybean: new developments in identification of allergenic proteins, cross-reactivities and hypoallergenization technologies.* Critical reviews in food science and nutrition, 2007. **47:** 127-43.
10. GMOCompass. *Crops: Soybeans.* 2008 [cited 22-05-2012]; Available from: http://www.gmo-compass.org/eng/grocery_shopping/crops/19.genetically_modified_soybean.html.
11. Vollmann, J., Fritz, C.N., Wagentristl, H., Ruckenbauer, P., *Environmental and genetic variation of soybean seed protein content under Central European growing conditions.* Journal of the Science of Food and Agriculture, 2000. **80:** 1300-1306.
12. ENSA. *List of members.* [cited 22-05-2012]; Available from: http://www.ensa-eu.org/public/en/about_ensa_members_en.php.
13. Bäck, M., Köppl, H., Köppl, P., Krumphuber, C., Lembacher, F., Schally, H., Wasner, J., *Soja - Eine Kulturpflanze mit Geschichte und Zukunft.* Landwirtschaftskammer österreich, 2010.
14. Kaya, Y., Baltensperger, L., Nelson, L., Miller, J., *Maturity grouping in sunflower (Helianthus annuus L.).* Helia, 2004. **27:** 257-270.
15. Vollmann, J. and M. Menken, *Soybean: Breeding for organic farming systems,* in *Organic Crop Breeding,* E.T. Lammerts van Bueren and J.R. Myers, Editors 2012. p. 203-214.
16. Kumar, V., A. Rani, and G.S. Chauhan, *Nutritional Value of Soybean,* in *The soybean: botany, production and uses,* G. Singh, Editor 2010, CABI. p. 375-403.
17. Liu, K., *Soybeans as a Powerhouse of Nutrients and Phytochemicals*, in *Soybeans as Functional Foods and Ingredients*, K. Liu, Editor 2004, AOCS.
18. Endres, J.G., *Protein Quality and Human Nutrition*, in *Soy protein products: characteristics, nutritional aspects, and utilization. Revised and expanded edition*, J.G. Endres, Editor 2001, AOCS, p. 10-19.
19. Cui, Z., Wilson,R.F., Carter, T.E., James, A.U., Miyazaki, S., *Breeding Specialty Soybeans for Traditional and New Soyfoods*, in *Soybeans as Functional Foods and Ingredients,* K. Liu, Editor 2004, AOCS.
20. Obendorf, R.L. and S.M. Kosina, *Soluble Carbohydrates in Soybean*, in *Soybean-Biochemistry, Chemistry and Physiology*, N. Tzi-Bun, Editor 2011. p. 201-228.
21. Lühs, W. and W. Friedt, *The major oil crops*, in *Designer Oil Crops, Breeding, Processing and Biotechnology*, D.J. Murphy, Editor 1994, VCH. p. 5-71.
22. Cahoon, E.B., Clemente, T.E., Damude, H.G., Kinney, A.J., *Modifying Vegetable Oils for Food and Non-food Purposes,* in *Oil Crops* J. Vollmann and I. Rajcan, Editors 2009, Springer. p. 31-56.
23. Hornig A., L.E., Vollmann J., *Isoflavone concentration of soybean in Central Europe as determined by HPLC/UV analysis before and after acid hydrolysis.* Acta Alimentaria, 2011. **40:** 247-253.

24. Liu, K., *Soy Isoflavones: Chemistry, Processing Effects, Health Benefits, and Commercial production*, in *Soybeans as Functional Foods and Ingredients*, K. Liu, Editor 2004, AOCS.
25. ENSA. *Soy and lactose intolerance.* [cited 10-02-2012]; Available from: http://www.ensa-eu.org/public/en/ about_soy_lactose_en.php.
26. ENSA. *Soy and cow's milk protein allergy.* [cited 10-02-2012]; Available from: http://www.ensa-eu.org/public/en/about_soy_milk_en.php.
27. Europe, W.H.O. *Obesity facts and figures.* [cited 11-02-2012]; Available from: http://www.euro.who.int/en/ what-we-do/health-topics/noncommunicable-diseases/obesity/facts-and-figures.
28. ENSA. *UE & Soy Obesity.* [cited 11-02-2012]; Available from: http://www.ensa-eu.org/public/en/position_papers_en.php?topic_id=1&title=Obesity&navSwitch=.
29. FAO. *Livestock's log shadow: environmental issues and options.* 2006 [cited 01-06-2012]; Available from: ftp://ftp.fao.org/docrep/fao/010/a0701e/a0701e.pdf.
30. ENSA. *Sustainable Development.* [cited 11-02-2012]; Available from: http://www.ensa-eu.org/public/en/development_soya_en.php.
31. Soyatech. *Soyfood facts.* [cited 20-01-2012]; Available from: http://www.soyatech.com/soyfoods_facts.htm.
32. Liu, K., *Edible Soybean Products in the Current Market*, in *Soybeans as Functional Foods and Ingredients*, K. Liu, Editor 2004, AOCS.
33. Mohamed, A. and R.S. Mentreddy, *Vegetable Soybeans as a Functional Food*, in *Soybeans as Functional Foods and Ingredients*, K. Liu, Editor 2004, AOCS.
34. Nout, M.J.R. and J.L. Kiers, *Tempeh as a Functional Food*, in *Soybeans as Functional Foods and Ingredients*, K. Liu, Editor 2004, AOCS.
35. Chen, P., D*eveloping high quality identity-preserved soybean for the specialty soyfood market*, in *Production and Marketing of Identity-Preserved Soybean,* American Soybean Association: MO, USA., 2004, p. 23-31.
36. Panthee, D.R., *Varietal Improvement in Soybean*, in *The soybean: botany, production and uses*, G. Singh, Editor 2010, CABI. p. 92-112.
37. AGES. Österreichische Sortenliste 2012 für landwirtschaftliche Pflanzenarten und Gemüsearten. 2012 [cited 06-06-2012]; Available from: http://www.baes.gv.at/pflanzensorten/oesterreichische-sortenliste/.
38. Mechtler, K., *Sorteneignung für die Speisesojaproduktion (Soybean varieties for production of soya food)*, in *„Vom Lebensmittel zum Genussmittel – was essen wir morgen?“*, Arbeitsgemeinschaft für LebensMittel-Veterinär- und Agrarwesen, 2010.
39. Esteve Agelet, L. and C.R. Hurburgh Jr., *A Tutorial on Near Infrared Spectroscopy and Its Calibration*. Critical Reviews in Analytical Chemistery, 2010. **40:** 240-260.
40. Utz, H.F., *PLABSTAT - Plant Breeding Statistical Program Version 3A,* 2005, Institut für Pflanzenzüchtung, Staatgutforschung und Populationsgenetik, Universität Hohenheim: Stuttgart, Germany.
41. SAS, SAS STAT User's Guide. Release 6. 03, 1988, SAS Institue: Cary, NC, USA.
42. Cober, E.R. and H. D *Voldeng, Developing High-Protein, High-Yield Soybean Populations and Lines.* Crop Science, 2000. **40:** 39-42.
43. Wilcox, J.R. and R.M. Shibles, *Interrelationships among Seed Quality Attributes in Soybean.* Crop Science, 2001. **41:** 11-14.
44. Scott, R.A. and K.D. Kephart, *Selection for yield, protein, and oil in soybean crosses between adapted and introduced parents*. Field Crops Research, 1996. **49:** 177-185.
45. *Analysis of ash and minerals.* [cited 31-01-2012]; Available from: http://people.umass.edu/~mcclemen/581Ash&Minerals.html.
46. Wikipedia. *Tocopherols.* [cited 01-06-2012]; Available from: http://en.wikipedia.org/wiki/Tocopherol.