Phenotypic and Genomics-assisted breeding of soybean for Central Europe: From Environmental Adaptation to Tofu Traits

Dissertation to obtain the doctoral degree of Agricultural Sciences (Dr. sc. agr)

Faculty of Agricultural Sciences

University of Hohenheim

Institute of Plant Breeding, Seed Science and Population Genetics, State Plant Breeding Institute

submitted by

Alena Katharina Kurasch

from *Stuttgart*

2021

This thesis was accepted as a doctoral dissertation in fulfillment of the requirements for the degree "Doktor der Agrarwissenschaften" (Dr. sc. Agr./ Ph. D. in Agricultural Sciences) by the Faculty of Agricultural Sciences at the University of Hohenheim, on December 08th, 2021.

Day of oral examination: April 22nd, 2022

Examination Committee:

Vice-Dean/Head of Committee:	Prof. Dr. U. Ludewig
1st examiner and reviewer:	Prof. Dr. T. Würschum
2nd examiner and reviewer:	Prof. Dr. F. Ordon
3rd examiner:	Prof. Dr. R. Vögele

Table of Contents

[1]Kurasch, A. K., Hahn, V., Leiser, W. L., Vollmann, J., Schori, A., Bétrix, C.-A., Mayr, B., Winkler, J., Mechtler, K., Aper, J., Sudaric, A., Pejic, I., Sarcevic, H., Jeanson, P., Balko, C., Signor, M., Miceli, F., Strijk, P., Rietman, H., Muresanu, E., Djordjevic, V., Pospišil, A., Barion, G., Weigold, P., Streng, S., Krön, M., and Würschum, T. (2017) Identification of mega-environments in Europe and effect of allelic variations at maturity E loci on adaptation of European soybean. *Plant, Cell & Environment*, *40*, 765–778. doi: 10.1111/pce.12896

[2] Kurasch, A. K., Hahn, V., Leiser, W. L., Starck, N., & Würschum, T. (2017). Phenotypic Analysis of Major Agronomic Traits in 1008 RILs from a Diallel of Early European Soybean Varieties. *Crop Science*, *57*, 726–738. doi: 10.2135/cropsci2016.05.0318

[3] Kurasch, A. K., Hahn, V., Miersch, M., Bachteler, K., & Würschum, T. (2018). Analysis of tofu-related traits by a bench-scale tofu production method and their relationship with agronomic traits in European soybean. *Plant Breeding*, *137*(3), 271–282. doi: 10.1111/pbr.12581

[4] Kurasch, A. K., Leiser, W., Bachteler, K., Miersch, M., Hahn, V., & Würschum, T. (2018). Evaluation of the genetic architecture of tofu traits in soybean towards genomics-assisted breeding. *Plant Breeding*, *137*(6), 873–882. doi: 10.1111/pbr.12651

General Introduction

A brief History of Soybean

Today, soybean (*Glycine max* Merr.) is the world's most widely grown leguminous crop providing an important source of protein and oil for food and feed. While in the western part of the world, soybeans are often seen as a new crop, in fact, it has a long history. Historical records give evidence that soybean cultivation was known around 2510 before present and recent archeological discoveries on charred soybean seeds date back to around 5000-3000 years before present for samples from Japan, Korea and China implying that domestication occurred even earlier (Hymowitz & Shurtleff 2005, Sedivy et al. 2017, Lee et al. 2011). Thus, soybeans have a strong historical root in the Asian world and the versatility of soybean in preparing various soyfoods was perhaps the driving force that favoured its cultivation as an agricultural crop. Until today, food based on soybeans remains an important staple of the Asian diet. Soybean is now grown almost all over the world, even though it was distributed comparatively late over the continents. Around 2000 years ago soybean cultivation spread from China to Korea and Japan, and only in the beginning of the 18th century soybean seeds were introduced in Europe from China probably on several occasions. Well documented, on the contrary, was the introduction of soybean seeds in North America in 1765, from China via London by the seaman Samuel Bowen (Hymowitz & Harlan, 1983), however, only since the beginning of the 19th century soybean was grown as an important crop in America.

The predominant soybean producing countries, currently, are within the Americas with the U.S. leading, and then followed by Brazil, and Argentina (FAOSTAT, 2016). Together, the American countries account for 293.4 million tons of soybeans, corresponding to 87.6 % of the world production, whereas Asia's production share is only 8.6 %. The total world production of 334.9 million tons in 2016 was mainly increased by increasing the production area which is now at 121.5 million hectares. The main and most important importer by far of whole soybeans is China with 84 million tons, which is 25 % of the world production, followed by the Netherlands, Mexico, Spain and Germany. More important for European countries is the import of soybean cake as livestock feed.

In terms of numbers, the history of soybeans within Americas can be taken as a success story that led to the top producers for the world trade with a high economic value. Nevertheless, the success story of soybeans in America also has its downsides. During the last 30 years, the worldwide demand for soybeans mainly as livestock feed increased tremendously, thus the production area had also to increase rapidly, leading to a large-scale deforestation of rainforests in

South America. Together with the large-scale use of monoculture, leading to the exhaustion of the soil and the introduction of genetically modified soybeans to withstand herbicides which enabled on one hand a reduced tillage preserving the erosion of the soil, but on the other hand led to an increased input of herbicides on the plant and soils, soybean has won a bad reputation.

However, soybeans can offer a broad spectrum of benefits to the whole chain of production from farmer to final consumer. As a leguminous crop, soybean has the ability to fix atmospheric nitrogen through its symbiosis with the root nodule bacteria *Bradyrhizobium japonicum* contributing to a sustainable cultivation by reducing N fertilization while preserving the soil and even saving money (Reckling et al., 2016; Zander et al., 2016). The valuable seed components with the unique composition of high oil contents combined with very high protein contents offer a wide range of usage possibilities for food, feed, pharmaceutical and technical industry. In particular the high content of essential amino acids makes the soybean very valuable for the human alimentation and livestock feeding.

Regarding Europe, where the demand for locally produced agricultural products is still growing, soybean cultivation can even play an important role in a more sustainable agricultural system by increasing a local and regional production. In Europe, soybeans are cultivated in a more sustainable way preventing deforestation under strict environmental and social rules given by the European Union. Soybean enriches the crop diversity and widens the crop rotation, therefore rejuvenates the soil and contributes to a more diversified agricultural landscape. The agronomical advantages of soybean are also of great interest for the organic farming sector, where almost similar yields can be achieved compared to conventional cropped soybean (Pimentel, Hepperly, Hanson, Douds, & Seidel, 2005). While 78 % of the worldwide grown soybeans in 2016 were genetically modified mainly for herbicide tolerance (ISAAA, 2016), no approvals for genetically modified soybeans were granted for cultivation in the EU (European Commission, 2018). The increased consumer awareness in Europe even resulted in a regulation of the European Parliament (Regulation (EC) No 1830/2003) on the traceability and labeling of genetically modified organisms (GMOs) and the traceability of food and feed products produced from GMOs (Varzakas et al. 2007). By cultivating soybeans in Europe the risk of impurity of non-GM soybeans is minimized and the input of herbicides is reduced, because the conventional soybeans in Europe are not treated with the controversially discussed herbicide glyphosate. Additionally, by reducing soybean imports and increasing the production in Europe, costs and resources for transportation can be reduced as well as import dependencies. Altogether soybean grown in Europe can contribute to an agricultural change to a more sustainable agriculture which is for the benefit of the environment and the whole of society (Bues et al., 2013). Farmers in Europe are

recently expressing an increasing interest in soybean production and the great potential was also recognized by politicians recommending to substantially increase the EU's protein crop production mainly driven by the desire to reduce reliance on imports (Häusling 2014; de Visser et al. 2014). Even though the production area has doubled the last ten years to 5 million hectares grown in Europe in 2016 (FAOSTAT, 2016), and the production share increased from 1.6 % in 2006 to 3.1 % in 2016, soybean is still a minor crop in Central Europe. The main producer countries in Europe 2016 are situated in eastern and southern Europe with Italy, Serbia, France, Romania and Croatia (EUROSTAT, 2020).

Looking back at the history in Europe the potential was indeed recognized in earlier times by enthusiastically researchers and policy makers that started to promote soybeans. Most famous are the Austrian scientist Friedrich Haberlandt, often called the European soybean pioneer who started testing soybeans at the end of the 19th century and Sven Holmberg, a plant breeder from the south of Sweden who started a breeding program to develop very early soybean varieties in 1940. Holmberg successfully developed very early maturing varieties as for example 'Fiskeby V' that was even able to reach maturity at 58° latitude (Shurtleff & Aoyagi, 2015). Despite the successful efforts made by these researchers the cultivation and breeding fell almost dormant in the second half of the 20th Century in Central Europe.

So, what hampered the successful establishment of soybean in Central Europe?

Adaptation and Agronomic Traits

The key role for a successful establishment of soybean cultivation in Europe is adaptation of soybean varieties to the central European growing conditions. Soybean has in general a great range where it can be grown, from 55° northern hemisphere to 55° southern hemisphere, with temperate and tropical regions (Lieberei & Reisdorff, 2012). However, it is possible that under cooler temperate climates at higher latitudes soybean will not mature completely, which is a crucial prerequisite for profitable cultivation. Soybean adaptation, however, includes more than reaching maturity or early flowering, but combines the ability to harvest soybeans at the most favorable moment and obtain the highest possible grain yield at a certain environment. A well-adapted cultivar can therefore be defined as best exploiting the yield potential of a certain environment (Zhang et al., 2007). Thus, adaptation is the simultaneous improvement of precocity and yield which is a challenge since maturity and yield are positively correlated in soybean (Cober & Morrison, 2010). An outstanding example, anyway, for a successful adaptation to European growing conditions is maize where the yield could be highly increased during the last 50 years and the production areas could be expanded towards more unfavorable cooler climates in the

northern parts of Europe through selection (Tenaillon & Charcosset, 2011). Hence, a successful adaptation and expanding of production area might be possible for soybean as well by appropriate selection of superior genotypes, especially since soybean and maize demand similar requirements on environmental conditions.

Consequently, breeding of soybean cultivars with improved agronomic traits and, especially, adaptation to higher latitudes is necessary to establish soybean cultivation in Central Europe in the long term. For that, knowing and characterizing the target environments and the genetic material which is also covered by the term of adaptation is one prerequisite for selection of adapted material. A concept following this is the maturity group (MG) classification system established in the U.S., where regions and genetic material are classified in a corresponding way. The original classification from the U.S. suggested seven MGs designated as MG I-VII where each MG corresponds to an adaptation zone covering a narrow belt with around 200 km in a north south-direction. Soybean as a short-day plant exhibits sensitivity to photoperiod, which explains the limit of each MG in the geographical distribution to a narrow range of latitude (Scott & Aldrich, 1983; Zhang et al., 2007). Today, due to the north and south expansion of soybean growing areas, this system was extended by earlier MGs from ultra-early '0000' to early '0' and later MGs from 'VIII' to 'X' (Jia et al., 2014; Zhang et al., 2007). The system facilitates a judgement of the chances of success for new varieties for breeding and production purpose. Consequently the MG system was adopted by many scientists in different countries from Japan, India, Brazil, Argentina, Canada and Europe summarized by Liu et al. (2017). The MG system is partly adaptable for Chinese material because the material is more diverse in flowering time and reproductive period in relation to maturity date, as the material was adapted to multiple cropping systems where different ecotypes are used with either a shorter vegetative period but longer reproductive period suited for northern spring planted areas or the other way round with a longer vegetative period but shorter reproductive period suited for double cropping areas (Yuesheng et al. 2006).

In Europe, the most important MGs cover a range from '000' to 'II', while the ultra-early '0000' is of minor importance so far due to the low yield potential, however, it might be important in expanding the growing area. Besides the photoperiod sensitivity, the temperature demand of soybean is also limiting its distribution. Therefore, Canada uses additionally Crop Heat Units to classify an adaptation zone and corresponding MG.

For the latitudinal adaptation to long day conditions from Central to Northern Europe, an adapted early flowering and maturity time is of crucial importance for a profitable cultivation.

The key traits flowering and maturity are quantitatively inherited and photoperiod responsiveness remains a key factor besides temperature sensitivity, regulating the key traits. To date, at least 11 loci are known that show an association with flowering and maturity date: *E1* and *E2* (Bernard, 1971), *E3* (Buzzell, 1971), *E4* (Buzzell & Voldeng, 1980), *E5* (McBlain & Bernard, 1987), *E6* (Bonato & Vello, 1999), *E7* (Cober & Voldeng, 2001), *E8* (Cober et al., 2010), *E9* (Kong et al., 2014), *E10* (Samanfar et al., 2017) and *J* (Ray et al., 1995). Except for *E6* and *E9* the dominant alleles of *E1* to *E10* confer to a later flowering and maturity phenotype. In recent years, genes underlying the maturity loci have been molecularly characterized, identified and their functions have been described for *E1-E4*, *E9*, *E10*, *J* and *Dt1* (Harada et al., 2011; B. Liu et al., 2010; Lu et al., 2017; Matsumura et al., 2008; Samanfar et al., 2017; Tian et al., 2010; Watanabe et al., 2009, 2011; Xia et al., 2012; Zhao et al., 2016), whereas the genes underlying *E6*, *E7* and *E8* remain unknown.

The most important loci for an early flowering and maturity are *E1-E4* and the various allelic combinations condition soybean flowering and maturity time and therefore strongly contribute to the wide adaptability (Jiang et al., 2014; Tsubokura et al., 2014; M. Xu et al., 2013). So far, the classification of lines into a MG can only be done by either a standard variety representing a MG or by testing a new line at many locations corresponding to a MG zone. Trying to connect *E* genes and MGs is just a logical consequence which was done in our study across Central Europe and meanwhile several studies were conducted trying to connect the allelic combination of *E1-E4* to MG for different regions and germplasms (Langewisch et al., 2017; X. Liu et al., 2017; Miladinović et al., 2018).

Since adaptation is not only reaching maturity but includes also the best exploitation of the yield potential of a certain environment, it also covers a broad range of agronomic traits. Because the climatic environment is always changing, adaptation is never attained but is more a continuous process including in addition to yield and maturity also agronomic traits. Looking again at the United States, soybean yields have largely increased with on average 23.4 kg ha^{-1} per year from 1924 to 2011, but with higher rates from 1972 to 1997 (Rowntree et al., 2013; Specht et al., 1999; Wilson et al., 2014). The yield gain is to one half attributable to the genetic advances and to the other half to the advances in agronomic practices. The agronomic improvements are mainly earlier planting dates, narrower row spacing, higher seeding rates, improved weed control and herbicide use as well as reduced harvest losses due to improved harvest machines (Rowntree et al., 2013). In more detail, the genetic gain improvement is most likely due to physiological improvements (Koester et al., 2014) that allowed increasing yield by producing more seeds per plant rather than increasing the seed size as well as improved pest resistance, decreased pod

shattering and decreased lodging (Specht et al., 1999) traits which can be also attributed to adaptation.

For Europe, a yield increase can therefore be expected for the future, as well, mainly due to better adapted material provided by plant breeding and of course by agronomic and technical improvements in cultivation. Soybean breeding has to face adaptation in two steps. First, identifying material that enables an adequate flowering and maturity, followed by selection of lines with highest yield potential. High yield without appropriate maturity time is useless for a farmer as well as a variety that ripens very early but has a very low yield potential as for example 'Fiskeby V'. However, lines like that can be very valuable as genetic resources to incorporate precocity in a breeding pool to create variation.

A further step to increase the farmer's adoption of soybean cultivation in Central Europe is to provide not only adapted cultivars but cultivars with superior value, as for example quality traits important for feeding purpose since this is the major usage in Europe. Protein and oil content are the most important quality traits targeted for selection. Most of the worldwide-cultivated soybean cultivars are primarily bred for high oil content (Sato et al., 2014) while in Europe soybean is not as interesting to produce oil since sunflower and oil seed rape are the major well-established oil seed crops. The main interest for European farmers is thus a high protein content as source for food and feed. Establishing our own Central European soybean breeding programs is therefore of great importance to meet farmer and market needs.

Tofu-omics

Soy-based food plays a pivotal role in the Asian cuisine offering very diverse products ranging from fresh green soybean known as edamame, fermented soybean products as soy-sauce, tempeh and natto up to soymilk and tofu, which are the most prominent products in the western part of the world. Tofu is enjoying ever greater popularity, as it is one of the best sources of plant protein with additional health benefits, rich in essential amino acids, beneficial lipids, vitamins, and minerals, as well as other bioactive compounds, such as isoflavones, soyasaponin, and others, (Lima et al., 2017; Zhang et al., 2018). Due to the rising trend of vegetarian and vegan diets in Europe often combined with a lifestyle of health and sustainability (LOHAs) with increased demand for locally produced products, tofu suitability of soybean seeds comes more and more into focus as a breeding target. Tofu manufacturers show an increased interest in locally grown soybean seeds as raw material to fit the market demand for local, sustainable and traceable produced tofu. Thus, plant breeding has to provide not only well adapted varieties with good agronomic and quality properties but also provide varieties well-suited to the further processing

into soymilk and tofu. Therefore, a good knowledge about the breeding target, how to assess it and how it is inherited is crucial.

Tofu can be classified in two major types, silken/soft tofu and firm tofu. However, firm tofu can vary in its consistency from very soft to very hard depending on the consumer preferences. Tofu produced for the European market is commonly of a very firm texture whereas the traditional firm tofu in Asia is often softer.

The main steps in firm tofu processing include the soaking of the raw soybeans, grinding, heating of the slurry, filtering, coagulation and finally pressing of the curd. Unfortunately, each single step from soaking time, soymilk or slurry heating time, heat exposure, stirring speed, time and temperature for coagulation, choice of coagulant, pressing time and weight, affects the quality of the final tofu. These physical parameters including also the lab condition can be best described and summarized as extrinsic factors influencing the final tofu product in its quality and yield. Factors influencing the tofu texture and yield that are related to the seed characteristics, such as protein concentration or composition can be classified as intrinsic factors. The seed characteristics, in turn, are influenced by the genetics, by the location and environment during cultivation, (Jaureguy et al., 2011; Kumar et al., 2006; Poysa et al., 2006) and by the post-harvest storage of the seeds (Hou & Chang, 2004). Among the intrinsic factors the protein composition is regarded as the most important and direct factor influencing the final tofu quality, however, the formation of tofu is also influenced by lipids, carbohydrate and phytic acids or its salts. Taken together, tofu texture and yield are affected by many very different factors and research on tofu processing has progressed in recent years as reviewed by Zhang et al. (2018). Less studies are available that deal with the genetic contribution to the variation for breeding purpose or analyzing the genetic architecture of traits.

Hence, plant breeding faces two challenges, first to evaluate and establish an efficient method to assess tofu quality which mainly includes the tofu texture and yield and second to get the information of the genetic part affecting the tofu properties of a soybean line. The requirements for a suitable test to quantify tofu texture and yield for a breeding program are quite different from the fine-scale optimization for commercial production of tofu. The method must be high-throughput allowing evaluating many breeding lines with a small number of seeds since only a limited number of seeds are available in early stages of a breeding program and tofu processing is a destructive process. In addition, the method has to be quick and has to be relevant to the end-use product tofu, but most important and that's also the biggest difference to a commercial tofu production, the method must allow for discriminating among breeding lines for tofu production from seed grown over multiple environments (Reid & Cober, 2018).

Thus, a quick, high-throughput small-scale method giving reliable results is needed to evaluate breeding lines. While the method is mainly dealing with extrinsic factors which have to be minimized and standardized, so that they do not contribute to the variation of the tofu-related traits, we further want to partition the variation corresponding to the intrinsic factors into effects derived from the genetics, the environment and the interaction of them.

In our study a small-scale method was established and used by Taifun Tofu in a previous project (Wilbois et al., 2014) that is oriented at the real production at Taifun Tofu using the traditional Nigari ($MgCl_2$ magnesium chloride) plus $CaSO_4$ Calcium sulfate as coagulant and a very small amount of only 80 g seeds.

The traits taken into account in our study were weight gain after soaking, tofu yield traits such as soymilk yield, tofu weight and tofu yield which gives the tofu to soybean ratio corrected for the proportion of used soymilk, and as quality criterion the tofu hardness was measured. While the tofu texture is of major importance for the acceptance of the end consumer, the tofu yield and soaking traits are of major importance for the tofu manufacturer to be most efficient in production. Soaking as the initial step can take up to 24 hours depending on the production method and it could be of an economical interest for the manufacturer to shorten this time by selecting appropriate varieties that need less time for soaking up to a final moisture content of 120 % (Pan & Tangratanavalee, 2003).

Phenotypic and Genomics-assisted Selection

Phenotypic and genomics-assisted selection is the basic application of quantitative genetics to crop improvement which is driven by the variables that describe response to selection. The response to selection (R) is defined as the realized average difference between the parent generation and the next generation. R is the product of the selection intensity i which is the standardized selection differential applied to the selection units, the heritability h^2 and the square root of the phenotypic variance σ_P which is the phenotypic standard deviation of the target trait for the selection unit in the reference population, highlighting the importance of the variance components estimation. Thus, R is determined by the genotypic variance which depends on the trait and its genetic nature and the accuracy of the trait assessment, by the phenotypic standard deviation of the trait which can be modified by choosing appropriate trial designs depending on location, years and replications, and R is determined by the selection intensity, where larger populations can result in a larger response to selection, but also in a larger effort. Taken together, a desirable breeding population enabling successful selection is therefore characterized by a high mean combined with a large genetic variance for the target traits. Rearranging the response to selection leads to the commonly known breeder's equation that describes the genetic gain. The

formula is very central in plant breeding to quantify the success of selection (Cobb et al., 2019). The breeder's equation can be arranged for phenotypic selection $\Delta G_p = \frac{i*h^2*\sigma_P}{t}$ and likewise for genomic selection as $\Delta G_A = \frac{i*r_A*\sigma_A}{t}$ where t is the time to complete one cycle of the breeding program (Voss-Fels, Cooper, & Hayes, 2019). The genomic breeder's equation ΔG_A corresponds to the genetic gain of genomic selection in terms of estimated genotypic value. The only differences to the classical breeder's equation are the replacement of heritability by the prediction accuracy r_A, and the replacement of the phenotypic standard deviation of the target trait by σ_A which is the additive genetic standard deviation of the target trait for the selection unit in the training population. The breeder's equation is a good starting point serving as mental framework in the evaluation and establishment of a breeding program because the parameters of the equation articulate what a breeder aims to manipulate as part of the crop improvement process.

The quality of the phenotypic data is influenced by the quality and reliability of the data collection, and the field design including multi locations and years. Testing genotypes by increasing the number of test locations while decreasing the number of replicates per environment using unreplicated or partially replicated (p-rep) designs was shown to be an appropriate strategy for enhancing the data quality (Moehring, Williams, & Piepho, n.d.) due to capturing a larger proportion of the genotype-by-environment interaction. All this information and parameters can be used for an optimum allocation of resources and to design a breeding program, to decide where and when to test and select for which trait. Altogether, a proper phenotypic data analysis and most precise phenotypic data are a crucial prerequisite to decide the selection strategy for different target traits including genomics-assisted approaches and thus to design an optimal soybean breeding program (Bernal-Vasquez et al., 2014).

With the development of molecular markers and therefore the ability to profile the DNA of a genotype it has become obvious to use them for the identification and selection of superior plants on the genetic level instead of selecting on the phenotypic level. Many genomic tools are nowadays available where two main genomics-assisted selection strategies can be distinguished: First, marker-assisted selection (MAS) which requires the detection of marker-trait associations by a QTL mapping approach where molecular markers are mapped near or within specific loci with phenotypic effects followed by a selection of individuals that carry favorable alleles for the trait of interest; and secondly, genomic selection (GS) where the information of all available markers is exploited simultaneously to predict a breeding value which was originally suggested by Meuwissen et al. (2001). Today, these applications are easily available mainly due to the tremendous advances of marker systems in the last decades with two main advancements, the

development of high-throughput technologies for marker genotyping and the development of statistical methods and computer software for implementing genomic procedures (Barabaschi et al., 2015; Bernardo, 2008; Mammadov et al., 2012; Varshney et al., 2014). Since the development of the first molecular marker systems RFLPs in the 1980's the development has changed from fragment based marker systems towards single nucleotide polymorphisms (SNPs) which has become the standard molecular marker system for genomics-assisted selection strategies because of their high frequency of occurrence in the genomes (Xu & Crouch, 2008). Genotyping-by-sequencing (GBS) was shown to be an easy and cost effective method for many species to provide up to hundreds of thousands of genome-wide SNP markers, due to the simultaneous SNP discovery and genotyping instead of the classical two-step process where SNP discovery is followed by an assay design (Elshire et al., 2011; Heslot et l., 2013; Poland & Rife, 2012). Especially, in combination with a reference genome, which is available for soybean (Schmutz et al., 2010), GBS is a powerful tool, since ordering and imputing low coverage marker data is much more straightforward (Poland & Rife, 2012). One of the most important advantages of GBS is furthermore the bias reduction due to the de novo SNP discovery and genotyping compared to array-based genotyping assays with pre-defined SNP panels which is highly biased towards a particular set of germplasm (Elbasyoni et al., 2018; Poland & Rife, 2012).

QTL mapping can be used in an explorative way with the aim to study the genetic architecture of a quantitative trait or in an applied way with the aim to identify markers which can be used for marker-assisted selection (Bernardo, 2008). Studying the genetic architecture of a trait is fundamental to decide which selection strategy to follow. Marker-assisted selection for example can only outperform phenotypic selection for traits that are expensive, laborious or time-consuming to evaluate and for identified QTL that contribute a substantial proportion of the genotypic variance. The difficulty here is that most of the target traits are complex inherited and show a low heritability which was shown to be very important for an efficient marker-assisted selection. However, a prominent example of a successful application of marker-assisted selection in soybean is the resistance to soybean cyst nematode SCN (Concibido et al., 1996) where molecular markers linked to SCN resistance QTL have been routinely used to introduce SCN resistance into elite soybean lines (Cahill & Schmidt, 2004).

As the phenotyping of tofu-related traits is complex, genomics-assisted approaches would be of great benefit compared to conventional phenotypic selection. Thus, QTL mapping to study the genetic architecture is the first step towards genomics-assisted selection and to estimate the success for the different genomics-assisted selection strategies. For traits that are highly complex and thus controlled by many small-effect QTL each contributing only a very low proportion of

the genotypic variance, genomic selection has proven to be more efficient compared to MAS as it captures the information of all available markers simultaneously throughout the genome by calculating a genomic estimated breeding value (GEBV) (Heffner, Sorrells, & Jannink, 2009; Meuwissen et al., 2001). The main principle of GS is to establish in a first step a prediction model by associating marker information with phenotypic information using a training set of individuals that have been genotyped and phenotyped. The second step is then to apply the prediction model to individuals that have been genotyped but not phenotyped and to calculate GEBVs for these untested selection candidates. GS has been broadly applied to various traits and crop species including soybean (Barabaschi et al., 2015), illustrating the potential of GS for soybean breeding (Bao et al., 2014; Duhnen et al., 2017; Jarquín et al., 2014; Shu et al., 2013; Zhang et al., 2016).

Aims of the study

The study covers various aspects relevant for the design of a European soybean breeding program. In particular, the objectives were to

- evaluate growing areas and the prerequisites for adaptation and study the maturity E genes

- evaluate variance components for different agronomic and quality traits, as well as specific food-related traits

- evaluate the application of genomics-assisted methods to study the genetic architecture and the use of genomic selection strategy for tofu-related traits

The general aim of this work was to increase our knowledge on parameters important for the establishment and the long-term success of Central European soybean breeding programs and how to implement them.

Identification of mega-environments in Europe and effect of allelic variation at maturity *E* loci on adaptation of European soybean[1]

Alena K. Kurasch , Volker Hahn, Willmar L. Leiser, Johann Vollmann, Arnold Schori, Claude-Alain Bétrix, Bernhard Mayr, JohannaWinkler, Klemens Mechtler, Jonas Aper, Aleksandra Sudaric, Ivan Pejic, Hrvoje Sarcevic, Patrice Jeanson, Christiane Balko, Marco Signor, Fabiano Miceli, Peter Strijk, Hendrik Rietman, Eugen Muresanu, Vuk Djordjevic, Ana Pospišil, Giuseppe Barion, Peter Weigold, Stefan Streng, Matthias Krön & Tobias Würschum

The original publication is available at:

https://onlinelibrary.wiley.com/doi/epdf/10.1111/pce.12896

Abstract:

Soybean cultivation holds great potential for a sustainable agriculture in Europe, but adaptation remains a central issue. In this large mega-environment (MEV) study, 75 European cultivars from five early maturity groups (MGs 000–II) were evaluated for maturity-related traits at 22 locations in 10 countries across Europe. Clustering of the locations based on phenotypic similarity revealed six MEVs in latitudinal direction and suggested several more. Analysis of maturity identified several groups of cultivars with phenotypic similarity that are optimally adapted to the different growing regions in Europe. We identified several haplotypes for the allelic variants at the *E1*, *E2*, *E3* and *E4* genes, with each *E* haplotype comprising cultivars from different MGs. Cultivars with the same *E* haplotype can exhibit different flowering and maturity characteristics, suggesting that the genetic control of these traits is more complex and that adaptation involves additional genetic pathways, for example temperature requirement. Taken together, our study allowed the first unified assessment of soybean-growing regions in Europe and illustrates the strong effect of photoperiod on soybean adaptation and MEV classification, as well as the effects of the *E* maturity loci for soybean adaptation in Europe.

Phenotypic Analysis of Major Agronomic Traits in 1008 RILs from a Diallel of Early European Soybean Varieties[2]

Alena K. Kurasch, Volker Hahn, Willmar L. Leiser, Norbert Starck, and Tobias Würschum

Abstract:

The leguminous crop Glycine max (L.) Merr., commonly known as soybean, holds the title of the most significant crop worldwide. Despite this, Europe relies heavily on soybean imports, making it crucial to improve cultivars via breeding for further expansion of soybean cultivation within the continent. Our study evaluated 1008 F5:8 recombinant inbred lines from five early-maturing European soybean varieties in multilocation field trials for grain yield, thousand-kernel weight, plant height, protein content, and oil content. Our findings unveiled significant genotypic variances, high heritabilities (h2 > 0.7), and transgressive segregation for all traits. Results also highlighted the complicated interplay between grain yield, plant height, and maturity but showed promise for breeding adapted and high-yielding varieties. Interestingly, maturity had no effect on protein and oil content, but both showed high negative correlation (r ~ −0.9). Therefore, maximizing grain yield may be an optimal strategy for soybeans intended for feeding purposes, while lines with high protein content could be desirable for food-grade varieties, particularly for tofu production. Overall, our results on phenotypic variation, variance components, heritabilities, and trait correlations could aid in soybean breeding directed at Central Europe.

[2] Kurasch, A. K., Hahn, V., Leiser, W. L., Starck, N., & Würschum, T. (2017). Phenotypic Analysis of Major Agronomic Traits in 1008 RILs from a Diallel of Early European Soybean Varieties. *Crop Science*, *57*, 726–738. doi: 10.2135/cropsci2016.05.0

Analysis of tofu-related traits by a bench-scale tofu production method and their relationship with agronomic traits in European soybean[3]

Alena K. Kurasch, Volker Hahn, Martin Miersch, Kristina Bachteler, Tobias Würschum

The original publication is available at:

Abstract:

The rising interest of European tofu manufacturers in local soybean cultivation requires further understanding of the impacts of genetics and environment on the quality of soy milk and tofu, as well as their correlation with agronomic traits. In this study, 215 recombinant inbred lines derived from two populations were evaluated for tofu traits in a bench-scale laboratory across three locations. The heritabilities of most evaluated tofu traits were moderately high, with significant genotypic and location variances. Population-dependent network analysis suggested the potential for improving tofu-related traits in European soybean, but agronomic traits were not associated. The bench-scale tofu production method is useful in breeding programs, but automation is necessary to reduce errors from laboratory staff.

---

[3] Kurasch, A. K., Hahn, V., Miersch, M., Bachteler, K., & Würschum, T. (2018). Analysis of tofu-related traits by a bench-scale tofu production method and their relationship with agronomic traits in European soybean. Plant Breeding, 137(3), 271–282. doi: 10.1111/pbr.12581

Evaluation of the genetic architecture of tofu traits in soybean towards genomics-assisted breeding[4]

Alena K. Kurasch, Willmar L. Leiser ,Kristina Bachteler, Martin Miersch , Volker Hahn, Tobias Würschum

The original publication is available at: https://onlinelibrary.wiley.com/doi/epdf/10.1111/pbr.12651

Abstract:

Developing soybean cultivars with enhanced tofu suitability is vital to tofu manufacturers. However, assessing traits linked to tofu production is arduous, expensive, and time-consuming. The aim of this study was to explore the prospective marker-based strategies for selecting improved tofu quality. We evaluated three RIL populations for various tofu-related traits and detected one QTL at most per trait in each population, accounting for 10.9% to 74.2% of genetic variation. We failed to discover any QTL for tofu yield and tofu weight. The QTL for tofu hardness, tofu value, and soymilk weight were specific to each population and could potentially regulate the water-soluble protein proportion. Our QTL mapping outcomes suggest that tofu traits are predominantly quantitative. Genomewide prediction results exhibited promise for seed quality and soaking traits, but for tofu traits, accuracies must improve as phenotyping practices get refined.

[4] Kurasch, A. K., Leiser, W., Bachteler, K., Miersch, M., Hahn, V., & Würschum, T. (2018). Evaluation of the genetic architecture of tofu traits in soybean towards genomics-assisted breeding. Plant Breeding, 137(6), 873– 882. doi: 10.1111/pbr.12651

General Discussion

Increasing the genetic gain by phenotypic and genomic selection

The main objective for the establishment of a soybean breeding program for Central Europe is to select superior plants and provide them to the farmer. To define what a superior plant is for European farmers, breeding goals have to be prioritized. The most important objective is therefore the environmental adaptation in terms of maturity, especially if the goal is to broaden the production area to higher latitudes as for example Germany, Poland, Belgium, the Netherlands and even Denmark. So, the environmental adaptation is the highest priority since this is right now the limiting factor for a successful production in the northern locations mentioned above. Besides maturity, of course grain yield is highly relevant for farmers' adoption particularly in regions where soybean has not been grown before. A stable and reliable high grain yield can only be reached when the agronomic traits as for example plant height and lodging resistance are taken into account during the selection process. Additional benefits for the farmers are quality traits that generate a greater added value as for example high oil content or high protein content. Producing soybean for the human consumption can be very interesting for farmers as it is in general very well paid as it requires special demands. Thus, varieties well suited for soymilk and tofu production are of major interest for growers and manufactures. Taken together, the first priority is the environmental adaptation, followed by grain yield and agronomic traits, and then quality traits (e.g oil and protein content) and finally specific traits like tofu suitability as it is a small but promising niche. Increasing the selection response and thereby the genetic gain by phenotypic and genomic selection is the key in designing a successful breeding program with respect to the prioritized objectives.

Characterization of Breeding Germplasm

The first step in the establishment of a breeding program is to know the material and the target environment to increase the adaptation of soybean to European growing conditions. Based on this information a breeding program can be implemented and the breeding goals can be formulated.

The germplasm used in our study represents an unselected breeding population developed by the single-seed-descent method derived by half-diallel crossing of five common European soybean varieties (Figure 1a). The half-diallel crosses allowed the creation of several biparental populations that are connected to each other by common parents with in total 1008 $F_{5:8}$ recombinant inbred lines (RILs). The eight biparental populations can easily be distinguished in a

PCoA plot, lying in between the corresponding parental lines (Figure 1), showing different diversity with populations P2×P4 and P1×P3 as rather narrow whereas the remaining populations show a broader spectrum in between the parental lines.

Besides parents P1 and P3, the five parental lines are farthermost, thus, representing diverse material with P2 and P5 the most diverse varieties among the parental lines consistent with the PCoA of the previous published study (Kurasch, Hahn, Leiser, Starck, & Würschum, 2017). The present breeding germplasm was used as the base population for the different studies conducted and published (Kurasch, Hahn, Leiser, Starck, et al., 2017; Kurasch, Hahn, Miersch, Bachteler, & Würschum, 2018; Kurasch, Leiser, et al., 2018).
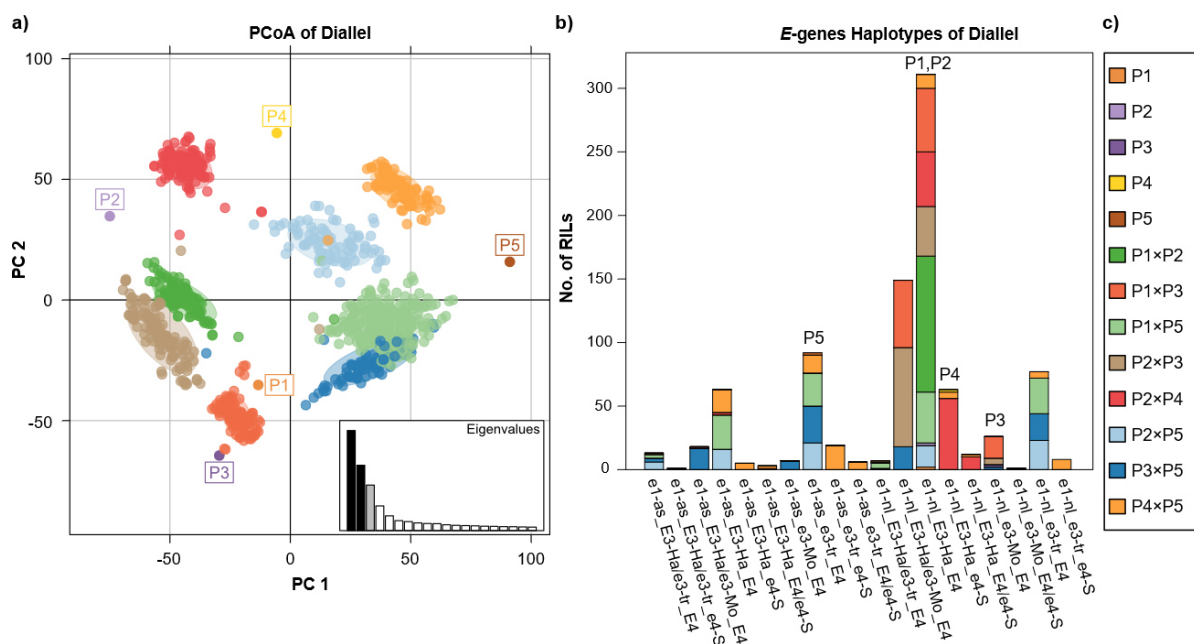

Figure 1: a) Principal coordinate analysis of the eight populations; b) Frequencies of E-genes haplotypes depending on population indicated by colored stacked bars; c) Color legend with P1 Gallec, P2 Primus, P3 Protina, P4 Sultana, P5 Sigalia.

Studying the allelic combination of the E genes will help to better understand the adaptation in terms of maturity to specific environments and might help to achieve appropriate maturity.

Even though the chosen parental varieties were from two early MGs (00, 000) the derived populations segregated strongly for maturity as was shown in Kurasch, Hahn, Leiser, Starck, et al. (2017). A closer look at recent analysis of the E genes ($E1$, $E2$, $E3$ and $E4$) of the whole population revealed 19 different E haplotypes explaining the strong segregation for maturity (Figure 1b). The $E2$ locus was monomorphic for the non-functional recessive allele $e2$-ns. Among the 19 haplotypes 9 haplotypes contain one heterozygous locus for either $E4/e4$-S, $E3$-$Ha/e3$-tr or $E3$-$Ha/e3$-Mo leaving 10 fixed haplotypes. Compared to the 75 varieties evaluated in 2017 ranging from MG 000 to II where in total 10 haplotypes were identified with only 6 haplotypes

from the two earliest MGs 000 and 00 (Kurasch, Hahn, Leiser, Vollmann, et al., 2017), we found more diverse haplotypes within the RILs. The most abundant haplotype *e1-nl/e2/E3-Ha/E4* can be found in one third of the analyzed RILs, since two parents P1 and P2 carrying this haplotype. Furthermore, we could even identify RILs with the rare photoperiod-insensitive allele combination *e1-nl/e2/e3-tr/e4* which derived from the cross of the relatively late variety P5 with early variety P4.

Recently, several studies were published examining the four maturity genes and their allelic diversity in germplasm from diverse geographic origins, from North America, China and Europe in relation to the MG classification and the relevance for adaptation to a specific latitude (Langewisch et al., 2017, 2014; Li et al., 2017; X. Liu et al., 2017; Miladinović et al., 2018). Miladinović et al. (2018) reported two main haplotypes for European varieties with either *e1as/E2/E3/E4* or *e1-as/E2/e3-tr/E4*, however these haplotypes refer probably to MG0 or later. They identified only one haplotype with the non-functional *e1-nl* allele within the European germplasm under study. These findings may reflect the higher frequencies of varieties belonging to MG later then MG0 since the biggest production areas are more in the southern eastern part of Europe and by now only a limited number of early and very early varieties are registered in the EU or European countries. Miladinović et al. (2018) suggested *e1-as/e2/E3/E4* as best adapted haplotype for environments of Central Eastern Europe as it was the most abundant haplotype and highest yielding within the Serbian varieties. As it was the most abundant haplotype as well in the study of 75 varieties, but covering at least four MGs from MGII to MG00, assuming two falsely classified varieties of MG000 with 'Perla' and 'Diamant' (Kurasch, Hahn, Leiser, Vollmann, et al., 2017), this might be a very versatile and interesting haplotype for further studies. For a better adaptation to higher latitudes in Central-Northern Europe *E* haplotypes containing the *e1-nl* allele will probably be more suitable as all varieties from MG000 consist of haplotypes with the non-functional *e1-nl* allele (Kurasch et al., 2017).

Taken together, the recessive *e2* allele together with the non functional *e1-ns* allele seem to be important alleles for adaptation to high latitude and cold regions like northern parts of Europe with early MGs '000 and '00' (Jia et al., 2014; Kurasch, Hahn, Leiser, Vollmann, et al., 2017; Tsubokura et al., 2014). The impact of different *E3* alleles is not as clear, as several alleles exist in the present breeding germplasm as well in the study from 2017, with no phenotypic differentiation for maturity. Interestingly, the variety 'Protina' carries the rare allele *e3-Mo* which was until now only detected in the Japanese cultivar Moshido Gong 503 by Harada et al. (2011) indicating a Japanese origin in the pedigree of the variety 'Protina'.

In the previous study a complex relationship among plant height, grain yield and maturity was reported (Kurasch, Hahn, Leiser, Starck, et al., 2017) and the new analysis of the *E* genes revealed now that the *E1* locus strongly affects plant height with the *e1-as* allele leading to significantly taller plants compared with *e1-nl* allele (Figure 2; shown by the notched boxes). This becomes even clearer when comparing all haplotypes. *E3* and *E4* do not have a clear effect on plant height. Nevertheless, among the *e1-nl* haplotypes there are some RILs showing a rather high plant height, especially for *e1-nl/e2/E3-Ha/e4* which is the most abundant haplotype. The large variation of plant height within this haplotype is very likey influenced by the genetic background of the RILs with specific plant height-related loci. The effect of the *E1* alleles on grain yield is not as pronounced as for plant height but is still significant according to the notched boxplots (Figure 3). Comparing the *E* haplotypes for grain yield we can identify RILs with rather high grain yield among haplotypes containing *e1-nl*, indicating the importance of the genetic background and the complex inheritance of grain yield which is generally subject to environmental factors. Our findings suggest a pleiotropic effect of the *E1* locus related to grain yield and plant height, confirming previous reports that flowering, maturity and plant height are controlled by shared major genes (Cober & Morrison, 2010; Fang et al., 2017; Lee et al., 1996).
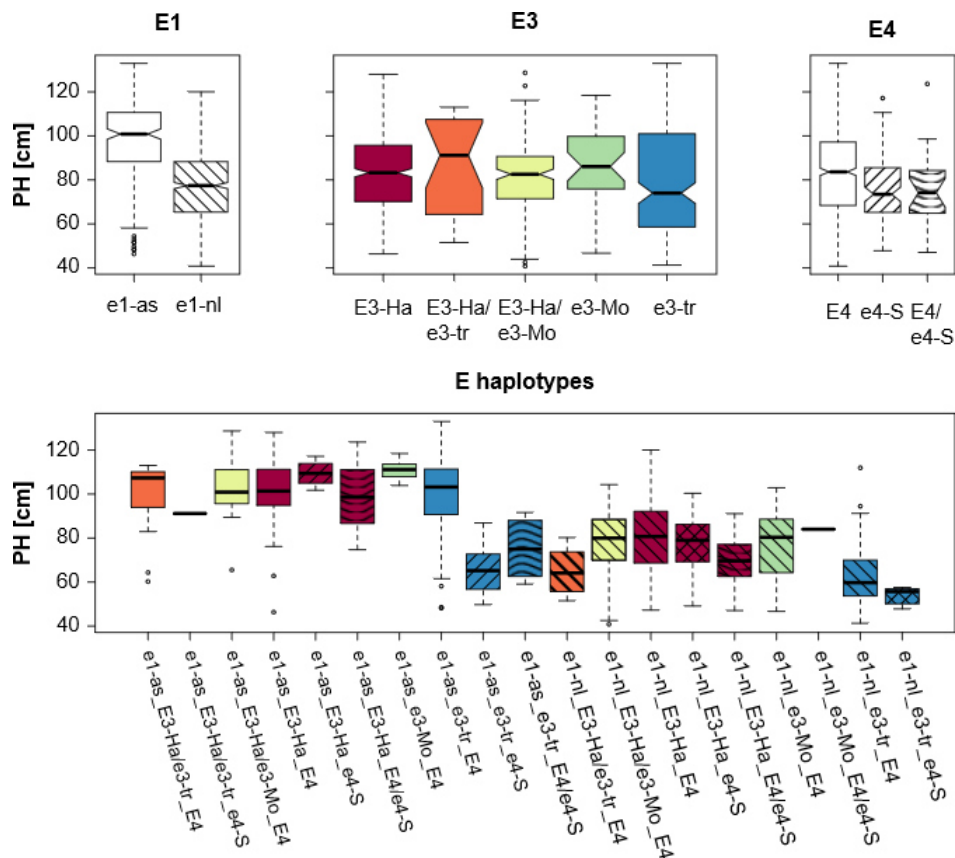

Figure 2: Distribution of plant height (PH) depending on E-haplotype.

Fang et al. (2017) also reported pleiotropic effects for *E1*, but mainly for *E2* we could not detect this due to the monomorphic *E2* locus in our material. The effect of the *E2* locus might be stronger than that of the *E1* locus, but since we only have the recessive *e2* allele in our germplasm, the effect of *E1* was obviously not masked by *E2*. To conclude, even though the *E1* locus has a strong effect on plant height, more plant height-related genes are involved independent from the maturity or *E* genes.

The limit of our study is the missing data for the maturity in days to maturity; however, we can classify the RILs into the three different trials as has been done in the previous study (Kurasch, Hahn, Leiser, Starck, et al., 2017) according to the maturity dates from 2013 into early, mid-early and late. Plant height in general is positively correlated with maturity and grain yield so the RILs with higher plant height even with *e1-nl* allele are assumed to be later, which was confirmed by figure 4 (Appendix).

In summary, there are three common conclusions or consents from the different studies including ours, first the earlier maturing soybeans require more recessive non-functional *E* genes
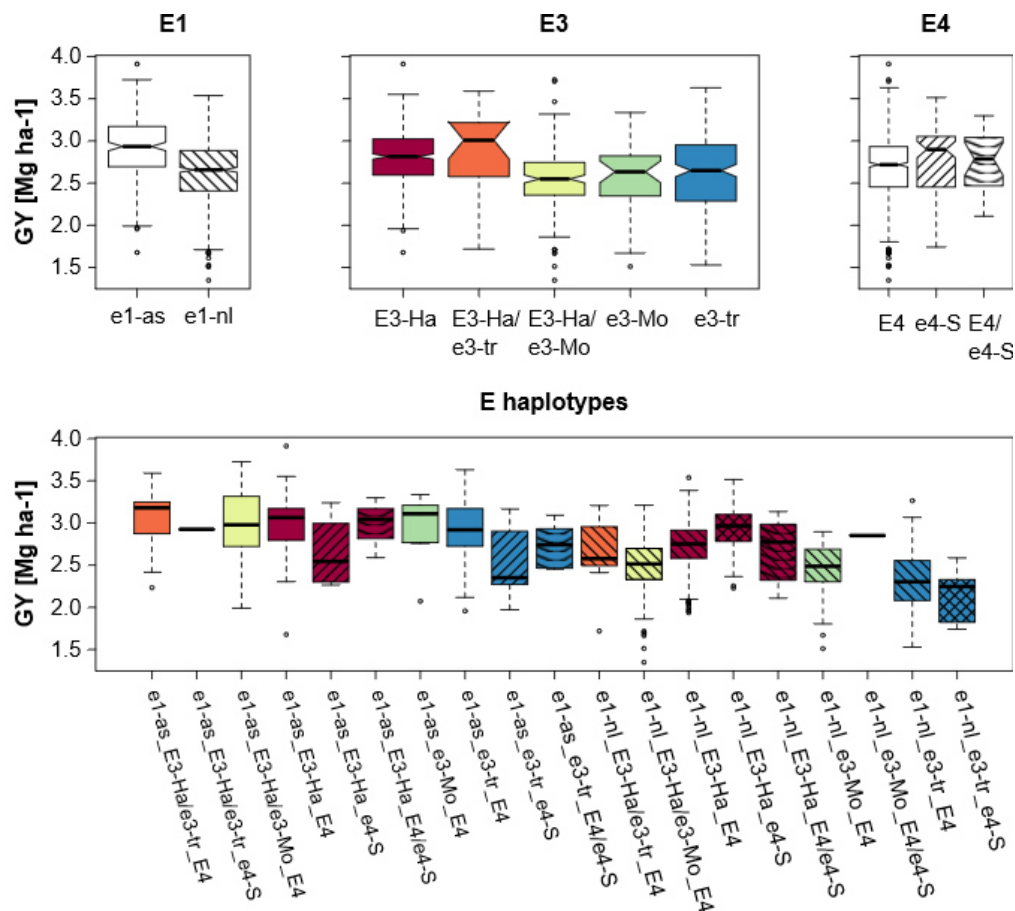

Figure 3 : Distribution of grain yield (GY) depending on E-haplotype.

alleles to fit to higher latitude environments being photoperiodic insensitive leading to earlier phenotypes, second it is not possible to predict or classify soybean with E haplotypes into specific MG and third, more molecular mechanism are involved as for example temperature sensitivity or other maturity-related genes. Results achieved so far by several research groups do not necessarily allow predictions of the MG classification based on E haplotypes even though there are strong tendencies (Langewisch et al., 2014; Li et al., 2017; Liu et al., 2017; Miladinović et al., 2018)

Characterization of environments

The classification of varieties into MGs helps the breeder and the farmer to understand the behavior or performance of varieties in a specific environment. Consequently it is useful to classify genotypes and characterize environments in a reciprocal way so that an appropriate maturity of a variety can be determined for a target environment in order to maximize the yield potential of the variety and the target environment.

The five mega environments identified in 2017 (Kurasch et al., 2017) give a first overview of growing zones for different MGs across Europe. We characterized which MG and which E gene haplotype works in which environment and found that this is mainly determined by the latitude. The results might be very useful for grower as well as for breeders to help defining target environments and thereby improve the adaptation. However, the lack of data concerning maturity and yield in combination with E genes hampers a clear conclusion on which haplotype or even MG is best adapted to which European growing conditions.

Especially for northern locations differences in temperature sensitivity and cold tolerance become more important. Perla and Diamant two varieties from Serbia were very early in southern regions, most probably due to the temperature sensitivity which promoted an earlier maturation, whereas with the E genes haplotype containing more dominant alleles, the variety was not photoperiodic insensitive enough for high latitude and/or the heat sums were not enough leading to the early maturing in the southern region but very late in northern locations. In the southern region the temperature might have been the driving force to fasten the maturation, whereas at higher latitudes combined with lower temperatures environment, the photoperiod hampered the early maturation.

The yield potential of a line is not only affected by maturity class but also by several different environmental factors, meaning the yield potential of a line is directly affected by the yield potential of an environment. That's why it is of great importance to characterize environments as specific as possible by envirotyping. According to Y. Xu (2016) environmental factors can be classified into five major groups, climate, soil, crop canopy, crop management and companion

organisms, each containing several subgroups that describe important environmental factors affecting plant growth and development. Even concentrating on the major group climate demands a precise dissection of complex environmental factors such as temperature accumulation, solar radiation, cold or low temperature risk, drought or rainfall distribution for both target environments and specific genotypes. Prerequisites for such an analysis are very detailed environmental data as for example daily climate data linked to trial locations which are often not available; the same is true for our conducted studies. High-throughput genotyping, phenotyping, and envirotyping should aim towards multi environmental trials in applied plant breeding as reliable data basis for selection and tackling genotype-by-environment interactions. Given these data, crop growth models can be used to dissect environmental factors and thereby predict yield and other traits (Cooper, Technow, Messina, Gho, & Totir, 2016).

Recent studies were published evaluating several environmental factors on different soybean traits from seed yield to quality traits in France and Germany (Boulch, Elmerich, Djemel, & Lange, 2021; Sobko et al., 2020). Both identified precipitation as main environmental factor determining seed yield by evaluating varieties of MG00 and MG000. Whereas Sobko et al. (2020) found also solar radiation to be significantly positive correlated with seed yield, Boulch et al. (2021) found the duration of the flowering to physiological maturity period under rainfed conditions with an optimum around 60 days as very important parameter affecting seed yield. According to Boulch et al. (2021) the period from flowering to physiological maturity is strongly correlated with cold stress, maximum temperature average, precipitation and water stress index. But of course the period from flowering to physiological maturity is as well determined by genotype, either directly by the *E* gene haplotypes (Kurasch, Hahn, Leiser, Vollmann, et al., 2017) or by the response of the genotype to the given environmental factors as chilling tolerance and sensitivity to water stress. Extending the period by an earlier planting as suggested by Boulch et al. (2021) could be an option by selecting genotypes that are more robust to earlier planting, through an improved cold tolerance and faster emergence at cooler temperatures.

Thus, besides flowering and maturity there are three main characteristics that are important for a successful adaptation to European target environments which are precipitation, temperature accumulation and cold tolerance. These traits are probably the most limiting factors for the broadening of the growing area and have to be addressed by plant breeding.

Implementation into breeding programs

The next step is to implement and discuss how the results achieved regarding adaptation can be integrated into a soybean breeding program or how it can be optimized with the new knowledge.

The starting point is therefore the method of soybean line development which is typically based on the single seed descent method as described in table 1. SSD was firstly proposed by Goulden (1941) with the aim to rapidly inbreed populations before evaluating individual lines in contrast to to pedigree method where the evaluation and selection takes place at the same time as the inbreeding to reach homozygosity. The main advantages of SSD are the rapid generation advance due to the use of off-season nurseries, the maintenance of an unbiased broad germplasm base until the evaluation of the lines starts; additionally the method is very labor- and time-efficient and allows a relatively easy handling of large numbers of lines. Extensive testing might be at least two testing steps with 6 and more locations.

Table 1: Classical SSD soybean breeding scheme (Example)

Year	Season	Activity	Harvest
N	Summer	Cross	Bulk
	Winter	F_1	Bulk
N+1	Summer	Grow F_2	Harvest one F_3 seed per plant and all seeds are bulked
	Winter	Grow bulk of F_3	Harvest one F_4 seed per plant and all seeds are bulked
N+2	Summer	Grow bulk of F_4	Harvest one F_5 seed per plant and all seeds are bulked
	Winter	Grow bulk of F_5 seed	Harvest individual plants
N+3	Summer	Grow $F_{5:6}$ lines in rows	Select among rows and harvest selected rows in bulk
	Winter		
N+4	Summer	begin testing of F_5 derived lines in several locations	

*pop size will decrease with each generation, due to lack of germination, lack of seed set, etc.

The first and most important decision in designing or optimizing a breeding program is the definition of target environments. Since we now know roughly which MG and E haplotype matures in which MEV due to the map we established in 2017, the target environment can be defined according to the map. Europe has to be split probably at least according to the MEV identified in 2017 and breeding programs should be divided according to the MEVs. Within one MEV up to three MGs could be cultivated, depending on the farmers' risk-taking, since earlier varieties will not have the highest yield potential compared to later varieties but will be more stable over years and at marginal sites. If the target environment is Central- Northern Europe, for example, the breeding goal is to select a superior soybean line which is best adapted to that target environment and the tested lines very likely should be of MG 000 containing *e1-nl*. Important characteristics of the testing environments are the clear differentiation among lines, especially under maturity evaluation and the clear representation of the target environment, to ensure the selection of the right material.

Starting from the first cross of two varieties, after defining the target environment and thereby the breeding goal regarding adaptation, the choice of the parental lines is the most crucial decision to take for a breeder. It is important to choose the parents in the most targeted way thus knowing the E genes of varieties by using an appropriate marker system is very advantageous. From the knowledge of the parental E gene haplotypes the segregation is predictable, thus controllable and many possibilities how to use this information are conceivable. Different breeding strategies can be deduced. Crossing two elite lines while staying in the same E haplotype is an easy way to prevent a strong segregation and therefore the planning of the testing location is simplified. As we have seen the choice of the parents based only on the MG is not an appropriate method since the segregation for maturity was very strong (Kurasch, Hahn, Leiser, Starck, et al., 2017).

Choosing lines with the same or almost the same E haplotype would be a good strategy for selecting elite lines and probably go fast by using more locations without the risk that the material is not adapted. As a rule of thumb a population size of around 200 lines in the F_5 generation is often sought, which means to choose more plants in the F_2 stage since due to the lack of germination or seed set the number of lines will decrease from F_2 to F_5 (table 1). Depending on the similarity of the parental lines it is also possible to adapt the number of F_2 plants, if the parents are very similar, the number can be reduced and therefore probably the number of locations can be increased for the first testing stage.

Another strategy is to cross two diverse lines regarding E haplotypes with the aim for example to combine earliness with higher yield potential but both known elite material. By choosing the parents according to the E haplotypes the segregation can be controlled. The more diverse the parents, stronger segregation will occur, the more similar the E haplotypes the less segregation will occur. The number of F_2 plants very likely has to be increased the more diverse the parents are to find the appropriate E haplotypes in the progenies most suitable to the target environment or to decide which E haplotypes to test in which environment. This means in effect that the progenies have to be tested with the use of molecular markers to know the E haplotypes before starting the extensive testing. A good example is the cross between P5 and P4 where the progenies segregated very strongly due to the diverse E haplotypes of the parents. However, the very rare recessive haplotype occurred in the progenies suggesting that these lines are more suitable to northern environments where photoperiod insensitivity is a prerequisite, whereas the progenies carrying the *e1-as* alleles are supposed to be later maturing and other locations should have been used to test them or discard them depending on the breeding goal. Knowing the E haplotypes enables also to cross certain parents with the aim to identify a very specific E gene

combination, as for example the very rare recessive *E* gene combination. Therefore the population size has to be large enough to identify these haplotype. Being able to analyze the *E* haplotypes enlarges the germplasm of usable cultivars or accession from all over the world and even almost all maturity groups. It enables to exchange lines from similar target populations of environments which could be from Canada, North China and Japan. This might be interesting not only for adaptation but as well to transfer specific traits into the European germplasm as for example tofu traits from Japanese cultivars or resistance traits. According to studies with worldwide material, the Chinese material at least regarding the *E2* genes is more similar to our European material than the American germplasm (Langewisch et al., 2017; Li et al., 2017; Miladinović et al., 2018; Zhai et al., 2014). Therefore, the use of genetic resources might be more easy and effective in a breeding program.

In the study from 2017 (Kurasch, Hahn, Leiser, Starck, et al., 2017) the maturity of the progenies was estimated by the phenotypic performance of the parents, which resulted in a field trial that could not be harvested due to the unexpected large differences in maturity, a case which was already predicted as a possible scenario when selection is based only on phenotype (Zhai et al. (2014). Predict the material by the *E* haplotypes of the parents or even genotype the material before testing is, thus, much more efficient and helps to take the right decision for selecting the test location or making a selection decision and testing only the lines with most promising *E* haplotype. So, knowing the allelic state of parental lines can help to control the outcome of a certain cross to achieve appropriate maturity. Since different alleles of an *E* gene lead to similar phenotypes regarding maturity it might be difficult to fix the allelic state by phenotypic selection leading to heterozygous allele combinations at the *E* genes. It might be advantageous to know the allelic state to keep diversity in the population due to closely linked genes that may contribute to other traits resulting in differences in the phenotype (Tardivel, Sonah, Belzile, & O'Donoughue, 2014).

Potential of phenotypic and genomic selection

Compared to the adaptation traits, quality traits as protein content and oil content are less affected by the environment showing therefore high heritabilities which makes them easier to select for either by phenotypic or genomic selection. Quality traits that are easy to screen and show high heritability and therefore a high prediction accuracy as protein content in our study will have a high response to selection either by phenotype or genome. The potential for GS is therefore high for the quality traits protein and oil content (Duhnen et al., 2017; Kurasch, Hahn, Leiser, Starck, et al., 2017; Kurasch, Leiser, et al., 2018) but the advantage over phenotypic

selection is not as high as for more complex traits. The main advantage of GS over phenotypic selection in this case might be the possible time reduction during the breeding cycle and the resource optimization (Technow, Bürger, & Melchinger, 2013). However, the breeding method which is mainly SSD has to be adapted in order to fully exploit the advantage of time reduction. The main advantage of the SSD method was the separation of inbreeding and the evaluation with a rapid generation advance before testing the almost homozygous lines. But with the use of the genomic tools, this might not anymore be a benefit since it is not possible to reduce the time for inbreeding. Time reduction can take place due to less testing steps by replacing first evaluation by GS and only test the most promising lines in a more intensive way. The disadvantage of this would be not to reduce the time for the inbreeding until first testing and therefore all F_2 plants would be advanced until $F_{5:6}$ and kept until the first evaluation. Advance all lines until $F_{5:6}$ without selection is costly knowing only 10-20 % will advance afterwards during the first field tests. One possibility might be to evaluate on genomic level at an early stage as for example F_3 or F_4 and only advance the best ones based on genomic value until $F_{5:6}$. This would enable to allocate the resources in different ways by increasing the population size of F_2 generation in order to apply more strict selection intensity and to test less lines at more locations and therefore more intensively leading to higher selection gain. A radical method to reduce the time for inbreeding during the breeding process is described by Cobb et al. (2019) called the rapid cycle recurrent selection where $F_{2:3}$ lines are chosen as parents with the aim to improve the breeding value of a population meaning to split the breeding process into parents selection with population enhancement on one side and commercial line development with highly inbred lines on the other side. Therefore the time for inbreeding during the breeding process is reduced to almost zero but requires a special process to extract commercial lines. This allows reducing the cycle time up to one year and therefore increasing the genetic gain drastically per year and totally. The combination of rapid cycle recurrent selection with a GS approach might be a process to further examine and evaluate.

Even though the potential for oil and protein using GS might be high, the negative correlation with grain yield has to be taken into account. As in other crops, e.g. wheat, the simultaneous improvement of protein content and grain yield is a challenge due to the negative correlation (Michel et al., 2019b; Rapp et al., 2018). Depending on the breeding goal or market specification (eg. feed vs. food) different approaches can be used to tackle the problem. For soybean which is mainly used as feed, the protein content is not as important as for food-grade soybeans for two reasons. First, compared to all other crops soybean already has the highest protein content and secondly it is only important for farmers that are feeding the whole soybean since for crushed

soybean the protein is already concentrated, thus, the protein content of raw soybean is of minor importance. Two different strategies can be distinguished where either grain yield is of higher importance with the aim to select high yielding lines while maintaining a sufficiently high protein content, e.g. for feeding purposes, or protein content is of higher importance with the aim to select high-protein lines while keeping an acceptably high grain yield, e.g. for food purpose. In the first case we could show that using the protein yield per ha can be a good tool to increase the yield and keep the protein content at an optimum level since protein yield is highly correlated with grain yield. Rapp et al. (2018) also concluded that selecting on protein yield will mainly increase grain yield in wheat, whereas the grain protein deviation (GPD) suggested by (Monaghan, Snape, Chojecki, & Kettlewell, 2001), which are the residuals of the regression of protein content on grain yield, mainly increases the protein content, but a combination of both indices could balance the simultaneous selection of protein content and grain yield. Michel et al. (2019b) also compared different selection methods for a simultaneous selection of yield and protein content in wheat by using grain yield, protein content, protein yield and four different restriction indices that either hold one of the tree traits stable or increase one of the three traits. Supporting Rapp et al. (2018), Michel et al. (2019b) concluded that a large genetic improvement can be achieved by simultaneous genomic selection for grain yield and protein content when combining the selection indices. Michel et al. (2019b) showed in addition a strong improvement of prediction accuracy of preliminary yield trials when combining phenotypic and genomic information in a genomics-assisted selection approach for the single traits as well as for the selection indices highlighting the importance to use phenotypic and genomic selection in an integrated way to have best results and selection responses.

Both showed that it is possible or even easy to select for either quality or grain yield by using restriction indices that hold one trait stable while increasing the other. Thus, GPD would be a good index to apply for food-grade soybean. Although high protein contents are demanded by tofu producers we could show neither a clear correlation with tofu yield traits nor tofu quality traits as for example tofu hardness (Kurasch, Hahn, et al., 2018). Since the actual protein content is not explaining other tofu traits, the protein subunit composition might be of greater importance for the tofu texture (Zhang et al. 2018). Thus, a strategy could be the investigation of the proportion of the storage proteins glycenin and β-conglycinin by a QTL analysis and then use these QTL to select for a specific protein subunit composition. However screening the proportion of the storage proteins glycenin and β-conglycinin is much more laborious compared to the easy screening of the protein content by NIRS technology and in addition the storage proteins might not explain the total variation of tofu texture and other tofu traits.

Instead of selecting or focusing on a very specific protein composition it might be advantageous to test and select for end-use traits (Reid & Cober, 2018). Because finally it is not only the goal to select lines with a specific protein composition but with an improved end use trait like tofu yield, tofu quality or for example baking quality in bread wheat (Michel et al., 2018, 2019a). The specific protein composition will probably explain a certain amount of the variation to the tofu traits. By phenotyping the end-use trait other influencing factors will be captured even though they remain unknown. Based on an example of salinity tolerance Cobb, Biswas, & Platten, (2019) compare in their review a QTL controlling 20 % of the variation in visual injury symptoms in the target environment under the stress of interest with a QTL controlling 20 % of the variation in leaf sodium content, which is associated with salinity tolerance, and come to the conclusion that it is preferable and more reliable to phenotype and select the actual trait under selection, which is in this case the visual injury symptoms, thus the end-use trait. This highlights the importance of the actual production of test tofu samples and the screening of the actual tofu traits. Focusing on a specific physiolocigal pathway is dangerous because it is only limited in the proportion of explained variation and other reasons will be neglected that in sum contribute to the actual target trait.

Due to the complexity of the tofu traits and the laborious evaluation, these traits should profit the most from genomic selection. Even though the results of the prediction accuracies for tofu traits in our study were moderate to low, genomic selection can be a promising tool. The method with the GBS TASSEL pipeline and prediction model as well as the quality of the genotypic data was proven to be good since the mean prediction accuracy for protein content was rather high and showed less variation. The difference between protein content, tofu yield traits and tofu hardness is that protein content is easy to screen and shows a high heritability whereas tofu yield traits and tofu hardness are very difficult to screen because of the small-scale production of tofu samples which was shown to be highly influenced by the lab and the location (Kurasch, Hahn, et al., 2018). The main bottleneck for GS in general is therefore the production of highly precise phenotypic data which has become a major issue in the research community (Bernal-Vasquez et al., 2014; Cobb, Declerck, Greenberg, Clark, & McCouch, 2013; Voss-Fels et al., 2019). It is deceptive to assume that the effort of phenotyping can be reduced when using a genomics-based breeding approach, as you have to constantly improve the phenotyping because GS depends crucially on the quality of the data, either phenotypic, genotypic or the size of the training population. Phenotypic data were in the beginning of GS probably one of the most underrated factors among the mentioned factors. Training sets for GS have to be regularly updated with new lines coming from the own breeding program or with external lines that should be incorporated

in the breeding program to have a good representation of the breeding material and thus maintain high accuracies. Much effort has to be put to a smaller and more targeted quantity of samples. In return, the tofu samples should be produced by well-trained staff to ensure tofu samples not impacted by the lab for a precise evaluation of tofu yield and quality. In addition, it might be possible to find some major QTL explaining more variation of the tofu yield and quality which can be used to improve the GS model by giving special weight to these markers (Bernardo, 2014). Several studies have shown that weighting QTL in a GS model increase the predictive performance (Boeven et al., 2016; Jähne et al, 2019; Moore et al., 2017). However, Bernardo (2014) recommended that weighted QTL used in GS models should account for more than 10 % of the genetic variance to be advantageous.

Establishing a good training set for tofu GS across the whole breeding program by phenotyping all new line candidates that come from one breeding cycle thereby improving the training set regularly and perform QTL analysis to detect specific markers in special tofu crosses can, thus, be a valuable combination to identify superior tofu soybean varieties. This enables the application of GS for tofu traits to the whole breeding program and not only focusing on specific crosses. In a breeding program crosses are conducted with a specific aim which can be staying in a certain maturity or making crosses to become earlier or to increase the tofu suitability. But tofu suitability is really a value-added trait and not a must-have trait, but is essential for a specific market, therefore, it is much more beneficial to screen the whole breeding program instead of performing a lot of crosses for this specific target. Furthermore, the sensory traits like tofu flavor become very important which are very difficult to screen and should also be incorporated in the phenotyping strategy.

Conclusions

Finally, putting all pieces together, our results can be used to design a meaningful breeding program. The must-have traits like adapted maturity (to enable harvesting) and grain yield can be addressed by fitting the right maturity into the right environment using E gene analysis and environmental analysis. Incorporating all the information from mega environments and maturity E gene haplotype into designing a valuable test location network allows an optimum allocation of resources. Thus, the outcome is much more accurate and precise; the yield potential can be fully exploited. Further methods like envirotyping or the use of crop models can help to better understand the soybean lines not only with respect to the maturity but also in their response to environmental factors, and the interaction at each developmental stage. E genes analysis can also be used to choose the right parents for crossing to get more targeted crosses and populations.

For the value-added traits like protein content and tofu traits, GS approaches seem to be worthy. Genotyping-by-sequencing worked very well for genomic selection for protein content, but in future will also be very interesting for tofu traits. The benefits of weighted GS by combining QTL analysis and GS should improve tofu trait prediction. Combining environmental analysis, *E*-gene analysis, GS approaches (QTL and GS), and tofu phenotyping will help to become more accurate and targeted in the way of selection thereby increasing the genetic gain. However, the breeding scheme has to be adapted in the way to become faster to be able to benefit mostly in an increase of the genetic gain. For example by a faster recycling of lines into new crosses or by an earlier genomic selection step enabling only to test the most promising lines based on GEBV. Combining the results helps to optimize the resources by using more targeted locations enabling identification of more adapted lines leading to higher heritabilities due to reduction of genotype-by-environment interaction. With all the results produced and when its going even more into detail for the environments like envirotyping or crop models as well as the knowledge about the *E* genes, a breeder can be much more precise and focused meaning placing the right lines into the right environment. The more targeted and specific, the more complex a breeding program gets, which requires adequate tools for data management including phenotypic, genotypic and environmental data, good digital tools for data collection, and data analysis to enable a quick and precise breeding decision, for the most important steps in a breeding program that are choosing parents, making crosses, choosing the test network, identifying and selecting superior individuals based on phenotype and genotype.

Summary

Soybean (*Glycine max* Merr.) is one of the major crops in the world providing an important source of protein and oil for food and feed; however it is still a minor crop in Central Europe. Soybean cultivation can play an important role in a more sustainable agricultural system by increasing local and regional protein production in Europe. The demand for locally produced soybean products is still growing in Europe. The key for a successful establishment of soybean cultivation in Europe is adaptation of soybean varieties to the Central European growing conditions. For the latitudinal adaptation to long-day conditions in Central to Northern Europe, an adapted early flowering and maturity time is of crucial importance for a profitable cultivation. The key traits flowering and maturity are quantitatively inherited and mainly affected by photoperiod responsiveness and temperature sensitivity. The most important loci for an early flowering and maturity are *E1-E4* and the various allelic combinations condition soybean flowering and maturity time and therefore strongly contribute to the wide adaptability (Jiang et al., 2014; Tsubokura et al., 2014; M. Xu et al., 2013).

Besides the main usage as protein source for animal feeding, soybean is also a very valuable source for human consumption. Soy-based food plays a pivotal role in the Asian cuisine offering very diverse products with tofu as most prominent product. Tofu is enjoying ever greater popularity in Europe, as it is one of the best sources of plant protein with additional health benefits, rich in essential amino acids, beneficial lipids, vitamins, and minerals, as well as other bioactive compounds, such as isoflavones, soyasaponin, and others, (Lima et al., 2017; Zhang et al., 2018). Thus, plant breeding has to provide not only well-adapted varieties with good agronomic and quality properties, but also provide varieties well-suited to the further processing into soymilk and tofu. Therefore, a good knowledge about the breeding target, how to assess it and how it is inherited is crucial.

In a first large mega-environment (MEV) study with 75 European varieties from five early maturity groups (MGs 000–II) grown at 22 locations in 10 countries across Europe, six mega-environments in latitudinal direction were revealed. Analysis of maturity identified several haplotypes for the allelic variants at the *E1*, *E2*, *E3* and *E4* genes, with each *E* haplotype comprising varieties from different MGs and vice versa varieties with the same *E* haplotype can exhibit different flowering and maturity characteristics. However, several groups of varieties with phenotypic similarity were identified that are optimally adapted to the different growing regions in Europe.

Besides adapted maturity to Central Europe a good knowledge about inheritance of agronomic traits is crucial for successful breeding. By evaluating 1008 $F_{5:8}$ recombinant inbred lines derived

from an incomplete half-diallel cross of five early-maturing European soybean culitvars evaluated in mulit-location field trials, significant genotypic variances, high heritabilities ($h^2 > 0.7$), and transgressive segregation were observed for grain yield, thousand-kernel weight, plant height, protein content, and oil content. Further, a complex relationship between grain yield, plant height, and maturity was revealed, but nevertheless indicated that breeding of adapted and high-yielding varieties is feasible. Results on available phenotypic variation, variance components, heritabilities, and trait correlations can expedite soybean breeding targeted at Central Europe.

Based on 215 recombinant inbred lines derived from two populations and grown at three locations that were evaluated for tofu traits in a bench-scale tofu laboratory, significant genotypic variance components were observed, but an even stronger contribution of the location, resulting in moderately high heritabilities with $h^2 > 0.6$. A network analysis of the evaluated tofu traits showed in general no association with any of the agronomic traits, indicating no obstacle in improving agronomic traits and tofu suitability simultaneously.

As phenotyping of tofu-related traits is cumbersome, costly and time-consuming, genomics-assisted approaches would be of great benefit compared to conventional phenotypic selection. To investigate the potential of marker-based approaches to assist selection for tofu quality, the genetic architecture of the tofu traits was investigated in the three recombinant inbred line populations evaluated for several tofu-related traits. QTL identified for soymilk weight, tofu hardness and tofu value were population-specific and explained only a small to moderate proportion of the genetic variation and no QTL were found for tofu weight and tofu yield. The results of the QTL mapping illustrated that the tofu traits are of a highly quantitative nature. Results from the genome-wide prediction were promising for seed quality traits and soaking traits, but for the tofu traits, the prediction accuracies have to be increased further based on improvements of the phenotyping procedure.

The conducted studies covered a broad range of aspects relevant to improve a soybean breeding program. By combining environmental analysis, *E*-gene analysis, genomic approaches (QTL mapping and genomic prediction), and tofu phenotyping, breeder decisions become more accurate and targeted in the way of selection thereby increasing the genetic gain. In addition, combining the results of the different aspects helps to optimize the resources of a breeding program. Increasing the knowledge about the different aspects from environment to tofu QTL enables a breeder to be more precise and focused. But the more targeted and specific, the more complex a breeding program gets, which requires adequate tools to handle all the different information in a meaningful and efficient way to enable a quick and precise breeding decision.

Zusammenfassung

Die Sojabohne (*Glycine max* Merr.) ist eine der wichtigsten Nutzpflanzen der Welt und stellt eine wichtige Protein- und Ölquelle für Lebens- und Futtermittel dar; in Mitteleuropa spielt die Sojabohne jedoch immer noch eine untergeordnete Rolle im Anbau. Der Sojabohnenanbau kann eine wichtige Rolle in einem nachhaltigeren Agrarsystem spielen, indem er die lokale und regionale Proteinproduktion in Europa steigert. Die Nachfrage nach lokal produzierten Sojabohnenprodukten wächst in Europa weiter. Der Schlüssel für eine erfolgreiche Etablierung des Sojaanbaus in Europa ist die Anpassung der Sojasorten an die mitteleuropäischen Anbaubedingungen. Für die Breitenanpassung an Langtagbedingungen in Mittel- bis Nordeuropa ist eine angepasste frühe Blüte- und Reifezeit von entscheidender Bedeutung für einen ertragreichen Anbau. Die Schlüsselmerkmale Blüte und Reife werden quantitativ vererbt und hauptsächlich durch die Photoperioden- und Temperaturempfindlichkeit beeinflusst. Die wichtigsten Genorte für eine frühe Blüte und Reife sind *E1-E4*. Die verschiedenen Allelkombinationen bedingen die Sojabohnenblüte und Reifezeit und tragen daher stark zur breiten Anpassungsfähigkeit bei (Jiang et al., 2014; Tsubokura et al., 2014; M. Xu et al., 2013).

Neben der Hauptverwendung als Proteinquelle für die Tierfütterung ist Soja auch eine sehr wertvolle Quelle für die menschliche Ernährung. Lebensmittel auf Sojabasis spielen eine zentrale Rolle in der asiatischen Küche, die sehr unterschiedliche Produkte anbietet, wobei Tofu das wichtigste Produkt ist. Tofu erfreut sich in Europa immer größerer Beliebtheit, da er eine der besten pflanzlichen Proteinquellen mit zusätzlichem Gesundheitsnutzen ist, reich an essentiellen Aminosäuren, nützlichen Lipiden, Vitaminen und Mineralstoffen sowie anderen bioaktiven Verbindungen wie Isoflavonen, Sojasaponin und andere (Lima et al., 2017; Zhang et al., 2018). Daher muss die Pflanzenzüchtung nicht nur gut angepasste Sorten mit guten agronomischen und qualitativen Eigenschaften liefern, sondern auch Sorten, die sich für die Weiterverarbeitung zu Sojamilch und Tofu gut eignen. Gute Kenntnisse über das Zuchtziel, wie es zu beurteilen ist und wie es vererbt wird, sind daher entscheidend.

In einer ersten großen Mega-Environment (MEV)-Studie mit 75 europäischen Sorten aus fünf Frühreifegruppen (MGs 000–II), die an 22 Standorten in 10 europäischen Ländern angebaut wurden, wurden sechs MEVs in Breitenrichtung entdeckt. Die Analyse der Reife identifizierte mehrere Haplotypen für die Allelvarianten an den *E1*-, *E2*-, *E3*- und *E4*-Genen, wobei jeder *E*-Haplotyp Sorten von verschiedenen Reifegruppen umfasst und umgekehrt Sorten mit demselben *E*-Haplotyp unterschiedliche Blüte- und Reifeeigenschaften aufweisen können. Es wurden jedoch mehrere Sortengruppen mit phänotypischer Ähnlichkeit identifiziert, die optimal an die unterschiedlichen Anbaugebiete in Europa angepasst sind.

Neben einer an Mitteleuropa angepassten Reife sind gute Kenntnisse über die Vererbung agronomischer Merkmale für eine erfolgreiche Züchtung entscheidend. Bei der Bewertung von 1008 rekombinanten $F_{5:8}$-Inzuchtlinien, die aus einer unvollständigen Halb-Diallel-Kreuzung von fünf frühreifenden europäischen Sojabohnensorten stammten, die in Feldversuchen mit mehreren Standorten bewertet wurden, wurden signifikante genotypische Varianzen, hohe Heritabilitäten ($h^2 > 0{,}7$) und transgressive Segregation beobachtet für Kornertrag, Tausendkorngewicht, Pflanzenhöhe, Proteingehalt und Ölgehalt. Darüber hinaus wurde ein komplexer Zusammenhang zwischen Kornertrag, Pflanzenhöhe und Reife festgestellt, aber dennoch gezeigt, dass die Züchtung von angepassten und ertragsstarken Sorten möglich ist. Ergebnisse zu verfügbaren phänotypischen Variationen, Varianzkomponenten, Heritabilitäten und Merkmalskorrelationen können eine gezielte Sojabohnenzüchtung in Mitteleuropa beschleunigen.

Basierend auf 215 rekombinanten Inzuchtlinien, die aus zwei Populationen stammten und an drei Standorten gezüchtet wurden und die in einem Tofu-Labor im Labormaßstab auf Tofu-Merkmale untersucht wurden, wurden signifikante genotypische Varianzkomponenten beobachtet, aber ein noch stärkerer Beitrag des Standorts, was zu mäßig hohen Heritabilitäten führte mit $h^2 > 0{,}6$. Eine Netzwerkanalyse der bewerteten Tofu-Merkmale zeigte im Allgemeinen keine Assoziation mit einem der agronomischen Merkmale, was darauf hindeutet, dass es kein Hindernis für die gleichzeitige Verbesserung der agronomischen Merkmale und der Tofu-Eignung gibt.

Da die Phänotypisierung von Tofu-bezogenen Merkmalen umständlich, kostspielig und zeitaufwändig ist, wären genomisch-basierte Ansätze im Vergleich zur konventionellen phänotypischen Selektion von großem Vorteil. Um das Potenzial von marker-basierten Ansätzen zur Unterstützung der Selektion auf Tofuqualität zu untersuchen, wurde die genetische Architektur der Tofumerkmale in den drei rekombinanten Inzuchtlinienpopulationen untersucht, die auf verschiedene Tofu-bezogene Merkmale untersucht wurden. Die identifizierten QTL für Sojamilchgewicht, Tofufestigkeit und Tofuwert waren populationsspezifisch und erklärten nur einen kleinen bis moderaten Anteil der genetischen Variation und es wurden keine QTL für Tofugewicht und Tofuertrag gefunden. Die Ergebnisse der QTL-Kartierung zeigten, dass die Tofu-Merkmale stark quantitativ vererbt werden. Die Ergebnisse der genomweiten Vorhersage waren für Samenqualitätsmerkmale und Einweichmerkmale vielversprechend, während für die Tofumerkmale die Vorhersagegenauigkeiten aufgrund von Verbesserungen des Phänotypisierungsverfahrens weiter erhöht werden müssen.

Die durchgeführten Studien deckten ein breites Spektrum von Aspekten ab, die für die Verbesserung eines Sojabohnenzuchtprogramms relevant sind. Durch die Kombination von

Umweltanalyse, *E*-Gen-Analyse, genomischen Ansätzen (QTL-Mapping und genomische Vorhersage) und Tofuphänotypisierung werden Züchterentscheidungen genauer und zielgerichteter in der Selektion, wodurch der Zuchtfortschritt erhöht wird. Darüber hinaus hilft die Kombination der Ergebnisse der verschiedenen Aspekte, die Ressourcen eines Zuchtprogramms zu optimieren. Die Erweiterung des Wissensstands über die verschiedenen Aspekte von der Umwelt bis zum Tofu-QTL ermöglicht es einem Züchter, präziser und fokussierter zu sein. Doch je gezielter und spezifischer, desto komplexer wird ein Zuchtprogramm, das adäquate Werkzeuge benötigt, um mit all den unterschiedlichen Informationen sinnvoll und effizient umzugehen, um damit dann eine schnelle und präzise Zuchtentscheidung zu ermöglichen.

References

Bao, Y., Vuong, T., Meinhardt, C., Tiffin, P., Denny, R., Chen, S., … Young, N. D. (2014). Potential of Association Mapping and Genomic Selection to Explore PI 88788 Derived Soybean Cyst Nematode Resistance. *The Plant Genome*, 7(3), 1–13. https://doi.org/10.3835/plantgenome2013.11.0039

Barabaschi, D., Tondelli, A., Desiderio, F., Volante, A., Vaccino, P., Valè, G., & Cattivelli, L. (2015). Next generation breeding. *Plant Science*, 242, 3–13. https://doi.org/10.1016/j.plantsci.2015.07.010

Bernal-Vasquez, A.-M., Möhring, J., Schmidt, M., Schönleben, M., Schön, C.-C., & Piepho, H.-P. (2014). The importance of phenotypic data analysis for genomic prediction - a case study comparing different spatial models in rye. *BMC Genomics*, 15(1), 646. https://doi.org/10.1186/1471-2164-15-646

Bernard, R. L. (1971). Two Major Genes for Time of Flowering and Maturity in Soybeans. *Crop Science*, 11, 242–244. https://doi.org/10.2135/cropsci1971.0011183X001100020022x

Bernardo, R. (2008). Molecular Markers and Selection for Complex Traits in Plants: Learning from the Last 20 Years. *Crop Science*, 48(5), 1649. https://doi.org/10.2135/cropsci2008.03.0131

Bernardo, R. (2014). Genomewide Selection when Major Genes Are Known. *Crop Breeding & Genetics*, 54(1), 68–75. https://doi.org/https://doi.org/10.2135/cropsci2013.05.0315

Boeven, P. H., Longin, C. F. H., Leiser, W. L., Kollers, S., Ebmeyer, E., & Würschum, T. (2016). Genetic architecture of male floral traits required for hybrid wheat breeding. *Theoretical and Applied Genetics*, 129(12), 2343–2357. https://doi.org/10.1007/s00122-016-2771-6

Bonato, E. R., & Vello, N. A. (1999). E6 , a Dominant Gene Conditioning Early Flowering. *Genetics and Molecular Biology*, 22(2), 229–232.

Boulch, G., Elmerich, C., Djemel, A., & Lange, B. (2021). Evaluation of soybean (Glycine max L.) adaptation to northern European regions under different agro-climatic scenarios. *In Silico Plants*. https://doi.org/https://doi.org/10.1093/insilicoplants/diab008

Bues, A., Preissel, S., Reckling, M., Zander, P., Kuhlman, T., Topp, K., … Murphy-Bokern, D. (2013). *The environmental role of protein crops in the new common agricultural policy*.

Buzzell, R. I. (1971). Inheritance of a soybean flowering response to fluorescent-daylength conditions. *Canadian Journal of Genetics and Cytology*, (13), 703–707. https://doi.org/10.1139/g71-100

Buzzell, R. I., & Voldeng, H. D. (1980). Inheritance of insensitivity to long daylength. *Soybean Genet. Newsl*, 7, 26–29.

Cahill, D. J., & Schmidt, D. H. (2004). Use of Marker Assisted Selection in a Product Development Breeding Program. *Proceedings of the 4th International Crop Science Congress*, 1–9. Retrieved from www.cropscience.org.au

Cobb, J. N., Declerck, G., Greenberg, A., Clark, R., & McCouch, S. (2013). Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype-phenotype relationships and its relevance to crop improvement. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, 126(4), 867–887. https://doi.org/10.1007/s00122-013-2066-0

Cobb, J. N., Juma, R. U., Biswas, P. S., Arbelaez, J. D., Rutkoski, J., Atlin, G., … Ng, E. H. (2019). Enhancing the rate of genetic gain in public-sector plant breeding programs: lessons from the breeder's equation. *Theoretical and Applied Genetics*, 132(3), 627–645. https://doi.org/10.1007/s00122-019-03317-0

Cober, E. R., & Voldeng, H. D. (2001). A new soybean maturity and photoperiod-sensitivity locus linked to E1 and T. *Crop Science*, 41(3), 698–701. https://doi.org/10.2135/cropsci2001.413698x

Cober, Elroy R., Molnar, S. J., Charette, M., & Voldeng, H. D. (2010). A New Locus for Early Maturity in Soybean. *Crop Science*, 50(2), 524. https://doi.org/10.2135/cropsci2009.04.0174

Cober, Elroy R., & Morrison, M. J. (2010). Regulation of seed yield and agronomic characters by

photoperiod sensitivity and growth habit genes in soybean. *Theor. Appl. Genet.*, *120*(5), 1005–1012. https://doi.org/10.1007/s00122-009-1228-6

Concibido, V. C., Denny, R. L., Lange, D. A., Orf, J. H., & Young, N. D. (1996). RFLP Mapping and Marker‑ Assisted Selection of Soybean Cyst Nematode Resistance in PI 209332. *Crop Science*, *36*, 1643–1650. https://doi.org/https://doi.org/10.2135/cropsci1996.0011183X003600060038x

Cooper, M., Technow, F., Messina, C., Gho, C., & Totir, L. R. (2016). Use of Crop Growth Models with Whole-Genome Prediction: Application to a Maize Multienvironment Trial. *Crop Science*, *56*(5), 2141. https://doi.org/10.2135/cropsci2015.08.0512

de Visser, C. L. M., Schreuder, R., & Stoddard, F. (2014). The EU's dependency on soya bean import for the animal feed industry and potential for EU produced alternatives. *OCL*, *21*(4), D407. https://doi.org/10.1051/ocl/2014021

Duhnen, A., Gras, A., Teyssèdre, S., Romestant, M., Claustres, B., Daydé, J., & Mangin, B. (2017). Genomic Selection for Yield and Seed Protein Content in Soybean: A Study of Breeding Program Data and Assessment of Prediction Accuracy. *Crop Science*, *57*(3), 1325. https://doi.org/10.2135/cropsci2016.06.0496

Elbasyoni, I. S., Lorenz, A. J., Guttieri, M., Frels, K., Baenziger, P. S., Poland, J., & Akhunov, E. (2018). A comparison between genotyping-by-sequencing and array-based scoring of SNPs for genomic prediction accuracy in winter wheat. *Plant Science*, *270*, 123–130. https://doi.org/10.1016/j.plantsci.2018.02.019

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS One*, *6*(5), e19379. https://doi.org/10.1371/journal.pone.0019379

European Commission. (2018). Community register of GM food and feed (Regulation EC 1829/2003). Retrieved November 22, 2018, from http://ec.europa.eu/food/dyna/gm_register/index_en.cfm

EUROSTAT. (2020). Rape, turnip rape, sunflower seeds and soya by area. Retrieved January 19, 2021, from https://ec.europa.eu/eurostat/databrowser/view/tag00100/default/table?lang=en

Fang, C., Ma, Y., Wu, S., Liu, Z., Wang, Z., Yang, R., … Tian, Z. (2017). Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. *Genome Biology.* https://doi.org/10.1186/s13059-017-1289-9

FAOSTAT. (2016). FAO Statistical Database. Retrieved November 16, 2018, from http://www.fao.org/faostat/en/#data/QC/visualize

Goulden, C. (1941). Problem in plant selection. In *Proc. Seventh Int.Genetical Congress (1939)* (pp. 132-133.). Edinburg.

Harada, K., Watanabe, S., Zhengjun, X., Tsubokura, Y., Yamanaka, N., & Anai, T. (2011). Positional Cloning of the Responsible Genes for Maturity Loci E1, E2 and E3 in Soybean. In *Soybean - Genetics and Novel Techniques for Yield Enhancement*. InTech. https://doi.org/10.5772/21085

Häusling, M. (2014). *The EU protein deficit: what solution for a long-standing problem? (2010/2111(INI)).*

Heffner, E. L., Sorrells, M. E., & Jannink, J. (2009). Genomic Selection for Crop Improvement. *Crop Science*, *49*(February), 1–12. https://doi.org/10.2135/cropsci2008.08.0512

Heslot, N., Jannink, J. L., & Sorrells, M. E. (2013). Using genomic prediction to characterize environments and optimize prediction accuracy in applied breeding data. *Crop Science*, *53*(3), 921–933. https://doi.org/10.2135/cropsci2012.07.0420

Hou, H. J., & Chang, K. C. (2004). Storage conditions affect soybean color, chemical composition and tofu qualities. *Journal of Food Processing and Preservation*, *28*(6), 473–488. https://doi.org/10.1111/j.1745-4549.2004.24015.x

Hymowitz, T., & Harlan, J. R. (1983). Introduction of soybean to North America by Samuel Bowen in 1765. *Economic Botany*, *37*(4), 371–379. https://doi.org/10.1007/BF02904196

Hymowitz, T., & Shurtleff, W. R. (2005). *Debunking Soybean Myths and Legends in the Historical and*

Popular Literature.

ISAAA. (2016). *Global Status of Commercialized Biotech/ GM Crops: 2016. ISAAA Brief* (Vol. 52). Ithaca, NY.

Jähne, F., Balko, C., Hahn, V., & Würschum, T. (2019). Cold stress tolerance of soybeans during flowering : QTL mapping and efficient selection strategies under controlled conditions, (March), 1–13. https://doi.org/10.1111/pbr.12734

Jarquín, D., Kocak, K., Posadas, L., Hyma, K., Jedlicka, J., Graef, G., & Lorenz, A. (2014). Genotyping by sequencing for genomic prediction in a soybean breeding population. *BMC Genomics*, *15*(1), 740. https://doi.org/10.1186/1471-2164-15-740

Jaureguy, L. M., Chen, P., & Scaboo, A. M. (2011). Heritability and correlations among food-grade traits in soybean. *Plant Breeding*, *130*(6), 647–652. https://doi.org/10.1111/j.1439-0523.2011.01887.x

Jia, H., Jiang, B., Wu, C., Lu, W., Hou, W., Sun, S., … Han, T. (2014). Maturity group classification and maturity locus genotyping of early-maturing soybean varieties from high-latitude cold regions. *PloS One*, *9*(4), e94139. https://doi.org/10.1371/journal.pone.0094139

Jiang, B., Nan, H., Gao, Y., Tang, L., Yue, Y., Lu, S., … Liu, B. (2014). Allelic combinations of soybean maturity Loci E1, E2, E3 and E4 result in diversity of maturity and adaptation to different latitudes. *PloS One*, *9*(8), e106042. https://doi.org/10.1371/journal.pone.0106042

Koester, R. P., Skoneczka, J. A., Cary, T. R., Diers, B. W., & Ainsworth, E. A. (2014). Historical gains in soybean (Glycine max Merr.) seed yield are driven by linear increases in light interception, energy conversion, and partitioning efficiencies. *Journal of Experimental Botany*. https://doi.org/10.1093/jxb/eru187

Kong, F., Nan, H., Cao, D., Li, Y., Wu, F., Wang, J., … Liu, B. (2014). A New Dominant Gene Conditions Early Flowering and Maturity in Soybean. *Crop Sci.*, *54*(6), 2529. https://doi.org/10.2135/cropsci2014.03.0228

Kumar, V., Rani, A., Solanki, S., & Hussain, S. M. (2006). Influence of growing environment on the biochemical composition and physical characteristics of soybean seed. *Journal of Food Composition and Analysis*, *19*(2–3), 188–195. https://doi.org/10.1016/j.jfca.2005.06.005

Kurasch, A. K., Hahn, V., Leiser, W. L., Starck, N., & Würschum, T. (2017). Phenotypic Analysis of Major Agronomic Traits in 1008 RILs from a Diallel of Early European Soybean Varieties. *Crop Science*, *57*, 726–738. https://doi.org/10.2135/cropsci2016.05.0318

Kurasch, A. K., Hahn, V., Leiser, W. L., Vollmann, J., Schori, A., Bétrix, C.-A., … Würschum, T. (2017). Identification of mega-environments in Europe and effect of allelic variations at maturity E loci on adaptation of European soybean. *Plant, Cell & Environment*, *40*, 765–778. https://doi.org/10.1111/pce.12896

Kurasch, A. K., Hahn, V., Miersch, M., Bachteler, K., & Würschum, T. (2018). Analysis of tofu-related traits by a bench-scale tofu production method and their relationship with agronomic traits in European soybean. *Plant Breeding*, *137*(3), 271–282. https://doi.org/10.1111/pbr.12581

Kurasch, A. K., Leiser, W., Bachteler, K., Miersch, M., Hahn, V., & Würschum, T. (2018). Evaluation of the genetic architecture of tofu traits in soybean towards genomics-assisted breeding. *Plant Breeding*, *137*(6), 873– 882. https://doi.org/10.1111/pbr.12651

Langewisch, T., Lenis, J., Jiang, G. L., Wang, D., Pantalone, V., & Bilyeu, K. (2017). The development and use of a molecular model for soybean maturity groups. *BMC Plant Biology*. https://doi.org/10.1186/s12870-017-1040-4

Langewisch, T., Zhang, H., Vincent, R., Joshi, T., Xu, D., & Bilyeu, K. (2014). Major soybean maturity gene haplotypes revealed by SNPViz analysis of 72 sequenced soybean genomes. *PLoS ONE*, *9*(4), e94150. https://doi.org/10.1371/journal.pone.0094150

Lee, G. A., Crawford, G. W., Liu, L., Sasaki, Y., & Chen, X. (2011). Archaeological soybean (Glycine max) in East Asia: Does size matter? *PLoS ONE*. https://doi.org/10.1371/journal.pone.0026720

Lee, S. H., Bailey, M. A., Mian, M. A. R., Shipe, E. R., Ashley, D. A., Parrott, W. A., … Boerma,

H. R. (1996). Identification of quantitative trait loci for plant height, lodging, and maturity in a soybean population segregating for growth habit. *Theoretical and Applied Genetics*, *92*(5), 516–523. https://doi.org/10.1007/BF00224553

Li, J., Wang, X., Song, W., Huang, X., Zhou, J., Zeng, H., … Qiu, L. (2017). Genetic variation of maturity groups & four E genes in the Chinese soybean mini core Collection. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0172106

Lieberei, R., & Reisdorff, C. (2012). *Nutzpflanzen*. Georg Thieme Verlag.

Lima, A., Oliveira, J., Saúde, F., Mota, J., & Ferreira, R. (2017). Proteins in Soy Might Have a Higher Role in Cancer Prevention than Previously Expected: Soybean Protein Fractions Are More Effective MMP-9 Inhibitors Than Non-Protein Fractions, Even in Cooked Seeds. *Nutrients*, *9*(3), 201. https://doi.org/10.3390/nu9030201

Liu, B., Watanabe, S., Uchiyama, T., Kong, F., Kanazawa, A., Xia, Z., … Abe, J. (2010). The Soybean Stem Growth Habit Gene Dt1 Is an Ortholog of Arabidopsis TERMINAL FLOWER1. *PLANT PHYSIOLOGY*. https://doi.org/10.1104/pp.109.150607

Liu, X., Wu, J., Ren, H., Qi, Y., Li, C., Cao, J., … Gai, J. (2017). Genetic variation of world soybean maturity date and geographic distribution of maturity groups. *Breeding Science*. https://doi.org/10.1270/jsbbs.16167

Lu, S., Zhao, X., Hu, Y., Liu, S., Nan, H., Li, X., … Kong, F. (2017). Natural variation at the soybean J locus improves adaptation to the tropics and enhances yield. *Nature Genetics*. https://doi.org/10.1038/ng.3819

Mammadov, J., Aggarwal, R., Buyyarapu, R., & Kumpatla, S. (2012). SNP markers and their impact on plant breeding. *International Journal of Plant Genomics*, *2012*, 728398. https://doi.org/10.1155/2012/728398

Matsumura, H., Liu, B., Abe, J., & Takahashi, R. (2008). AFLP mapping of soybean maturity gene E4. *Journal of Heredity*, *99*(2), 193–197. https://doi.org/10.1093/jhered/esm114

McBlain, B. A., & Bernard, R. L. (1987). A new gene affecting the time of flowering and maturity in soybeans. *The Journal of Heredity*, *78*, 160–162.

Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. (2001). Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics*, *157*, 1819–1829. Retrieved from http://www.genetics.org/content/genetics/157/4/1819.full.pdf

Michel, S., Kummer, C., Gallee, M., Hellinger, J., Ametz, C., Akgöl, B., … Buerstmayr, H. (2018). Improving the baking quality of bread wheat by genomic selection in early generations. *Theoretical and Applied Genetics*, *131*(2), 477–493. https://doi.org/10.1007/s00122-017-2998-x

Michel, S., Löschenberger, F., Ametz, C., Pachler, B., Sparry, E., & Bürstmayr, H. (2019a). Combining grain yield, protein content and protein quality by multi-trait genomic selection in bread wheat. *Theoretical and Applied Genetics*. https://doi.org/10.1007/s00122-019-03386-1

Michel, S., Löschenberger, F., Ametz, C., Pachler, B., Sparry, E., & Bürstmayr, H. (2019b). Simultaneous selection for grain yield and protein content in genomics-assisted wheat breeding. *Theoretical and Applied Genetics*, *132*(6), 1745–1760. https://doi.org/10.1007/s00122-019-03312-5

Miladinović, J., Ćeran, M., Đorđević, V., Balešević-Tubić, S., Petrović, K., Đukić, V., & Miladinović, D. (2018). Allelic Variation and Distribution of the Major Maturity Genes in Different Soybean Collections. *Frontiers in Plant Science*. https://doi.org/10.3389/fpls.2018.01286

Moehring, J., Williams, E., & Piepho, H. (n.d.). Efficiency of augmented p-rep designs in multi-environmental trials. *Theoretical and Applied Genetics*, *2014*. Retrieved from https://link.springer.com/article/10.1007/s00122-014-2278-y

Monaghan, J. M., Snape, J. W., Chojecki, A. J. S., & Kettlewell, P. S. (2001). The use of grain protein deviation for identifying wheat cultivars with high grain protein concentration and yield. *Euphytica*, *122*(2), 309–317. https://doi.org/10.1023/A:1012961703208

Moore, J. K., Manmathan, H. K., Anderson, V. A., Poland, J. A., Morris, C. F., & Haley, S. D. (2017). Improving Genomic Prediction for Pre-Harvest Sprouting Tolerance in Wheat by

Weighting Large-Effect Quantitative Trait Loci. *Crop Science*, *57*(3), 1315-1324. https://doi.org/https://doi.org/10.2135/cropsci2016.06.0453

Pan, Z., & Tangratanavalee, W. (2003). Characteristics of soybeans as affected by soaking conditions. *LWT - Food Science and Technology*, *36*(1), 143–151. https://doi.org/10.1016/S0023-6438(02)00202-5

Pimentel, D., Hepperly, P., Hanson, J., Douds, D., & Seidel, R. (2005). Environmental, Energetic, and Economic Comparisons of Organic and Conventional Farming Systems. *BioScience*, *55*(7), 573–582.

Poland, J. A., & Rife, T. W. (2012). Genotyping-by-Sequencing for Plant Breeding and Genetics. *The Plant Genome Journal*, *5*(3), 92. https://doi.org/10.3835/plantgenome2012.05.0005

Poysa, V., Woodrow, L., & Yu, K. (2006). Effect of soy protein subunit composition on tofu quality. *Food Research International*, *39*(3), 309–317. https://doi.org/10.1016/j.foodres.2005.08.003

Rapp, M., Lein, V., Lacoudre, F., Lafferty, J., Müller, E., Vida, G., … Longin, C. F. H. (2018). Simultaneous improvement of grain yield and protein content in durum wheat by different phenotypic indices and genomic selection. *Theoretical and Applied Genetics*, *131*(6), 1315–1329. https://doi.org/10.1007/s00122-018-3080-z

Ray, J. D., Hinson, K., Mankono, J. E. B., & Malo, M. F. (1995). Genetic Control of a Long-Juvenile Trait in Soybean. *Crop Science*, *35*(4), 1001. https://doi.org/10.2135/cropsci1995.0011183X003500040012x

Reckling, M., Hecker, J.-M., Bergkvist, G., Watson, C. A., Zander, P., Schläfke, N., … Bachinger, J. (2016). A cropping system assessment framework—Evaluating effects of introducing legumes into crop rotations. *European Journal of Agronomy*, *76*, 186–197. https://doi.org/10.1016/J.EJA.2015.11.005

Reid, J. A. F., & Cober, E. R. (2018). A Small-Scale Tofu Test for Soybean Breeding Programs. *Canadian Journal of Plant Science*. Retrieved from www.nrcresearchpress.com

Rowntree, S. C., Suhre, J. J., Weidenbenner, N. H., Wilson, E. W., Davis, V. M., Naeve, S. L., … Conley, S. P. (2013). Genetic gain × management interactions in soybean: I. Planting date. *Crop Science*, *53*(3), 1128–1138. https://doi.org/10.2135/cropsci2012.03.0157

Samanfar, B., Molnar, S. J., Charette, M., Schoenrock, A., Dehne, F., Golshani, A., … Cober, E. R. (2017). Mapping and identification of a potential candidate gene for a novel maturity locus, E10, in soybean. *Theoretical and Applied Genetics*. https://doi.org/10.1007/s00122-016-2819-7

Sato, T., Van Schoote, M., Wagentristl, H., & Vollmann, J. (2014). Effects of divergent selection for seed protein content in high-protein vs. food-grade populations of early maturity soybean. *Plant Breeding*, *133*(1), 74–79. https://doi.org/10.1111/pbr.12138

Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., … Jackson, S. A. (2010). Genome sequence of the palaeopolyploid soybean. *Nature*, *463*(7278), 178–183. https://doi.org/10.1038/nature08670

Scott, S. R., & Aldrich, W. O. (1983). *Modern Soybean Production* (2nd ed.). Champaign, Ill.: S & A Publications.

Sedivy, E. J., Wu, F., & Hanzawa, Y. (2017). Soybean domestication: the origin, genetic architecture and molecular bases. *New Phytologist*. https://doi.org/10.1111/nph.14418

Shu, Y. J., Yu, D. S., Wang, D., Bai, X., Zhu, Y. M., & Guo, C. H. (2013). Genomic selection of seed weight based on low-density SCAR markers in soybean. *Genetics and Molecular Research*, *12*(3), 2178–2188. https://doi.org/10.4238/2013.July.3.2

Shurtleff, W., & Aoyagi, A. (2015). *History of Soybeans and Soyfoods in France (1665-2015)*. Lafayette, CA: Soyinfo Center.

Sobko, O., Stahl, A., Hahn, V., Zikeli, S., Claupein, W., & Gruber, S. (2020). Environmental Effects on Soybean (Glycine Max (L.) Merr) Production in Central and South Germany. *Agronomy*, *10*(12), 1847. https://doi.org/10.3390/agronomy10121847

Specht, J. E., Hume, D. J., & Kumudini, S. V. (1999). Soybean yield potential - A genetic and

physiological perspective. In *Crop Science*. https://doi.org/10.2135/cropsci1999.3961560x

Tardivel, A., Sonah, H., Belzile, F., & O'Donoughue, L. S. (2014). Rapid Identification of Alleles at the Soybean Maturity Gene E3 using genotyping by Sequencing and a Haplotype-Based Approach. *The Plant Genome*, *7*(2), 0. https://doi.org/10.3835/plantgenome2013.10.0034

Technow, F., Bürger, A., & Melchinger, A. E. (2013). Genomic prediction of northern corn leaf blight resistance in maize with combined or separated training sets for heterotic groups. *G3: Genes, Genomes, Genetics*, *3*(2), 197–203. https://doi.org/10.1534/g3.112.004630

Tenaillon, M. I., & Charcosset, A. (2011). A European perspective on maize history. *Comptes Rendus - Biologies*. https://doi.org/10.1016/j.crvi.2010.12.015

Tian, Z., Wang, X., Lee, R., Li, Y., Specht, J. E., Nelson, R. L., … Ma, J. (2010). Artificial selection for determinate growth habit in soybean. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1000088107

Tsubokura, Y., Watanabe, S., Xia, Z., Kanamori, H., Yamagata, H., Kaga, A., … Harada, K. (2014). Natural variation in the genes responsible for maturity loci E1, E2, E3 and E4 in soybean. *Ann. Botany*, *113*(3), 429–441. https://doi.org/10.1093/aob/mct269

Varshney, R. K., Terauchi, R., & McCouch, S. R. (2014). Harvesting the Promising Fruits of Genomics: Applying Genome Sequencing Technologies to Crop Breeding. *PLoS Biology*, *12*(6), e1001883. https://doi.org/10.1371/journal.pbio.1001883

Varzakas, T. H., Arvanitoyannis, I. S., & Baltas, H. (2007). The politics and science behind GMO acceptance. *Critical Reviews in Food Science and Nutrition*. https://doi.org/10.1080/10408390600762696

Voss-Fels, K. P., Cooper, M., & Hayes, B. J. (2019). Accelerating crop genetic gains with genomic selection. *Theoretical and Applied Genetics*, *132*(3), 669–686. https://doi.org/10.1007/s00122-018-3270-8

Watanabe, S., Hideshima, R., Zhengjun, X., Tsubokura, Y., Sato, S., Nakamoto, Y., … Harada, K. (2009). Map-based cloning of the gene associated with the soybean maturity locus E3. *Genetics*, *182*(4), 1251–1262. https://doi.org/10.1534/genetics.108.098772

Watanabe, S., Xia, Z., Hideshima, R., Tsubokura, Y., Sato, S., Yamanaka, N., … Harada, K. (2011). A map-based cloning strategy employing a residual heterozygous line reveals that the GIGANTEA gene is involved in soybean maturity and flowering. *Genetics*, *188*(2), 395–407. https://doi.org/10.1534/genetics.110.125062

Wilbois, K.-P., Spiegel, A.-K., Balko, C., Becker, H., Berset, E., Butz, A., … Zurheide, T. (2014). *Expansion of soybean cultivation in Germany through adaptation by breeding as well as optimization of crop production and processing technology*. Frankfurt am Main. Retrieved from www.orgprints.org/28484

Wilson, E. W., Rowntree, S. C., Suhre, J. J., Weidenbenner, N. H., Conley, S. P., Davis, V. M., … Casteel, S. N. (2014). Genetic gain × management interactions in soybean: II. nitrogen utilization. *Crop Science*. https://doi.org/10.2135/cropsci2013.05.0339

Xia, Z., Zhai, H., Liu, B., Kong, F., Yuan, X., Wu, H., … Harada, K. (2012). Molecular identification of genes controlling flowering time, maturity, and photoperiod response in soybean. *Plant Systematics and Evolution*, *298*(7), 1217–1227. https://doi.org/10.1007/s00606-012-0628-2

Xu, M., Xu, Z., Liu, B., Kong, F., Tsubokura, Y., Watanabe, S., … Abe, J. (2013). Genetic variation in four maturity genes affects photoperiod insensitivity and PHYA-regulated post-flowering responses of soybean. *BMC Plant Biology*, *13*(1), 91. https://doi.org/10.1186/1471-2229-13-91

Xu, Y. (2016). Envirotyping for deciphering environmental impacts on crop plants. *Theoretical and Applied Genetics*, *129*(4), 653–673. https://doi.org/10.1007/s00122-016-2691-5

Xu, Y., & Crouch, J. H. (2008). Marker-assisted selection in plant breeding: From publications to practice. *Crop Science*. https://doi.org/10.2135/cropsci2007.04.0191

Yuesheng, W., Jianbing, Q., Junyi, G., & Guangyuang, H. (2006). Classification and Characteristic of Maturity Groups of Chinese Landraces of Soybean [Glycine max (L.) Merr.]. *Genetic*

Resources and Crop Evolution, *53*(4), 803–809. https://doi.org/10.1007/s10722-004-5731-y

Zander, P., Amjath-Babu, T. S., Preissel, S., Reckling, M., Bues, A., Schläfke, N., … Watson, C. (2016). Grain legume decline and potential recovery in European agriculture: a review. *Agronomy for Sustainable Development*, *36*(2), 26. https://doi.org/10.1007/s13593-016-0365-y

Zhai, H., Lü, S., Wang, Y., Chen, X., Ren, H., Yang, J., … Xia, Z. (2014). Allelic variations at four major maturity e genes and transcriptional abundance of the E1 gene are associated with flowering time and maturity of soybean cultivars. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0097636

Zhang, J., Song, Q., Cregan, P. B., & Jiang, G.-L. (2016). Genome-wide association study, genomic prediction and marker-assisted selection for seed weight in soybean (Glycine max). *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, *129*(1), 117–130. https://doi.org/10.1007/s00122-015-2614-x

Zhang, L. X., Kyei-Boahen, S., Zhang, J., Zhang, M. H., Freeland, T. B., Watson, C. E., & Liu, X. (2007). Modifications of Optimum Adaptation Zones for Soybean Maturity Groups in the USA. *Crop Management*, *6*, 1–10. https://doi.org/10.1094/CM-2007-0927-01-RS

Zhang, Q., Wang, C., Li, B., Li, L., Lin, D., Chen, H., … Yang, W. (2018). Research progress in tofu processing: From raw materials to processing conditions. *Critical Reviews in Food Science and Nutrition*. https://doi.org/10.1080/10408398.2016.1263823

Zhao, C., Takeshima, R., Zhu, J., Xu, M., Sato, M., Watanabe, S., … Abe, J. (2016). A recessive allele for delayed flowering at the soybean maturity locus E9 is a leaky allele of FT2a, a FLOWERING LOCUS T ortholog. *BMC Plant Biology*, *16*(1), 20. https://doi.org/10.1186/s12870-016-0704-9

Declaration in lieu of an oath on independent work

Declaration in lieu of an oath on independent work according to Sec. 18(3) sentence 5 of the University of Hohenheim's Doctoral Regulations for the Faculties of Agricultural Sciences, Natural Sciences, and Business, Economics and Social Sciences

1. The dissertation submitted on the topic

 "Phenotypic and Genomics-assisted Breeding of Soybean for Central Europe:From Environmental Adaptation to Tofu Traits"

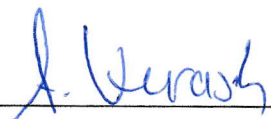is work done independently by me.

2. I only used the sources and aids listed and did not make use of any impermissible assistance from third parties. In particular, I marked all content taken word-for-word or paraphrased from other works.

3. I did not use the assistance of a commercial doctoral placement or advising agency.

4. I am aware of the importance of the declaration in lieu of oath and the criminal consequences of false or incomplete declarations in lieu of oath.

 I confirm that the declaration above is correct. I declare in lieu of oath that I have declared only the truth to the best of my knowledge and have not omitted anything.

Bodigheim, 10.4.23
Place, Date

Signature

Curriculum Vitae

Name:	Alena Katharina Kurasch
Date and place of birth:	9. March 1989 in Stuttgart
Marital Status:	Married, 3 children

Education

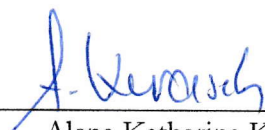|	
April 2015 - March 2018	Doctorate candidate in State Plant Breeding Institute (Prof. Dr. T. Würschum), University of Hohenheim, Germany
October 2012 - March 2015	M.Sc. studies and graduation, Agricultural Science, University of Hohenheim, Germany
October 2009 - September 2012	B. Sc. studies and graduation, Agricultural Biology, University of Hohenheim, Germany
July 2008	High school degree (Abitur), Eckenberggymnasium, Adelsheim, Germany

Professional Experience

since April 2018	Silage maize breeder, MAS Seeds Deutschland, Neckarmühlbach, Germany
April 2015 - March 2018	Research Assistant in State Plant Breeding Institute (Prof. Dr. T. Würschum), University of Hohenheim, Germany

10.4.23

Date,   Alena Katharina Kurasch
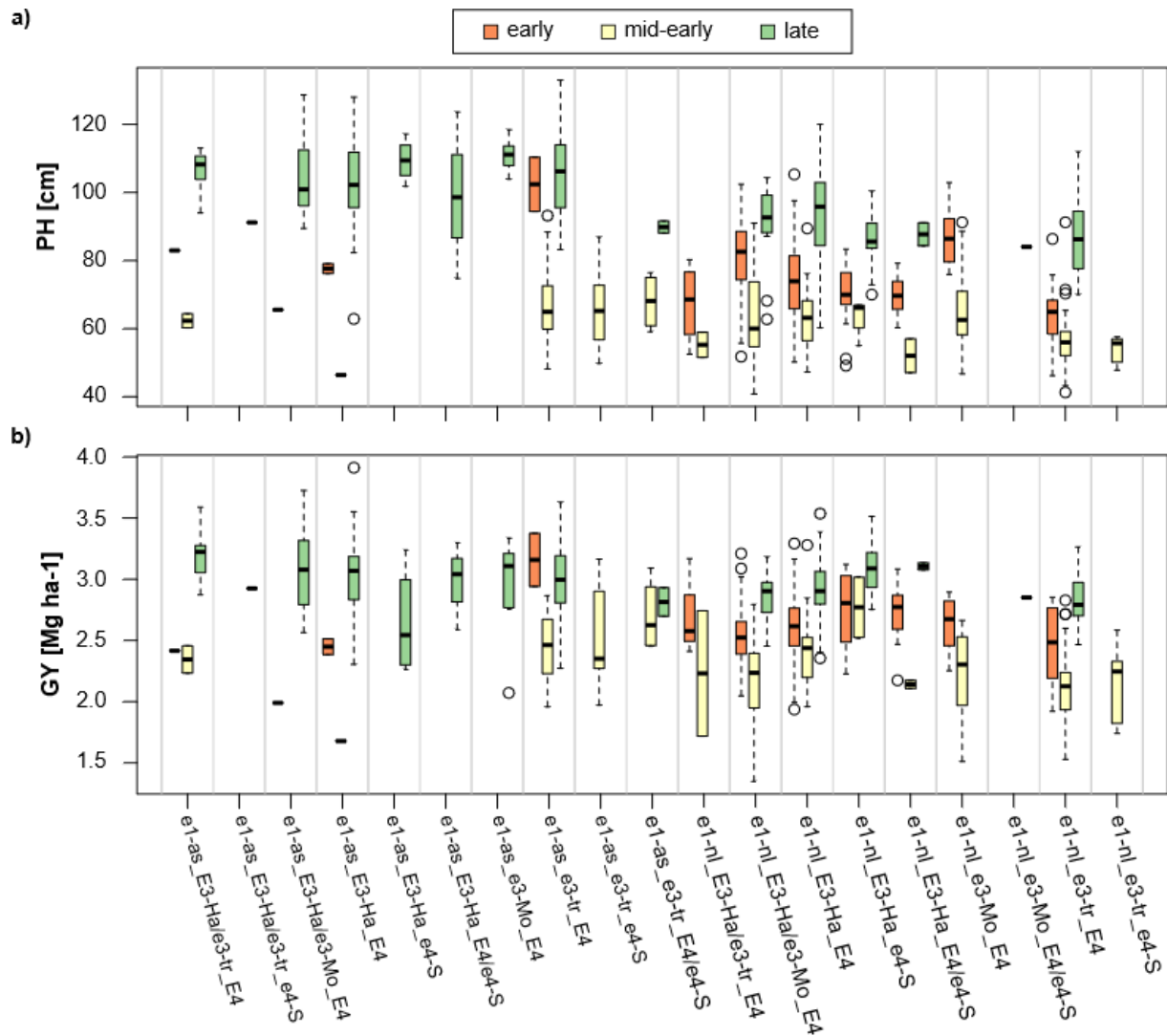
Appendix

Figure 4: Distribution of *E*-haplotypes depending on trial assignment from 2013 (Kurasch, Hahn, Leiser, Starck, et al., 2017) into early, mid-early and late for a) plant height and b) grain yield.