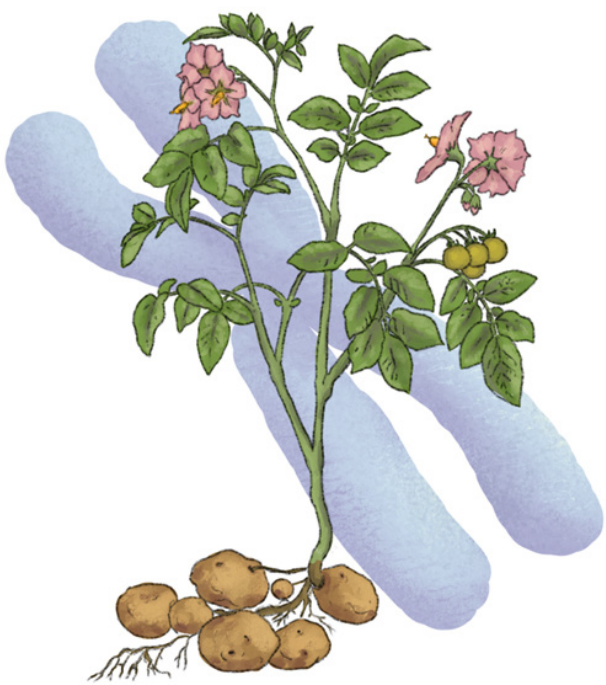
Acta Universitatis Agriculturae Sueciae

Doctoral Thesis No. 2021:80
Faculty of Landscape Architecture, Horticulture and Crop Production Science

Genomic-led potato breeding

Catja Selga



Genomic-led potato breeding

Catja Selga
Faculty of Landscape Architecture, Horticulture and Crop Production Science
Department of Plant Breeding
Alnarp

SLU
SWEDISH UNIVERSITY OF AGRICULTURAL SCIENCES

DOCTORAL THESIS
Alnarp 2021

Acta Universitatis agriculturae Sueciae
2021:80

Cover: A potato plant with one of its 48 chromosomes during the prophase of mitosis (Illustration by Louise Selga)

ISSN 1652-6880
ISBN (print version) 978-91-7760-837-0
ISBN (electronic version) 978-91-7760-838-7

Alnarp
Print: SLU Service/Repro, Alnarp 2021

Genomic-led potato breeding

Abstract

Most cultivated potatoes (*Solanum tuberosum* L.) are polyploids with tetrasomic inheritance ($2n = 4x = 48$), which makes breeding for new improved cultivars more complicated compared to diploid crops. In addition to this, the potato market is much more fragmented compared to other staple crops such as maize or wheat, where yield and dry matter are by far the most important traits. A conventional potato breeder needs to consider a number of important breeding traits, including plant maturity, tuber and cooking quality, host plant resistance to pests and pathogens, and starch content. In the scope of this thesis, genomic-led breeding approaches for tetraploid potato are explored using high throughput genotypic data acquired from single nucleotide polymorphism (SNP) markers. Quantitative trait loci (QTL) for host plant resistance to *Alternaria solani* are mapped to chromosomes 5 and 11 using a segregating bi-parental crossing population. The application of genomic selection (GS) is tested in a potato breeding program situated in Sweden by predicting genomic estimated breeding values (GEBVs) for eight important breeding traits, such as tuber yield and quality, and host plant resistance to *Phytophthora infestans.* The predictive ability of GEBVs across clonal generations in the breeding program is poor for most breeding traits, however, the approach of predicting GEBVs within and across half-sib families results in higher predictive ability. Lastly, the genetic diversity of the potato germplasm at the Nordic Genetic Resource Centre (NordGen) is compared to newly bred clones in Sweden. These results suggest a close genetic relationship between the accessions from NordGen (local farmer's cultivars and obsolete cultivars), and the modern breeding clones from the potato breeding program based in Sweden.

Keywords: GEBV, genomic selection, linkage analysis, population structure potato breeding, QTL, tetraploid

Author's, address: Catja Selga, Swedish University of Agricultural Sciences, Department of Plant Breeding, Alnarp, Sweden

Genomic-led potato breeding

Abstract

Den mesta av den odlade potatisen (*Solanum tuberosum* L.) är polyploid med tetrasomisk nedärvning ($2n = 4x = 48$), vilket bidrar till att förädling av nya förbättrade sorter är mer komplicerad än för diploida grödor. Förutom detta så är marknaden för potatis mycket mer fragmenterad i jämförelse med andra basgrödor så som majs och vete, där avkastning och torrvikt är de absolut viktigaste egenskaperna. En konventionell potatisförädlare måste ta många egenskaper i beaktande, så som mognadstid, knöl- och kokkvalitéer, värdplantans resistens mot skadedjur och sjukdomar, och stärkelseinnehåll. I denna avhandling utforskas genomiskt baserade förädlingsmetoder för tetraploid potatis genom användandet av högkapacitets genotypdata från enbaspolymorfi (SNP) markörer. Kvantitativ egenskaps loci (QTL) för plantans resistens mot *Alternaria solani* kartläggs till kromosomerna 5 och 11 genom användandet av en segregerande förädlingspopulation. Tillämpandet av genomisk selektion (GS) testas i ett svenskt potatisförädlingsprogram genom att beräkna genomiskt estimerade förädlingsvärden (GEFVs) för åtta viktiga förädlingsegenskaper, till exempel knölskörd och knölkvalité, och värdplantans resistens mot *Phytophthora infestans.* Graden av korrekt beräknade GEFVs över selektionsgenerationer i förädlingsprogrammet är låg för de flesta förädlingsegenskaper, dock verkar det bli högre korrelation när GEFVs beräknas inom och över familjer av halvsyskon. Slutligen undersöks och jämförs den genetiska diversiteten av potatisklonerna hos det Nordiska Genresurscentret (NordGen) med nya förädlingskloner i Sverige. Dessa resultat antyder att det genetiska avståndet mellan klonerna från genbanken och de nya förädlingsklonerna från det svenska förädlingsprogrammet är mycket litet.

Nyckelord: tetraploid, potatisförädling, QTL, GEFV, populationsstruktur, kopplingsanalys, genomisk selektion

Author’s address: Catja Selga, Swedish University of Agricultural Sciences, Department of Plant Breeding, Alnarp, Sweden

Dedication

To my sisters and to my sunshine, Julian

"[…]enemy of hunger,
in all
nations
you've planted
your victorious and ready
banner,
in frozen land or in the ground
of burning coastlines
your anonymous flower
has appeared,
announcing the thick
and steady
birth rate of your roots."

Pablo Neruda, Oda a la papa

Contents

List of publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I. Odilbekov, F.*, Selga, C., Ortiz, R., Chawade, A., Liljeroth, E. (2020). QTL mapping for resistance to early blight in a tetraploid potato population. *Agronomy*, 10 (5), 728.

II. Selga, C.*, Reslow, F., Pérez-Rodríguez, P., Ortiz, R. (in press). The power of genomic estimated breeding values for selection when using a finite population size in genetic improvement of tetraploid potato. *G3 Genes|Genomes|Genetics*

III. Selga, C., Chrominski P., Carlson-Nilsson, U., Andersson, M., Chawade, A., Ortiz, R.* Diversity and population structure of Nordic potato cultivars and breeding clones grown under long daylength in Europe (manuscript)

Papers I and II are reproduced with the permission of the publisher.

*Corresponding author

The contribution of Catja Selga to the papers included in this thesis was as follows:

I. Performed genetic analysis, contributed to final version of the manuscript.

II. Designed study together with co-authors, performed data collection and data analysis. Wrote final version of the manuscript with input of co-authors.

III. Designed study together with co-authors, contributed to data collection. Performed data analysis. Wrote final version of the manuscript with input of co-authors.

List of tables

List of figures

Illustrations created by Catja Selga unless specifially mentioned otherwise.

Abbreviations

AFLP	Amplified fragment length polymorphism
ATW	Average weight per tuber
AUDPC	Area under disease progress curve
BRR	Bayesian Ridge Regression
CIP	International Potato Centre
DArT	Diversity array technology
DNA	Deoxyribonucleic acid
EB	Early blight
GBS	Genotyping by sequencing
GEBV	Genomic estimated breeding value
GS	Genomic selection
GWAS	Genome-wide association study
LB	Late blight
LOD	Logarithm of odds
MAS	Marker-assisted selection
NGS	Next generation sequencing
NordGen	Nordic Genetic Resource Centre
PC	Principal coordinate
PCA	Principal component analysis

PCoA	Principal coordinate analysis
PCR	Polymerase chain reaction
QTL	Quantitative trait loci
RAPD	Random amplified polymorphic DNA
RFLP	Restriction fragment length polymorphism
SG	Specific gravity
Sli	S-locus inhibitor
SLU	Swedish University of Agricultural
SNP	Single nucleotide polymorphism
SSR	Simple sequence repeat
T_i	Clonal generation of selection, in the i:th cycle of selection
TN	Per plant tuber number
TPS	True potato seed
TW	Per plant tuber weight

1. Introduction

The world food shortage is increasing and this negative trend has escalated during the COVID-19 pandemic. Between 720 and 811 million people faced hunger in 2020, which is an increase of as many as 161 million people just from the year before (FAO *et al*., 2021). Ending world hunger and improving nutrition for all people requires transformation of food systems on a global level, and an important strategy includes improving agricultural technology and delivering related innovations. Increasing crop yield and crop adaptation to changing growth conditions is one of several means to ensure food security. After wheat and rice, potato is the most important crop in human diets worldwide, and is grown in all continents except Antarctica (FAOSTAT, 2021). Plant breeders have an important role to play in ensuring a sustainable potato production that meets consumer demands, by developing new cultivars with increased host plant resistance to pathogens and increased yields per hectare (Birch *et al.* 2012).

Producing enough food to ward off hunger has been an issue for humanity for the last 12 000 years. Today, most fertile land is already cultivated, and hence crop production will have to be improved by increasing crop yields instead of expansion of new agricultural lands (Borlaug, 2002). Since the agricultural revolution, deliberately or not, plant breeding has been practiced by farmers through the maintenance of individual plants with favourable traits for cultivation in the next season. After the rediscovery of Mendel's research on inheritance and the establishment of the understanding of quantitative genetics in the beginning of the 20th century, the efficiency of plant breeding has increased significantly (Bernardo, 2020). During the second part of the 20th century, the launching of public sector research programs and international agricultural research centres, led to the

development of many modern cultivars of various crops. These new cultivars were a huge part of the success of the so-called Green Revolution that reduced the hunger in Asia and Latin America starting in the 1960s (Qaim, 2020). One of the successful breeding methods of the Green Revolution-era was the introduction of dwarfing genes and other desirable traits (e.g. daylength insensitivity or host plant resistance to target pests) by access to large genetic stocks, particularly in cereals (Evenson and Gollin, 2003).

From the 1990s and onwards, genetics have come to play an increasing important role in plant breeding by the introduction of DNA markers (Gupta and Varshney, 2000) as indirect selection aids or to study genetic diversity. Today, genomics-based plant breeding techniques such as marker-assisted selection (MAS) and genomic selection (GS) are speeding up the genetic gain for many of our most important food crops. Compared to the advances in genetic gain in cereal crops, advances in genetic gain for potato is lagging behind (Slater *et al.*, 2016). The focus of this thesis is to explore how genomics-based methods may aid potato breeding.

2. Background

2.1 *Solanum tuberosum*

2.1.1 Origin and domestication

The cultivated potato (**Figure 1**), *Solanum tuberosum* (L.), originates from the Andes in South America. The potato was domesticated in Peru around 8 000–10 000 years ago from wild relatives, and early selection characters were large tubers, short stolons, low content of tuber glycoalkaloids based on a bitter taste, tuber shape and colour. The origin of the cultivated potato was most likely a diploid member of the *Solanum brevicaule* group (Spooner *et al.* 2005).

The potato can reproduce either sexually, through pollination of its flowers and the production of berries which set true seeds, or asexually through tuber propagation. The plants originating from tuber propagation are clones of the mother plant, while each seedling from a true seed is genetically unique. Both these modes of reproduction are practiced by potato breeders today, and were probably used in early domestication events as well (Bradshaw and Ramsay, 2009). The domestication of the potato probably proceeded quite slowly, as the crops earliest farmers were primarily herders of llama and alpaca. The Andean people most likely moved up and down the mountains with the changing of seasons, and tuberous plants such as the potato were planted at favoured sites and then harvested upon subsequent re-visits (Bradshaw and Ramsay, 2009). The raw potato tuber is indigestible, but the Andean people had several ways to prepare them. Just like today, boiling and roasting were

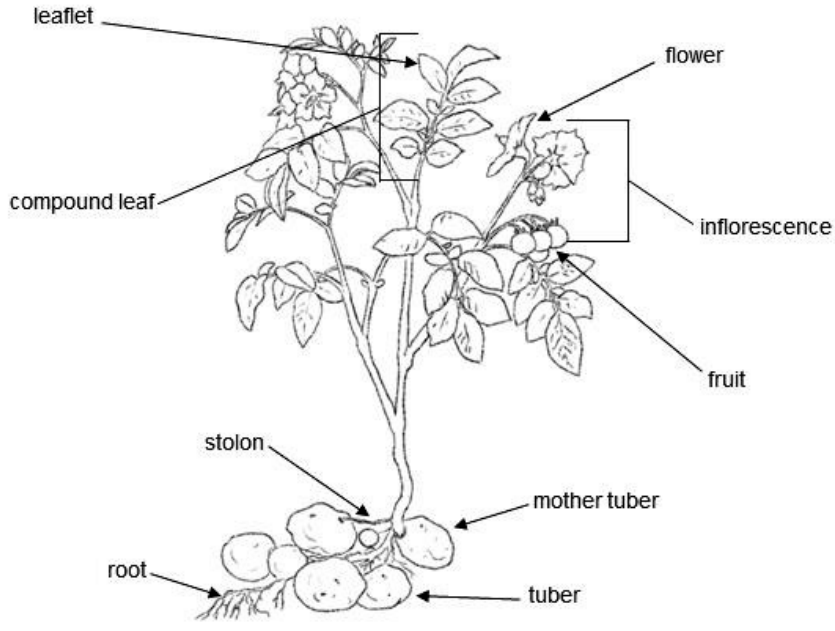


Figure 1. Schematic figure of a cultivated potato plant *Solanum tuberosum* (L.). Illustration: L. Selga

common practices. The preparation of chuño, an ancient method of freeze-drying potatoes, allowed the Andean people to store the potato for later consumption over an extended period. For ancient and modern people alike, potato tubers were an important source of nutrients. The tubers are rich in carbohydrates and provide important basic nutrients including, dietary fibre, vitamins and minerals (Zaheer and Akhtar, 2016).

In the 16th century, as the Spanish invaded South America, the potato was a widespread crop in the Andean region (Bradshaw and Ramsey, 2009). The Spanish brought the potato back with them to Europe, and the first recorded cultivation was on the Canary Islands in the middle of the 16th century (Hawkes and Francisco-Ortega, 1993). From early on, several different potato landraces were introduced to Europe, but eventually, landraces from Chile became to be particularly favoured as they were already adapted to long-day growth conditions as those found in European agriculture (Ames and Spooner, 2008). At its introduction to Europe, potato was grown as a medical plant and botanical curiosity (Burton, 1989). It was not until the end of the 17th century that the potato transitioned to being a major food crop,

when it was beginning to be cultivated in large scale on Ireland. The Irish people subsequently became reliant on the potato as their staple food, until the late blight (LB) epidemic in 1845, caused by the oomycete *Phytophthora infestans* (Bradshaw and Ramsay, 2009; Yoshida *et al.* 2013). In many parts of Europe, the cultivation of potato was introduced by military decree to evade famine during the 18th century (Burton, 1989; Bradshaw and Ramsay, 2009).

The cultivated potato was introduced from Europe to North America and some parts of western Asia early in the 17th century. Later, the potato crop has also been introduced to other parts of Asia – including China and India – Africa and Oceania (Bradshaw and Ramsay, 2009).

2.1.2 Tetraploid genetics

There are several levels of ploidy naturally occurring in wild potato species, but Andean farmers have (unintentionally) favoured tetraploid types. Presumably, the tetraploid types were favoured over diploid types due to them being superior in tuber productivity and other traits (Bradshaw and Ramsay, 2009). Wild tetraploid and hexaploid species often follow a diploid inheritance pattern. The cultivated tetraploid potato, however, exhibits a tetrasomic pattern of inheritance ($2n = 4x = 48$). A tetrasomic pattern of inheritance entails that all four sets of chromosomes are entirely homologous so that meiotic pairing within each group of the four homologues occurs completely at random (**Figure 2**). There are three ways in which tetrasomic inheritance differs from disomic inheritance; (1) double reduction, (2) the gametes are diploid ($n = 2x$), and (3) the deviation from the Hardy-Weinberg equilibrium following random mating. Double reduction is a phenomenon where chromosomes may associate in a quadrivalent instead of a bivalent. As a result of this, at the end of meiosis, sister chromatids may end up in the same gamete. The gametes being diploid allow for dominance and inbreeding effects to be passed on from parent to offspring. The deviation from Hardy-Weinberg is caused by the fact that the single locus equilibrium of genotype frequency is not achieved by one generation of random mating, and random mating will not remove the effects of inbreeding in a single generation (Bradshaw, 2007; Bradshaw, 2021).

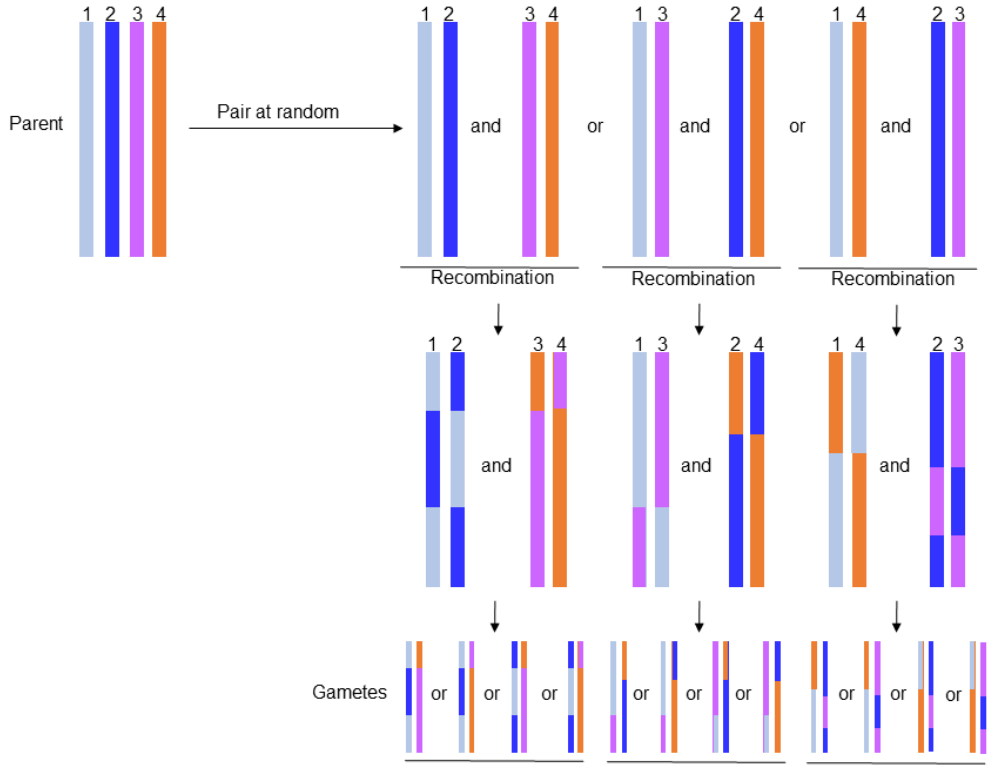


Figure 2. Chromosome pairing of four homologous chromosomes and reduction to gametes during meiosis in a tetraploid species with tetrasomic inheritance, assuming chromosomal segregation.

In addition to this, performing genetic analyses in an organism with tetrasomic inheritance means that there is less available software to use. Most software is developed with diploid organisms in mind, and the additional allelic states required is often an issue. It is also important to keep in mind the possible dominance and additive effects of the alleles at each individual locus. Researchers have sometimes solved this problem by "diplodizing" their data; i.e., reducing the format of five possible allelic states at each locus (AAAA, AAAO, AAOO, AOOO and OOOO) to three allelic states (AA, AO and OO). The practice of diploidizing data may, however, result in loss of vital information (Piepho and Koch, 2000).

For potato breeders, tetrasomic inheritance leads to favourable alleles often not being fixed in parental lines (homozygous), and a tremendous amount of segregation is observed in the offspring. This limits the genetic gain, as only a small number of the offspring will outperform the parents for all breeding traits (Vos, 2016).

2.1.3 Commercial production

Today, potatoes are grown in over 100 countries worldwide and in numerous environmental conditions including temperate, subtropical and tropical zones (**Figure 3A**) (Hijmans, 2001). China is the world's leading potato producing country (**Figure 3B**) accounting for about a quarter of the world total potatoes in 2019 (FAOSTAT, 2021). In Europe, the top producing country is Ukraine, producing on average 21.8 million tonnes per year (2009–2019). In contrast, Sweden produces on average 0.8 million tonnes per year (FAOSTAT, 2021).

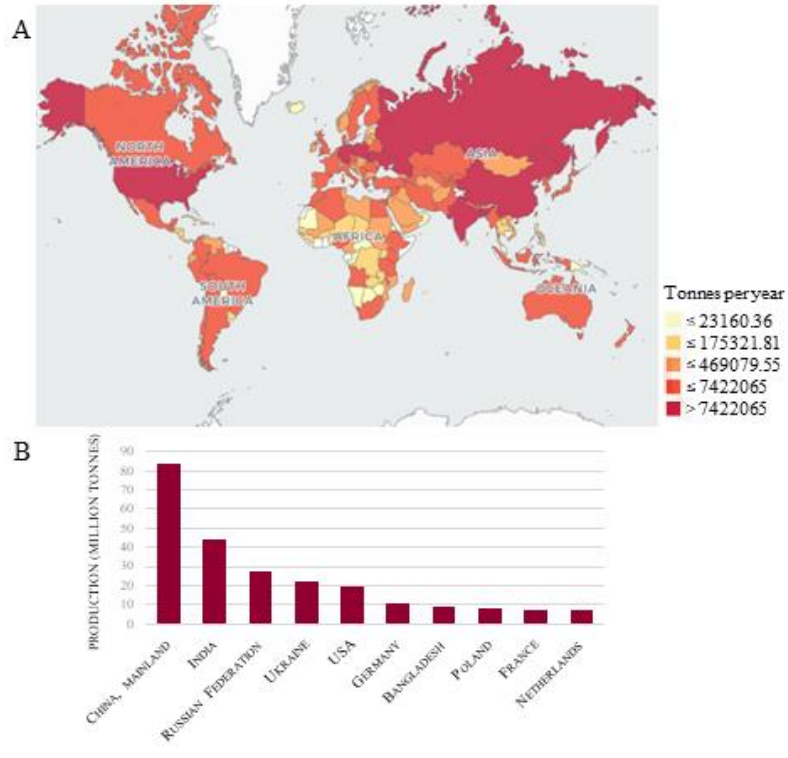


Figure 3. Potato production (2009–2019): average tonnes per yearworldwide (**A**) and top 10 producing countries (**B**). Source: FAOSTAT, 2021.

In lowland temperate regions such as Sweden, potato is grown as a summer crop. In some areas of mainland China, parts of equatorial South America, and East Africa, potatoes are grown around the year, while in other part of the world it is grown as a winter, autumn or spring crop. Some countries, such as Japan, practises a double cropping system in spring and autumn (Bradshaw, 2021). The length of the cropping season, as well as temperature and day-length, determines the potential yield of the potato plant. To accommodate the various growing conditions worldwide, different maturity types of potato are used, from late cultivars that need to utilise large quantity of light to early cultivars that display more efficient use of resources (Haverkort and Struik, 2015). The average cropping season for potato is 90–120 days long, but may be as short as 75 days in lowland subtropics and up to 180 days in the high Andes (Bradshaw, 2021). In Sweden, which is a relatively long country going from 69°N to 55°N, the cropping season may be up to 120 days in the south, and 90 days in the North (Ortiz *et al.* submitted).

The main component of the dry matter in potato tubers is starch, which provide a popular source of carbohydrates around the world. However, the tubers are also quite rich in proteins, with a good composition of amino acids and contain many vitamins and minerals. Potatoes are produced for a variety of different markets, for example, food products such as crisps, chips, canned potato, or industrial products such as starch production towards paper, textile and biodegradable plastics (Bradshaw and Ramsay, 2009).

2.2 Potato breeding

The potato market differs in a number of ways from the market of other major crops such as wheat, rice, maize and sugar beet. For major staple crops, the yield per hectare is the most important trait. Several additional traits – mainly tuber quality traits – are of importance in potato. In addition to this, there are several different markets segments for potato, such as the starch industry, crisp industry, chips industry and fresh potato consumption. All this adds up to that in a potato breeding program, many more traits have to be considered compared to the other major crops.

2.2.1 Historic breeding methods

During the 18th century, named cultivars from South America were introduced to Europe and North America (Bradshaw, 2021). However, directed potato breeding, through hybridization between cultivars, is first reported at the beginning of the 19th century (Knight, 1807). Up until then, the crop was improved through pollination by insects, and the subsequent plantings of the genetically unique seeds which arose from these natural crossings. The crosses were then maintained through clonal propagation of tubers, and the cultivars could be maintained. During the second part of the 19th century, the first successful cultivars from directed crosses were released from England. One of the first cultivars to be released was 'Magnum Bonum' (Bradshaw, 2021).

Genetic hybridization – the practise of breeding individuals from genetically distinct populations – was applied as a breeding method during the 19th century, and germplasm exchanges between North America and Europe occurred. Seedlings from berries derived from open pollination were commonly planted, and some very popular cultivars resulted from this practise, 'Burbank' (the source of the cultivar 'Russet Burbank') being one example (Ortiz, 2001).

With the rediscovery of Mendel's work on inheritance in the early 20th century, breeders could apply this knowledge when making crosses. Applications of this work proved difficult for potato though, as it was discovered to be tetraploid with tetrasomic inheritance during the 1930s and 1940s (Cadman, 1942; Bradshaw, 2021).

2.2.2 Current breeding methods

Since the beginning of the 20th century, many new breeding methods have been implemented by potato breeders. Today, most conventional breeding programs consist of a number of cycles of selections usually up to eight, which herein will be denoted as T_i where i is the i:th cycle of selection. It has been estimated that on average it takes 11 years from the initial cycle of selection to commercialisation of selected potato clones (Milbourne *et al.* 2007). The cycles start with the crossing of two parents, breeding clones or cultivars, often with complementing characteristics, and the subsequent production of seeds. The seeds are planted in greenhouses and small tubers,

sometimes referred to as tuberlings, are produced, one tuber per plant. These tubers, which makes up T_1 are planted and the plants and their subsequent tuber yield are scored for a number of traits. The harvested tubers are kept for planting next year, and hence the number of replicates for each breeding clone will be increased for each passing T_i, while the number of clones is reduced through the breeder's selection. In the 1960s, it was theorized that the number of seedlings in T_1 should be about 100,000, produced from 200 to 300 crosses (Bradshaw, 2009: Bradshaw, 2017). Though some breeders still operate their breeding programmes according to this idea, assessing 100,000 plants each year visually is time-intensive and has been deemed to be ineffective. To rationalize the selections in T_1, progeny testing may be carried out, to select among full-sib families in greenhouse conditions before the field assessments are to be carried out (Mackay, 2003; Bradshaw, 2017).

Wild potato species are sometimes used for introgression breeding of certain traits, as well as base broadening to increase the breeding germplasms' genetic diversity. In the 1980s, the use of haploids (individuals which have the same number of chromosome sets as a gamete) was proven to be an effective method of transferring genetic diversity and introduction of traits. The offspring of haploids may be tetraploid through sexual polyploidisation (Hermundstad and Peloquin, 1987; Ortiz *et al.* 2009). Breeding potatoes as a diploid, inbred-line based crop would make introgression from wild species more manageable (Jansky *et al.* 2016), and currently research on diploid potato breeding advanced greatly with the introduction of the dominant S-locus inhibitor (*Sli*) gene which leads to self-compatibility (Eggers *et al.* 2021). The transition of breeding potatoes as diploid, inbred lines has revived the interest in propagating plants from true potato seeds (TPS), which is the seed produced from a sexual cross, in contrast with vegetative tuber propagation (Jansky *et al.* 2016). Propagating plants from TPS was first researched as a mean of producing offspring free from tuber-borne diseases. Furthermore, the seeds were easier to store and handle, as TPS are much smaller than seed tubers and are less limited by physiological aging factors affecting tubers in storage. Despite these advantages, breeding for TPS cultivars has not become widespread due to the lack of offspring uniformity (Simmonds, 1997) and the difficulty to convince farmers to change their potato cropping systems (Chujoy and Cabello, 2007). However, with the current research on diploid potato breeding, TPS has gained interest again as

a means of maintaining the hybrid lines produced by diploid breeding (Jansky *et al.* 2016).

Just as for other crops, genetic modification and gene editing for crop improvement has been explored in potato, with several successful cases such as resistance to LB (Jo *et al.* 2014), increased resistance to potato viruses (Chung *et al.* 2013), altered glycoalkaloids content (Nakayasu *et al.* 2018) and modified starch properties (Andersson *et al.* 2017; Zhao *et al.* 2021). Other molecular based breeding methods, such as MAS and GS, are implemented in potato breeding to a limited extent, and will be covered in later sections of this thesis (2.5 and 2.6).

2.3 Potato breeding in Sweden

By 1658, potato was introduced to Sweden as a medical curiosity in the botanical garden in Uppsala, where it was given the name 'Peruvian nightshade' (translated from 'Peruviansk nattskatta') (Burton, 1989; Erjefält, 2001). The spread of the 'Peruvian nightshade' (or 'Tartuffel' as it was also known as) in Sweden was slow, up until agricultural pioneer Jonas Alströmer started advocating for the nutritional benefits of the crop in 1724. Alströmer adopted the English word for the crop, and called the plant 'Potatoes' (from which the Swedish word 'potatis' is derived) in publications used for convincing government authorities of the tubers' excellent properties. Despite the major efforts to popularize potato as a staple crop by the Swedish Academy of Science and the government, the introduction of cultivating potato for food was slow in Sweden during the 18th century. It was not until 1800 that potato cultivation became widespread among the Swedish people. The government handed out free seed to farmers, arranged harvest contests and promoted the benefits of replacing cereal with potato by printing a calendar (2nd most popular book in Sweden at the time) with information on how to use the tubers, one of which was the production of vodka (Erjefält, 2001).

Around the year 1900, the cultivation of the first named potato cultivars started in Sweden. Among the first cultivars to be grown were 'Early Rose', 'Magnum Bonum', 'Bintje' and 'King Edward VII'. Up until then, most potatoes were farmer's cultivars with descriptive names such as 'Light red

long', 'Blue potato' and 'Giant potato' (translated from 'Ljusröd lång', 'Blåpotatis' and 'Jättepotatis', respectively).

Targeted potato breeding began in 1903 by the Swedish Seed Association (Svalöf) situated in Svalöv. Their breeding targets where the same as those in place in the breeding program today: to produce high yielding cultivars with early maturity, consumption quality and resistance to pests and pathogens – especially host plant resistance to LB. In 1911, another company situated close to Svalöv in Landskrona, W. Weibulls began breeding potatoes. Up until the 1940s, the two potato breeding companies co-existed, until Weibulls terminated their Swedish potato breeding program. During the 1960s–1980s, the potato breeding enterprises at Svalöf were thriving. In the late 1950s, the breeding material was expanded with germplasm form Denmark, and from 1963 the company had 100 000 first year clones each year. During the 1970s, breeding material was further expanded with material from the International Potato Centre (CIP) in Peru. Many of these expansions were possible through major public and private investments. In 1993, the two companies Svalöf and Weibull merged, which included the potato breeding activities from Weibulls in the Netherlands. In 2006, the company decided to terminate the potato breeding program for the Swedish market, and the material was transferred to the Swedish University of Agricultural Sciences (SLU) in Alnarp (Erjefält, 2001; Eriksson et al. 2016).

2.3.1 The SLU breeding program: methods

The potato breeding program at SLU follows a traditional breeding approach, where selections are based on the breeder's eye and scores for a number of breeding traits. No kind of high throughput phenotyping methods, MAS or GS are practised in the programme. All breeding material is tetraploid, and the breeding parents are either released cultivars with desirable traits, or elite breeding lines from within the SLU program. The approximated number of clones, replicates and field sites for each clonal generation is presented in **Table 1**. The seed tubers in the first clonal generation (T_1) are produced in greenhouse from plants produced from seeds. From T_3 and onwards, seed tubers are produced in fields in Umeå in northern Sweden, to limit tuber borne pests. The breeding program uses three field sites for scoring plants in generations T_1 to T_8. Two field sites are located in Southern Sweden, Mosslunda (approximately, 55°98'N, 14°10'E) and Helgegården

(approximately, 56°02'N 14°07'E) and one in Northern Sweden, Umeå (approximately, 63°84'N 20°26'E). In Mosslunda, all generations are planted and scored for traits of interest. Helgegården and Umeå hosts plants from T_4 and onwards. Fungicides are only applied in the field in Helgegården. In the fields situated in Mosslunda and Umeå, only insecticides are applied (personal communication by Fredrik Reslow, practical breeder at the SLU potato breeding program).

Table 1. Summary of each clonal generation in the SLU potato breeding program with number of accessions, number of field sites and number of replicates (Fredrik Reslow, personal communication).

Clonal generation (T_i)	Number of clones	Number of field sites	Number of replicates
T_1	8000–9000	1	0 (1 plant)
T_2	250–300	1	0 (1 plot of 10 plants)
T_3	50–60	1	2 (2 plots of 10 plants)
T_4	25	3	2 (2 plots of 10 plants)
T_5	10	3	2 (2 plots of 10 plants)
T_6	5–7	3	2 (2 plots of 10 plants)
T_7	2–3	3	2 (2 plots of 10 plants)
T_8	1–2	3	2 (2 plots of 10 plants)

2.3.2 The SLU breeding program: target traits

The target market for the SLU breeding program is fresh potatoes. A number of traits are scored and weighted against each other when selections are made (personal communication by Fredrik Reslow, practical breeder at the SLU potato breeding program).

Tuber yield

Yield is one of the most important traits for any plant breeder. Potato tuber yield can be measured using different components including per plant, or plot, tuber weight (TW) and tuber number (TN) or as tuber weight per hectare in larger trials. Yield is a quantitative trait, and is highly affected by

the environment as well as the genotype (Bradshaw, 2021). At the SLU breeding program, tuber yield is studied from T_1 and onwards.

Tuber quality

For the fresh potato market, it is important to consider traits such as tuber size, shape, uniformity of size and shape, tuber eye depth, skin colour, flesh colour and scab infections. Tuber size, shape and eye depth are scored in T_1 as an overall assessment at harvest in field. Skin and flesh colour are both qualitative traits (Bradshaw, 2021), and the desired colours for the Swedish market are yellow to pink skin, and yellow to light yellow flesh. The desired tuber finish is round to slightly oval, smaller tubers and with shallow set eyes (personal communication by Fredrik Reslow, practical breeder at the SLU potato breeding program). Scab is visually determined at SLU, and the breeder does not distinguish between different pathogens – which may be caused by either fungi or bacteria (Burton, 1989).

Cooking qualities

As the main end use for the fresh potato market is cooking through boiling, the tuber quality after cooking is tested in the breeding program from T_5 and onwards. The desired tuber is neither too mealy, where the tuber falls apart, nor too firm. The breeder also excludes accessions that exhibit post-cooking blackening. This blue-greyish discolouration may occur as chlorogenic acid combines with iron to form a complex which oxidizes when cooling to the coloured ferri-dichlorogenic acid (Bradshaw, 2021). The flavour and texture of the tubers are also scored after boiling.

Starch content

Specific gravity (SG) is used as a non-destructive way of assessing tuber dry matter content. The major component of tuber dry matter is starch. Starch will affect the texture of cooked potatoes, and may give an indication to if the accession might be targeted for the starch production market. Dry matter content is a quantitative trait and is affected by both genotype and environment (Bradshaw, 2021). SG is scored from T_4 and onwards.

Glycoalkaloid content

Glycoalkaloids are present in all potato cultivars and are considered to be

toxic compounds. High concentrations of glycoalkaloids results in a bitter-tasting tuber, and concentrations above 20 mg 100 g^{-1} fresh weight are considered unsafe for human consumption. Introgression breeding with wild *Solanum* relatives often results in higher concentrations of glycoalkaloids (Bradshaw, 2021). At SLU, the glycoalkaloid content is scored in individuals in T_8.

Plant maturity
The growth season in Sweden is short, with long days, and therefore it is desirable to develop cultivars which can utilize these conditions with early maturity (Eriksson *et al.* 2016). At the SLU breeding programme, plant maturity is scored as number of days until flowering. Scoring takes place in T_2 and onwards. Early maturity is especially desirable in northern Sweden, where the growing period is short.

Late blight
The oomycete *Phytopthora infestans* is the causal agent of late blight, which may affect potato foliage, stems and tubers (Henfling, 1987). The disease is potentially the most serious one in potatoes worldwide, and the pathogen can spread in fields very quickly if left uncontrolled. As spraying is expensive, and the new populations of *P. infestans* often are resistant to the conventional fungicides, there is a need for resistant cultivars (Bradshaw, 2021). A correlation between foliar and tuber LB has been found in some germplasm (Stewart *et al.* 1994), however this is not true for all germplasm (Bradshaw, 2021). At SLU, foliar LB is scored from T_2 and onwards in Mosslunda and Umeå, where the disease is allowed to spread naturally. Tuber LB is scored for T_4 and T_5 in a laboratory setting through the inoculation of tubers with *P. infestans*. Several genes conferring resistance against LB have been found in wild *Solanum* relatives. Host plant resistance to LB is quantitative and controlled by several major QTL (Bradshaw, 2021).

Early blight
The fungus *Alternaria solani* is the causal agent of the potato disease early blight (EB). EB is a serious disease for potatoes grown in warm climates, and might be increasing in Europe and Sweden (Kapsa and Ozowski, 2012; Runno-Paurson *et al.* 2015). Just like LB, EB may affect both foliage and

tubers. EB is not scored for at the SLU potato breeding program, as the first symptoms usually occur at the end of the growth season and do not seriously affect the yield for table potato in Sweden (personal communication by Professor Erland Liljeroth, Department of Plant Protection Biology, SLU).

2.4 Genotyping methods for potato

Genotypic data from plants has become more accessible to plant breeders as the development of new DNA marker systems have advanced over the last three decades. DNA markers are defined as a nucleotide sequence where a polymorphism can be detected between different individuals. Below, a selection of DNA marker systems are discussed.

The first DNA marker technique developed was restriction fragment length polymorphism (RFLP). The principle of RFLP makers is that genomic DNA from different individuals is digested into DNA fragments of varying size by restriction enzymes. The size of the DNA fragment can be used as a means of distinguishing differences between individuals at specific loci, due to mutations in the enzyme restriction site. This method was applied when developing the first DNA-based genetic linkage map (Sambrook *et al.* 1975; Yang *et al.* 2013).

Random amplified polymorphic DNA (RAPD) is a marker system that is based on the polymerase chain reaction (PCR) technique. The RAPD markers depend on short primers where mutations within the target sites are visualised in a stained agarose gel (Williams *et al.* 1990; Welsh and McClelland 1990). Amplified fragment length polymorphism (AFLP) is another PCR-based marker system, and a continuation of the RFLP and RAPD techniques that is less labour intensive and more robust compared to its precursors (Zabeau and Vos, 1993; Vos *et al.* 1995).

With the increasing availability of DNA sequencing data –a technique which can identify the individual nucleotide bases and the order thereof– sequenced-based marker systems has been an increasingly popular genotyping method (Shendure *et al.* 2017).

Single nucleotide polymorphism (SNP) markers take advantage of the millions of single base pair changes present in a genome within species (Lander, 1996). There are several methods for genotyping using SNP markers, such as chip-based next generation sequencing (NGS) and genotyping by sequencing (GBS). There are several different NGS techniques, but they are all based on a similar methodology. DNA libraries are prepared, and randomly cut fragments of DNA are ligated with universal adapters at both ends. The sequencing is performed continuously where one or more nucleotides are incorporated which results in a release of signal that can be detected by the sequencer (Metzker, 2010). The predetermined set of SNP markers included in each chip-based NGS method, have a physical position that maps to the reference genome. The most current available SNP chip for potato contains 20 000 SNP markers (Vos *et al.* 2015). GBS is a NGS technique where SNPs are discovered at any site of the genome by the utilization of restriction enzymes, and this can be used to produce a set of SNP markers specific for the collection of individuals included in the study (Elshire *et al.* 2011).

Diversity array technology (DAtT) markers are becoming increasingly popular. DArT can detect variations in DNA bases without the need for sequence information (Wenzl *et al.* 2004), and microarrays containing DArT markers are highly reproducible for any species this technique is applied to.

Older marker systems, such as RFLP, RAPD and AFLP do not offer uniform distribution of markers across the genome, even at higher marker density (**Table 2**). This, together with the older marker systems often requiring intensive laboratory work and are time consuming and expensive has pushed more plant breeding researchers to utilize newer techniques such as SNP chips, and SNP discovery by GBS or DArT for applications in molecular breeding methods (Agarwal *et al.* 2008; Yang *et al.* 2013; Nadeem *et al.* 2018).

Table 2. Comparison of important characteristics of a selection of molecular markers commonly used for plant breeding research. After Nadeem et al. 2018.

Characteristics	RFLP	RAPD	AFLP	SNP	DArT
Co-dominant/Dominant	Co-Dominant	Dominant	Dominant	Co-Dominant	Dominant
Reproducibility	High	High	Intermediate	High	High
Polymorphism level	Medium	Very high	High	High	High
Required DNA quantity	High	Medium	Low	Low	Low
Marker index	Low	High	Medium	High	High
Cost	High	Less	High	Variable	Cheapest
Sequencing	Yes	No	No	Yes	Yes
PCR requirement	No	Yes	Yes	Yes	No
Visualization	Radioactive	Agarose gel	Agarose gel	SNP-VISTA	Microarray

2.5 Marker assisted selection and QTL mapping for potato

A potential method to speed up the process of plant breeding is through the application of molecular techniques. The use of diagnostic DNA markers, which associate genotype with phenotype, allows for more efficient screening of a large number of accessions at earlier stages of the breeding process. Genetically elite accessions can be identified, while accessions with undesired alleles can be discarded, even before a phenotype can be observed. Such a selection process is called marker-assisted selection (MAS), and has the potential to significantly reduce the amount of time and money required in conventional plant breeding.

Fixation of desired alleles is slow in potato breeding due to the nature of tetrasomic inheritance, and even slower when considering several unlinked QTL (Bradshaw, 2021). Hence, the implementation of MAS in potato breeding has been quite limited. However, there are a few examples of traits where MAS has been applied, such as resistance to potato virus Y and potato cyst nematode (Ortega and Lopez-Vizcon, 2012), resistance to *P. infestans* and potato virus X (Mori *et al.* 2015), tuber chip colour, tuber starch content and starch yield (Li *et al.* 2013). The first traits to be mapped were monogenic traits, such as flower colour (van Eck *et al.* 1993) and tuber skin

colour (van Eck *et al.* 1994). However, as these traits are easy to score phenotypically, the application of MAS for them is seldom pursued. In the early 1990s, resistance genes to major pathogens were mapped, for example *P. infestans* (El-Kharbotly *et al.* 1994), the root cyst nematode *Globodera rostochiensis* (Gebhardt *et al.* 1993) and the potato cyst nematode *Globodera pallida* (Rouppe van der Voort *et al.* 2000). To map the traits in the research previously mentioned, diploid potato crosses were used. Using diploid plants to construct genetic maps is a way to avoid complex analysis, and allows application of a larger number of software. Mapping traits using tetraploid potato started a bit later with resistance genes against potato virus Y (Brigneti *et al.* 1997), *P. infestans* (Li *et al.* 1998) and *G. pallida* (Bryan *et al.* 2002).

For a genetic marker to be of use in MAS, it is crucial that the marker has a significant association with a QTL for the trait of interest. First however, the QTL has to be identified. Since the 1980s the utilization of markers for quantitative traits has been ministered through the advances in genotyping methods and statistical methods for detecting QTL. Linkage analysis is a method widely used for detecting QTL, and is available for tetraploid potato through software such as TetraploidMap (Hackett and Luo, 2003), PERGOLA (Grandke *et al.* 2017) and Polymapr (Bourke *et al.* 2018). Linkage analysis requires biparental mapping populations, and is hence limited to the variation available therein. Genome-wide association studies (GWAS) is an alternative statistical method for QTL mapping. GWAS utilizes the association between markers and phenotypic traits, which arise from linkage disequilibrium across a diverse panel of individuals. Compared to linkage analysis, GWAS offers increased high-resolution mapping as historical recombination events and natural diversity may be exploited as compared to the limited recombination available in a single cross (Zhu *et al.* 2008). GWAS for tetraploid potato is available through the software GWASpoly (Rosyara *et al.* 2016).

Once a QTL has been mapped, it has to be verified in other independent populations to make sure that the QTL is still effective for individuals with different genetic backgrounds to where the QTL was identified. Subsequently, candidate DNA markers in close proximity to the QTL has to be identified, to ensure that the marker and QTL are inherited together and limit the number of individuals with recombination between marker and

QTL. To confirm that the selected marker is tightly linked to the QTL, the marker has to be tested in independent populations with genetic backgrounds different to where the QTL and marker was identified. (Collard and Mackill, 2008).

2.6 Genomic selection for potato

GS is a form of MAS where genome-wide marker effects are simultaneously estimated to predict breeding values of individuals (Meuwissen *et al.* 2001). In contrary to MAS, GS can be applied to traits with a complex genetic background, where many minor QTL affects the trait, and it has been estimated to accelerate breeding cycles and thus increasing the genetic gain per unit of time (Desta and Ortiz, 2014). In theory, GS has the potential to capture all the underlying information to explain the genetic variation of a trait, as it is assumed that all QTL are in linkage disequilibrium with at least one marker. The genome-wide marker effects are estimated through the regression of a trait of interest (phenotype) on all markers (genotype) in a set of individuals included in a so-called training population. This multiple regression analysis gives rise to a computational problem, as the number of markers (p) often far exceeds the number of individuals (n). The problem that arises is that the variance of the estimator (marker effect) increases with the increase of $p > n$. This computational problem of GS methods can be solved either by reducing the number of variables, or by shrinkage estimations, or a combination of both. By shrinking the estimates toward a fixed point, the variance of the estimator will decrease, however at the same time the estimator bias may increase. There are several procedures for shrinkage estimation, the most common types are penalized and Bayesian methods. Bayesian ridge regression (BRR) is a model that is a popular method for estimating marker effects. The BRR method uses a Gaussian prior and performs a homogenous shrinkage across all markers, which might not be optimal when, in reality, some markers will be linked to QTL while others are not. Other models use other priors that make them more flexible, and model optimization has to be performed to ensure the selection of a model that fits both the genetic background of the trait of interest and individuals included in the study (de los Campos *et al.* 2013). Model optimisation is explored by the inclusion of a so-called testing population. For the testing population, each individual is assigned a genomic estimated

breeding value (GEBV) based on their genotype and the marker effects which were estimated in the training population. The individuals in the testing population can either be a set of individuals from a different population than the training population, or a randomly drawn subset of individuals from the training population as is used in the case of cross-validation approaches. The accuracy of the model is evaluated by calculating the correlation of the phenotypes to the GEBVs, which yields model predictive ability. Sometimes the calculation of predictive ability also includes dividing the correlation value with the square root of the trait narrow-sense heritability. The model predictive ability will however also depend on a number of other variables, such as the genetic background of the trait studied, the relative size of the testing population compared to the training population and the population structure of the individuals. (Habier *et al.* 2007).

To use GS in a breeding approach, selections are made based on the individual's GEBV, without the necessity to phenotype an entire breeding population at all cycles of selection (**Figure 4**). GS was first implemented as a breeding method for dairy cattle (Hayes *et al.* 2009), and a number of studies have reported promising results for breeding of other animal and plant species, for example, pig (Knol *et al.* 2016), sheep (Daetwyler *et al.* 2012), wheat and maize (Crossa *et al.* 2010), soybean (Jarquín *et al.* 2014) or eucalyptus (Resende, 2012). For potato, GS is still in its infancy (Bradshaw, 2017; Bradshaw, 2021). Habyarimana *et al*. (2017) were the first to explore the GS method for potato breeding in practise. They studied the predictive ability of a number of breeding traits including tuber yield, tuber flesh colour and tuber dry matter content. For genotyping their selection of 190 potato cultivars and breeding clones, they used approximately 50 000 SNPs obtained from a DArT marker system. Habyarimana *et al.* (2017) concluded that it is possible to use GEBVs as a method for potato breeding based on their results. Since 2017, a relatively small number of papers on GS for potato breeding has been published. Sverrisdóttir *et al.* (2017) concluded that the GBS marker system works for estimating GEBVs in potato for selections in tuber starch content and chipping quality. The material in this publication further investigated by the inclusion of two other populations, where one was a distinct collection of cultivars and breeding clones from the

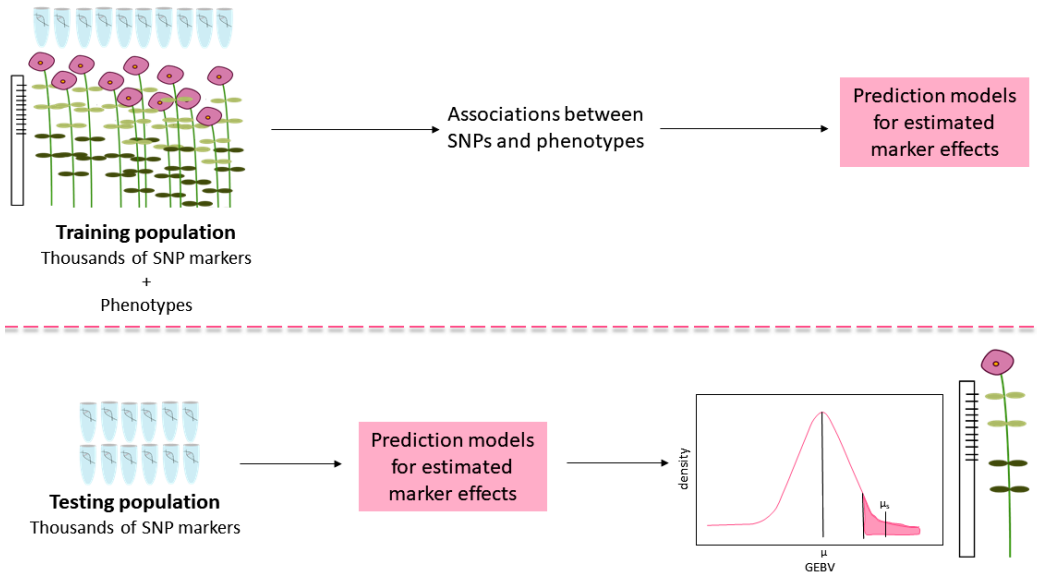


Figure 4. Genomic selection for plant breeding. Individuals comprising a training population are genotyped (with SNP markers) and scored for a phenotypic trait (e.g. plant height). A prediction model is constructed based on associations between SNPs and phenotype with estimated effects of each molecular marker. A separate set of individuals, for which only genotypic data is available, makes up the testing population. For each individual in the testing population, a genomic estimated breeding value (GEBV) is estimated based on the predicted marker effects from the model. Selections can subsequently be made to select individuals with the best performing GEBVs, and the population mean (μ) will move towards the selected mean (μ_s).

United Kingdom (Sverrisdóttir *et al.* 2018). Here, the authors studied the model predictive ability of tuber dry matter content and chipping quality across and within populations, and concluded that predictions across breeding populations are unreliable, but that an expansion of individuals included in the training population may increase model predictive ability. Stich and Van Inghelandt (2018), Enciso-Rodriguez *et al.* (2018) and Endelman *et al.* (2018) all used SNP markers from NGS chips for genotyping the populations in their GS research. Stich and Van Ingheland (2018) investigated GS as a breeding approach for selections in a number of breeding traits including resistance to late blight, plant maturity and tuber yield using a number of different models. They noted that the inclusion of diagnostic markers should be modelled as fixed effects, and that traits with a complex genetic background benefits from the use of models which accounts for additive and dominance effects (i.e., using another prior than the one used in BRR). Enciso-Rodriguez *et al.* (2018) also examined resistance to late

blight and common scab and found that some increases of model predictive ability might be achieved by using a model that accounts for additive as well as dominance effects. Endelman *et al.* (2018) used a different approach compared to the other GS research mentioned above, by using genomic covariance matrices in a mixed model instead of prediction of marker effects, for prediction of GEBVs for their traits of interest (specific gravity, fry colour and tuber yield). Genomic covariances matrices accounting for additive, digenic dominant and additive × additive epistatic effects were calculated and the predictive abilities among the models were compared. This study also included a comparison of using a relationship matrix from pedigree data, which were found to be as effective as the covariance matrices in the prediction of GEBVs for specific gravity. Recently, GS research on potato has continued to explore possible dominance effects on prediction of breeding traits and found that these models are not superior to models limited to additive effects regarding the model predictive ability (Amadeu *et al.* 2020).

It is complicated to make any comparisons between research available through the literature on GS as a breeding method for potato since several aspects differ among them, for example population structure, genotyping and phenotyping methods, and estimation of model predictive ability. However, all research indicate that GS is a promising tool for making potato breeding a more effective endeavour.

2.7 Potato germplasm potential: landraces and gene banks

The diversity of the cultivated potato is vast. The centre of origin, the Andes in South America, are home to 3000 of the 5000 known potato cultivars. Most of these cultivars belong to the species *S. tuberosum*, but other *Solanum* species are also cultivated (Burlingame *et al.* 2009; Zaheer and Akhtar, 2016). Many of these cultivars are landraces, which have been domesticated and adapted to the local climate over a long time-span through traditional agricultural practises. Alongside the cultivated potato species, 199 wild potato species are known, which grow on both of the American continents (Hijmans and Spooner, 2001). Both landraces and wild species are important

sources of genetic diversity for potato breeders. For example, many genes for resistance against *P. infestans* have been identified in wild potato species, and landraces have been actively used in potato breeding to improve genetic gain since the middle of the 19th century (Ortiz, 2001; van der Vossen *et al.* 2003; Bradshaw and Ramsay, 2005; Bradshaw *et al.* 2006; Foster *et al.* 2009; Bernal-Galeano *et al.* 2020).

Collections of potato landraces and wild relatives are kept in genebanks all over the world. The genebank at CIP in Peru hosts the largest collection of cultivars (from 2*x* to 5*x*), but there are many large collections of potato accessions at genebanks in other parts of the world as well (Ortiz, 2001). In Sweden and the other Nordic countries, the Nordic Genetic Resource Centre (NordGen) maintains collected potato germplasm. In addition to accessions of landraces and wild potato relatives, genebanks may also host cultivars released from breeding enterprises that are not used in agricultural practise anymore, and farmer's cultivars. A farmer's cultivar is an accession which have been grown and adapted to local growth conditions using traditional agricultural methods. The role of genebanks for species conservation and as a source of germplasm for crop improvement is well recognized, and the collections are often used as material in potato research (Hijmans *et al.* 2000; Bradshaw and Ramsey, 2005). Recently, potato genebank materials have been included in genetic diversity research using high throughput marker systems in CIP, Japan, China and USA (Ellis *et al.* 2018; Igarashi *et al.* 2019; Wang *et al.* 2019; Pandey *et al.* 2021). The knowledge accumulated from their research are important to potato breeders worldwide.

3. Thesis aims

Genetic gain, which can be defined as an increase of the target or desired trait per year, is limited in potato compared to other major crops (Rijk *et al.* 2013). There are two main reasons behind this limitation of genetic gain in potato: 1) the genetics of the species – potato is a highly heterozygous outcrossing crop displaying a tetrasomic pattern of inheritance – and 2) the potato market which is highly fragmented and requires consideration of many traits in addition to yield. There have been many research efforts to investigate how to improve the genetic gain in potato. In this thesis, the potential of genomic-led methods to aid potato breeding in Sweden is explored. The specific aims of this thesis were:

- To map the genetic background of host plant resistance to EB in foliage and tubers and identify novel QTL (Paper I)
- To explore the potential of GEBVs for selection of potato breeding traits such as tuber yield, tuber quality and host plant resistance to LB (Paper II)
    - Through genomic predictions across clonal generations (Paper II)
    - Through genomic predictions across and within full-sib families (Paper II)
- To explore the genetic structure of the NordGen potato germplasm as a population for potential genetic hybridization of the SLU potato breeding material (Paper III)

4. Methods

4.1 Plant material and genotyping

For mapping resistance to EB (paper I), a cross between a susceptible cultivar ('Matilda') and resistant cultivar ('Magnum Bonum') was made for maximum segregation of QTL. This resulted in a bi-parental crossing population containing 80 tetraploid offspring exhibiting widespread infection patterns to EB. Neither 'Matilda' nor 'Magnum Bonum' are frequently used as parents at the SLU potato breeding program. 'Matilda' is a cultivar of Swedish origin released in 1986 and primarily used for the fresh potato market (Hutten and Berloo, 2001). 'Magnum Bonum' is an old cultivar released in 1876 in Great Britain, which was a very popular fresh potato cultivar until the 1980s, when it was found that this cultivar had high amounts of glycoalkaloids (Branzell and Hellenäs, 1999). The breeding germplasm in paper II and III were collected from the SLU potato breeding program. The breeding population in paper II contains accessions from all clonal generations of selection present in the program at the time (year 2018); i.e., T_1–T_8. However, the T_1 generation is reduced in size and limited to only five bi-parental offspring. The parents of these five hybrid offspring were selected due to exhibiting high resistance to late blight, possibly from different R-gene sources. Some of these parents (and grandparents) are also deployed as parents for accessions in later clonal generations (T_2–T_8). They were, among others, 'Sarpo Mira', 'Bionica', '93-1015' and 'C08II69'. '93-1015' and 'C08II69' are both breeding clones generated at the breeding program, while 'Sarpo Mira' and 'Bionica' are cultivars released in the 2000s from European breeding programs (Hutten and Berloo, 2001). In

paper III individuals from T_4 to T_8 are included and their genetic diversity of was compared with germplasm from the *in vitro* potato collections from the Nordic Genetic Resource Centre (NordGen).

The plant material was genotyped using SNP chips, containing 10 000 SNP markers (paper I) or 20 000 SNP makers (papers II and III). Genotypic data from SNP chips is commonly used in research such as that of papers I and II, as this method is easily available and user friendly with a low frequency of missing genotypes compared to GBS. Even though SNP chips are often deployed for gene discovery studies such as paper I, the method may be less suitable for these types of studies due to ascertainment bias. This bias arises due to the SNP chip being developed using SNPs from a specific gene pool and subsequently applied to other, separate gene pools. This results in a limit in the variation that can be discovered when the gene pool used for developing the SNP chip does not match the experimental material (Lachance and Tishkoff, 2013).

DNA extraction, genotyping and allele scoring were conducted by the breeding company HZPC, Joure, The Netherlands (Paper I) and TraitGenetics GmbH, Gatersleben, Germany (Paper II & III). The genotypic data obtained was of high quality (no individuals had to be discarded due to a high degree of missing marker information) and marker reduction could be conducted following standard protocol of removing markers with a high degree of missing scores (< 10%) and removal of monomorphic makers (minor allele frequency > 0.5). Genotypic data were that of a tetraploid, exhibiting all five allelic states (AAAA, AAAO, AAOO, AOOO and OOOO).

4.2 Phenotyping

Data for several breeding traits was collected for the studies included in this thesis. All traits were collected from plants grown at field sites in Sweden. Yield is often considered the most important trait for plant breeders. Yield was measured at harvest using three different components – tuber weight (kg) per plant (TW), tuber number per plant (TN) and average weight (kg) per tuber (ATW). For the fresh potato market, tuber uniformity traits for traits such as shape, size and eye depth are of great importance. These traits were

measured using a discrete scale ranging from 1 (uniform) to 9 (non-uniform). Cooking-related quality traits were evaluated as tuber specific gravity (SG) and percent dry matter content. SG was recorded as the weight of tubers divided by the weight of tubers in water. Tuber dry weight was measured after water has been removed from the tuber, and the percentage of dry matter content was calculated as dry weight divided by fresh weight multiplied by 100 (Norell *et al.* 2016). Estimates of SG and percent dry matter content are known to be correlated, and SG was converted to percent dry matter content as described by Mosley and Chase (1993) for comparisons between SLU and NordGen material for paper III. Foliar resistance to early blight (EB) and late blight (LB) were measured in field throughout the growing period. EB and LB scores were calculated from the area under disease progress curve (AUDPC). AUDPC estimates the disease intensity over time and can thus summarise the possible polycyclic nature of the diseases (Forbes *et al.* 2014). Defoliation was scored along-side EB using the same scale. Both EB and LB can also affect potato tubers. Resistance to tuber EB was measured post-harvest through inoculation with the pathogen under controlled conditions, and scored as the volume of infected tuber tissue.

4.3 Linkage analysis

The 80 offspring from a bi-parental cross were scored for three traits: resistance to foliar EB, resistance to tuber EB and defoliation (Paper I). To identify QTL associated with these traits, TetraploidMap for Windows was used (Hackett *et al.* 2007). TetraploidMap (Hackett and Luo, 2003) and its successors TetraploidMap for Windows and TetraploidSNPMap (Hackett *et al.* 2017) are some of the most widely used software for linkage mapping and QTL identification in tetraploid potato. TetraploidMap has been used to map QTL for traits such as drought tolerance (Schumacher *et al.* 2021), chip colour (Rak *et al.* 2017) and yield (Bradshaw *et al.* 2008). TetraploidMap for Windows was developed before NGS technologies made SNP markers the most used genotyping method. The software was developed for genotypic data from AFLP and simple sequence repeat (SSR) markers, but can also be applied to SNP marker data. The crossing population used to map EB was genotyped using SNP markers. Two types of segregating markers were kept for linkage analysis, simplex (AOOO × OOOO) and duplex (AAOO × OOOO) markers. Marker segregation was estimated in the offspring, where

a simplex marker is expected to appear at a 1:1 ratio and duplex marker in a 5:1 ratio, with no double reduction (Hackett *et al.* 2001). In the next step, linkage maps for the two parents were constructed through cluster analysis, where the distance related to recombination frequencies is calculated. In each linkage group, the markers are then phased over the four copies of the chromosomes. After the development of ordered linkage maps, the QTL interval mapping for the three traits was carried out considering alleles inherited from each parent separately. The probability of a QTL at any location was expressed as the logarithm of odds ratio (LOD) score, which is the $\log_{10}$ of the ratio of the likelihood of the data given a QTL at this point to the likelihood of no QTL at this point. The significance of the QTL was determined through a 100-run permutation test. QTL were detected using a full model, where all QTL genotypes are considered, or one of several simpler models. For the simpler models, the QTL is controlled by a simplex or dominant duplex allele and the QTL genotypes are *Qq* and *QQ* or *Qq* respectively (Hackett *et al.* 2007). For simplex and dominant duplex models, it is possible to determine the QTL main effect, by calculating the mean phenotypic value of the offspring sharing the QTL genotype with the parent from which the marker was inherited.

4.4 Genomic prediction models

To test if GS could be applied as a method of selection for potato breeding, GEBVs were estimated from genomic prediction models for eight breeding traits for 200–669 potato accessions from the SLU breeding program. To estimate GEBVs the BRR model was fit:

$$\boldsymbol{y} = \mathbf{1}\boldsymbol{\mu} + \boldsymbol{X\beta} + \boldsymbol{e}, \qquad (1)$$

where $\boldsymbol{y}$ is a vector of phenotypes, $\boldsymbol{\mu}$ is the mean, $\boldsymbol{X}$ is a matrix of the markers used for genotyping, $\boldsymbol{\beta}$ is a vector of marker effects and $\boldsymbol{e}$ is a vector of normal deviates with mean 0. The modelling was performed using the package 'BGLR' in R 3.6.2 (Pérez and de los Campos, 2014; R Core Team 2021). 'BGLR' is by far the most widely used software for studies on GS in tetraploid potato (Habyarimana *et al.* 2017; Sverrisdóttir *et al.* 2017; Enciso-Rodriguez *et al.* 2018; Stich and Van Inghelandt, 2018; Sverrisdóttir *et al.* 2018; Caruana *et al.* 2019; Amadeu *et al.* 2020; Byrne *et al.* 2020) as it is

possible to apply to tetraploid genotypic data and includes a range of models for genomic prediction. The potato breeding population was partitioned into training and testing sets based on two criteria: (a) clonal generation of selection or (b) full-sib family. In the first approach (a) accessions which had undergone at least one cycle of selection (T_2 or T_{3+}, depending on phenotypic data availability) were used as a training population to estimate marker effects. This population was used to estimate GEBVs for the testing population comprised by unselected individuals (T_1 or T_2 depending on the trait of interest). For each partitioning, model 1 was fit. To estimate model predictive ability, the Pearson's correlation coefficient between observed and predicted values was calculated as:

$$\boldsymbol{r}\,(\boldsymbol{GEBV} : \boldsymbol{y}). \qquad (2)$$

For the second approach (b) two full-sib families from T_1 were used as training and testing populations respectively. Cross-validations within each family were also conducted, where the family was randomly partitioned using five-fold schemes, where one group acts as training and one as testing and the process was repeated until all individuals had a GEBV. The same cross-validation approach was tested on the whole breeding population (n=200–669, depending on phenotypic data availability) to allow for comparison with other GS research results. The model fit (1) and method to calculate model predictive ability (2) was the same for all methods.

4.5 Population structure estimates

Genetic markers have become a very popular tool to estimate population structure (Patterson *et al.* 2006). To estimate the population structure among the germplasms from the SLU potato breeding program and genebank material from NordGen several methods for population structure analysis was used. Principal coordinate analysis (PCoA) is a multivariate analysis which reduces the dimensions of the data while preserving the covariance among the samples. Unlike principal component analysis (PCA), which is also often applied when examining population structure, PCoA is based on a distance matrix. The PCoA is most often applied to variables with a quantitative or discrete scale, such as SNP markers. The data is reduced to a smaller set of principal coordinates (PCs) which each describe a portion of

the sample variation. The genotypes are projected in the space between the PCs, which is often reduced to maximum three PCs when visualized. For visualization, the most common method is a scatter plot, where the distance between genotype samples is considered to reflect the relationship among the individuals just like a PCA (Patterson *et al.* 2006). For calculating the distance between individuals, two methods were used – the Euclidian distance (paper II) and Nei's genetic distance (paper III).

To further study the population structure among the accessions from SLU and NordGen (paper III), two additional methods were used on the whole dataset. First, a heatmap based on Nei's genetic distance was generated. Secondly, the population structure was determined with the software 'STRUCTURE' (Pritchard et al. 2000) using an admixture model. 'STRUCTURE' is based on repeated cluster analyses where the log likelihood of *K* is estimated. The *K* value is the expected number of populations, or clusters, in your data (Pritchard *et al.* 2000; Evanno *et al.* 2005). 'STRUCTURE' allows for a more formal test of population structure compared the other methods used in this study.

The genetic relationship among NordGen accessions was also studied through a hierarchical cluster analysis based the Euclidian distance among the samples and visualised as a dendrogram. This method was not applied to the whole dataset, as the resulting dendrogram was too large for visualization.

5. Results and Discussion

5.1 Quantitative trait loci for resistance to early blight

A. solani causes EB in both potato foliage and tubers, and the occurance of the disease seem to be increasing in Europe (Kapsa and Ozowski, 2012; Runno-Paurson *et al.* 2015). Different potato cultivars exhibit different levels of resistance to EB, and it has been shown that late maturing cultivars have higher levels of resistance (Boiteux *et al.* 1995). However, as late maturing plants require an extended period of time in the field, the chance of other biotic and abiotic stresses affecting tuber yield and quality increases (Devaux *et al.* 2014; Kawar *et al.* 2018). Hence, breeding for host plant resistance would be one of the most effective ways to control EB in potato. Paper I includes the first QTL mapping of EB in tetraploid potato, and the QTL for foliar and tuber EB mapped could be used in developing markers for MAS of resistant cultivars.

Linkage maps were constructed for each of the two parents 'Matilda' and 'Magnum Bonum' using the 80 offspring. A majority of the 12 potato chromosomes were mapped over more than one linkage group. None of the 31 linkage groups for 'Matilda' or 36 groups for 'Magnum Bonum' mixed chromosomes, which could be confirmed by the physical positions of the SNP markers from the GGPv 3.0 SNP chip (Vos *et al.* 2015). The QTL analysis included three breeding traits: resistance to tuber EB, resistance to foliar EB and defoliation. Defoliation was studied to distinguish QTL controlling late maturity from QTL associated to foliar EB, as a strong correlation between the traits was identified in the crossing population

(**Figure 5**), and has previously been reported (Boiteux *et al.* 1995; Zhang, 2004). Since two parental linkage maps were used, it is not possible to distinguish if all QTL were unique, e.g. the two QTL for foliar EB on chromosome 12 might be the same QTL since they were found using both parental linkage maps (**Table 3**). The issue of developing separate linkage maps for the two parents could have been avoided if the software TetraploidSNPmap had been used in the analysis, where dosages from both parents can be estimated simultaneously (Hackett *et al.* 2017). The use of TetraploidSNPmap would also have allowed for usage of a higher number of markers for linkage mapping, and might have yielded a linkage map with

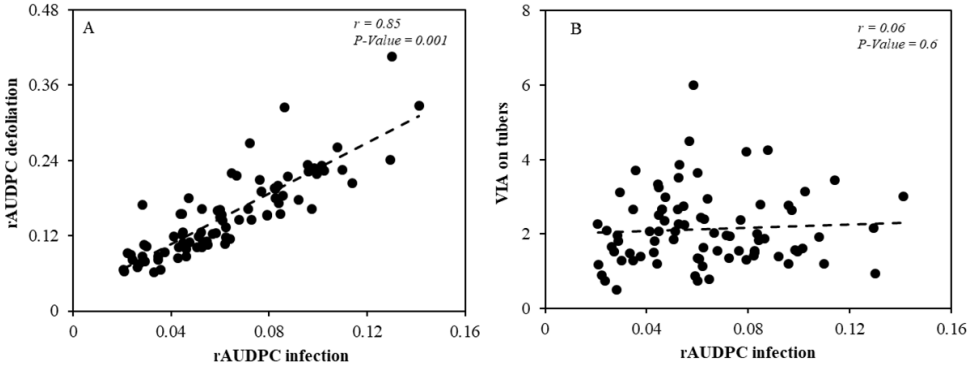


Figure 5. The association between (**A**) host plant resistance to foliar early blight (rAUDPC infection) and defoliation (rAUDPC defoliation), and (**B**) host plant resistance to foliar early blight (rAUDPC infection) and host plant resistance to tuber early blight (VIA). Correlations are estimated using Pearson's correlation coefficient. Reproduced from Odilbekov *et al.* 2020.

higher resolution.

In total, 33 significant QTL were detected for the three traits (**Table 2**): 11 for resistance to tuber EB, 11 for resistance to foliar EB, and 11 for defoliation. Out of the 11 QTL for resistance to foliar EB, two could be determined as independent from defoliation QTL. The QTL were detected using three different models: full, dominant duplex and simplex. For marker development for MAS, it is crucial that the SNP in closest proximity to the QTL is either dominant duplex or simplex, as the genotype of the QTL can be determined in these cases. Out of these two models, the simplex model is the most appropriate for marker development, as it only requires the introgression of a single copy of the allele. One of the two QTL for foliar resistance to EB was mapped to chromosome 5 using the simplex model, and

could be a candidate for marker development towards MAS. This QTL was found to have a positive effect on the offspring compared to the mapping parent, and is a good candidate for the search of genes conferring resistance to EB using transcriptomic analysis. Several of the QTL for tuber resistance to EB were found using a duplex dominant or simplex model. Host plant resistance to tuber EB has not been as intensely studied as foliar EB, but as there does not seem to be a correlation between the traits (Pearson's correlation coefficient (r) = 0.06, P = 0.35), it is of interest to study the genetic architecture of this trait separately. QTL for tuber EB were found on chromosomes 1, 2, 3, 4, 5, 8, 11 and 12, thereby suggesting that the trait is controlled by several QTL with small effects.

Table 3. The significant quantitative trait loci (QTL) found for three traits – tuber and foliar resistance to early blight (EB) caused by *A. solani* and defoliation. Mapping parent is the parent used to construct the linkage map the QTL was found in on chromosome (Chr) 1-12. The QTL interval position in centimorgan (cM) is estimated from the area of the linkage group where the genetic signal reached the significance threshold using. The physical size in base pairs (bp) is the estimated QTL interval using the flanking markers. QTL were deemed significant using a logarithm of the odds ratio (LOD) above 2. %PVE: percent phenotypic variance explained. The main Q effect was calculated using the SNP immediate to the QTL peak effect on the offspring phenotype compared to parent with corresponding genotype. Only significant effects ($p<0.05$) with greater (↑) and lower (↓) trait values in offspring are included; "-" indicates that no significant effect was detected in the offspring. Bold text identifies the foliar resistance to EB and defoliation QTL which were found to be independent from one another. After Odilbekov *et al.* 2020.

Trait	Mapping parent	Chr	QTL interval position cM	Physical size (bp)	LOD	PVE%	QTL genetic model	Marker closest to proposed QTL position	Homologous chr	Main Q effect
Tuber resistance	Magnum Bonum	1	56-60	50031365	2.492	8.19	full model			
		5	5.7-6.2	2451341	3.264	22.12	full model			
		8	6.6-34	46485633	3.687	50.6	duplex	snp_c2_28580	Q12	-
		11	19.5-39.7	2035814	3.215	48.70	full model			
		12	33.3-39.5	6007028	3.140	17.58	full model			
	Matilda	2	15.9-16	2504010	3.652	46.51	full model			
		3	25.2-34.7	11438315	2.524	12.82	simplex	PotVar0120608	C3	↑
		4	2.3-19.9	46382019	4.367	19.39	full model			
		8	52.3-68	6532990	2.816	20.52	duplex	PotVar0063945	Q13	↓
		11	15.4-20.4	1143780	3.045	30.23	duplex	PotVar0008448	Q23	↑
		11	11.7-18	3641233	3.069	35.96	duplex	snp_c2_49316	Q14	-
Foliar resistance	Magnum Bonum	1	0-26.4	65672744	2.439	13.06	duplex	PotVar0098794	Q14	↓
		5	**0-16.9**	**7341418**	**2.077**	**11.60**	**simplex**	**PotVar0026113**	**C1**	↑
		5	7.8-8.7	3035817	3.014	46.51	full model			
		5	0-28	3252350	9.857	43.36	simplex	PotVar0079374	C1	↓
		7	0-5.9	5308369	2.744	9.93	full model			
		12	7.7-8.5	1254555	2.134	35.81	duplex	PotVar0052987	Q13	-
	Matilda	5	19.8-20	13052157	2.717	17.16	full model			
		6	49.1-50	6406355	2.371	21.42	simplex	PotVar0087396	C2	-
		10	0-6.7, 10.9-30	33204844	3.093	15.68	duplex	PotVar0108340	Q24	↓
		11	**3.4-14.4**	**1549203**	**3.271**	**51.96**	**full model**			
		12	21.4-29	32470228	3.017	15.12	duplex	snp_c2_17613	Q34	↑
Defoliation	Magnum Bonum	**1**	**12.3-16.1**	**48738162**	**3.961**	**29.93**	**full model**			
		1	17.2-28.8	49687204	2.700	29.88	duplex	PotVar0044984	Q23	↓
		3	**0-2.1**	**17205892**	**3.327**	**42.61**	**full model**			
		3	**29.8-30.1**	**9970508**	**2.932**	**28.25**	**duplex**	PotVar0094921	**Q34**	-
		5	7.9-8.1	3035817	4.094	51.29	full model			
		5	0-28	3252350	8.104	36.87	simplex	PotVar0079374	C1	↓
		7	0-22	2135154	3.063	11.21	full model			
		8	**21.5-30**	**45302690**	**4.803**	**47.71**	**full model**			
		12	22.6-26.8	49762620	4.765	52.92	full model			
	Matilda	6	49.8-50	6406355	2.789	33.49	full model			
		10	15.5-16.3	433252	2.443	12.42	duplex	PotVar0108340	Q24	↓

5.2 Introducing genomic selection to a potato breeding program

In a conventional potato breeding program, the least robust phenotypic scores are obtained in the early clonal generations, where the number of individual breeding clones is very high, while the number of replicates for each clone is low (Gopal *et al.* 1992). To circumvent this issue, GEBVs may be applied for selection. GS models were generated with BRR for prediction of eight important breeding traits – TW, TN, ATW, LB, SG, and uniformity of tuber shape, size and eye depth – in early clonal generations (Paper II). The GEBVs were estimated by applying genomic prediction models in two distinct training populations: later clonal generations, or unselected full-sib families. The predictive ability of traits was low both for predictions across clonal generations (**Table 3**) and across full-sib families (**Table 4**), with exceptions for SG and LB. Both SG and LB have previously been reported to have moderate to high heritability (Ruttencutter *et al.* 1979; Haynes *et al.* 1995; Pajerowska-Mukhtar et al. 2009; Solano et al. 2014; Enciso-Rodriguez et al. 2018).

Table 4. Prediction ability (Spearman's correlation coefficient between observed and predicted phenotypes) for seven breeding traits; per plant tuber weight (TW), per plant tuber number (TN), average weight per tuber (ATW), specific gravity (SG) and uniformity of tuber size and shape and tuber eye depth. The model was validated either by training and testing populations divided by cycle of selection, or by randomly partitioned five-fold cross validations. *For the cross validation the total population divided into 80-20 fractions with accessions parted at random, mean prediction accuracies over 100 model runs. (Source: Selga *et al.* in press)

Training population	**Testing population**	**Breeding trait**						
		TW	**TN**	**ATW**	**SG**	**Size**	**Shape**	**Eye**
$T_{2\text{-}7}$	T_1	0.05	0.05	0.04				
$T_{3\text{-}7}$	T_2	0.07	0.04	0.18	0.43	0.16	0.03	0.15
80%*	20%*	0.75	0.72	0.39	0.62	0.17	0.045	0.15

This observation is in line with other research, as there is a correlation between highly heritable traits and prediction accuracy of GEBVs (Zhang *et al.* 2017). However, there is not a one-to-one relationship between trait predictive ability and trait heritability, and model prediction accuracies are mostly considerably lower than trait heritability (Kim *et al.* 2017). The broad-sense heritability estimates (H^2) for the SLU breeding population (Paper II) were calculated on a limited set of individuals (n = 38) for which data were available from all three field sites. H^2 for the yield traits TW, TN and ATW were found to be very high (0.64, 0.668 and 0.692, respectively) which is not in line with what have been previously found for potato. Yield is estimated to have low to medium narrow-sense heritability (Ortiz and Golmirzaie 2002; Schönhals et al., 2017), and is a classic example of a quantitative trait controlled by several minor loci (Bradshaw, 2021). Hence, the H^2 for these traits in our research may be overestimated, as it also includes non-additive effects.

Table 5. Prediction ability (Spearman's correlation coefficient between observed and predicted phenotypes) for four breeding traits; per plant tuber weight (TW), per plant tuber number (TN), average weight per tuber (ATW) and host plant resistance to late blight (LB). The model was validated by dividing accessions based on full-sib family or by randomly partitioned fivefold cross validation within each full-sib family. Cases with negative correlation; i.e., model not being able to predict trait are denoted with the symbol '-'. *For cross validation within the full-sib family, the accessions were divided at random and mean prediction abilities were estimated over 100 model runs. (Source: Selga *et al.* in press)

Training population	**Testing population**	**Breeding trait**			
		TW	**TN**	**ATW**	**LB**
T_1-C	T_1-D	0.088	-	0.101	0.29
T_1-D	T_1-C	0.069	-	0.080	0.31
T_1-C*		0.130	0.070	0.337	0.25
T_1-D*		0.300	0.260	0.182	0.16

The limited predictive ability across clonal generations could very well be a result of the high degree of population structure among the different populations (**Figure 6**). From the PCoA it became clear that the genetic overlap between T_1 and $T_{2–7}$ was very limited. The T_1 consisted of five half-

sib families, derived from a small set of parents, which makes the effective population size of T_1 smaller than later clonal generations, where the number of breeding parents are larger. The overlap between the five full-sib families was also remarkably small, possibly contributing to the low predictive ability across full-sib families seen for some traits (**Table 4**). For predictions across clonal generation, T_1 was used as testing population when predicting TW, TN and ATW. For predictions of SG, Eye, Shape and Size T_2 acted as testing population. The predictive ability for Eye, Shape and Size are also poor even though the population structure among later clonal generations was weak (**Table 3**).

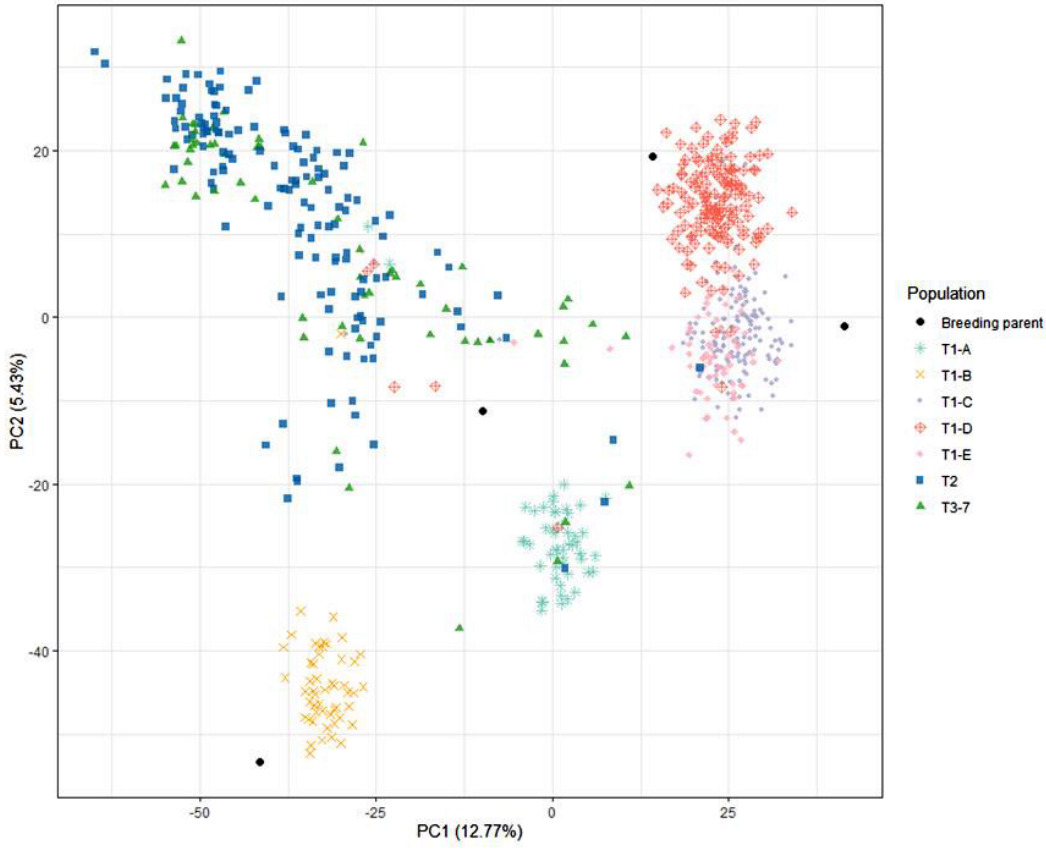


Figure 6. Principal coordinate analysis using principal coordinates (PCs) one and two based on the Euclidian genetic distance between 669 potato accession from the SLU breeding program. Cycles of selections (Ti) are grouped based on colours, T_1 – $T_{3–7}$. Five full-sib families of the T_1 are also separated by colour. (Source: Selga *et al.* in press)

A great problem when using later clonal generations from a crossbreeding program as a training population is that this population will naturally be smaller than the targeted testing population, since the number of breeding clones in each clonal generation is greatly reduced each year. Previous research on GS for potato, where training-testing populations were used, had a maximum of 30% of the total number of individuals as a testing set. In our study on predictions across clonal generations, these divisions are reversed, with the targeted testing population being far larger than the training population, e.g. the number of individuals in T_1 is 465 (our testing population), while the number of individuals in $T_{2\text{-}7}$ is 200 (our training population). This could be a problem, as the number of individuals over time in later cycles of selection will accumulate over time with the inclusion of historical data. The highest predictive abilities from this study were obtained from the within full-sib family predictions (**Table 4**). This approach could be implemented as a means of GS in early clonal generations were only a fraction of all individuals would be phenotyped and GEBVs can be acquired and used as a means of selection among the unphenotyped individuals of the same full-sib family.

5.3 Discovering the potential diversity of the potato germplasm in the Nordic genebank

For a plant breeder, knowledge of the diversity of their breeding germplasm is a fundamental tool for making informed decisions when breeding. Genebanks collects material such as landraces, farmer's cultivars, obsolete cultivars and breeding clones, and may provide accessions for broadening the genetic base of the breeding material. In paper III, the genotypic diversity of elite breeding material from SLU (breeding clones from T_3 and onwards) and germplasm from NordGen were examined for population structure together with a selection of cultivars grown today. The genotypes were grouped based on three of different criteria 1) the previously mentioned source of the accession (NordGen, SLU or other), 2) clonal type; i.e., farmer's cultivar, released cultivar, or breeding clone, and 3) country of origin.

No clear population structure was revealed based on the SNP marker data. The PCoA (**Figure 7A-C**) revealed a tendency of grouping among the earlier generations (T_3 and T_4) of breeding clones from SLU (**Figure 7B**), and hence

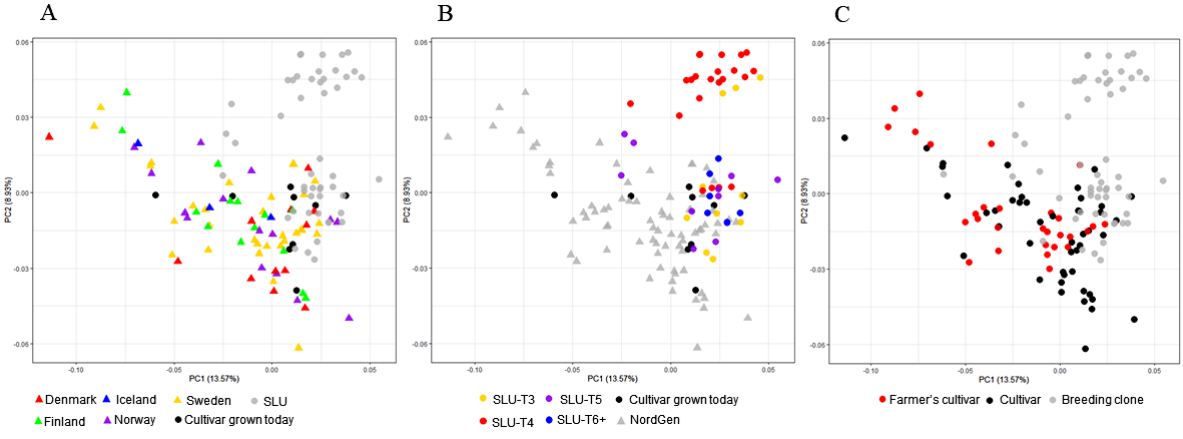


Figure 7. The genetic diversity of 133 accessions from SLU and NordGen. The principal component analysis uses the first two using principal coordinates (PCs) based on Nei's genetic distance of SNP markers among accessions. Coloured groupings are based on (**A**) country of origin for accessions from NordGen, (**B**) cycle of selection (T3 –T6+) for SLU breeding clones, and (**C**) clonal type. (Source: Paper III)

a tendency of breeding clones to cluster separately from the two other groups (**Figure 7C**). This grouping of SLU breeding material separately from the rest of the studied material was also visible in the results from STRUCTURE and in a heatmap visualisation (**Figure 8A-C**). According to STRUCTURE, the *K* with the maximum likelihood was 2, thereby suggesting two subpopulations among the genotypes included the study. However, the population structure found by STRUCTURE still suggests low population structure, as all individuals were an admixture of both subpopulations (**Figure 8A, Figure 8C**). According to the STRUCTURE results, farmer's cultivars clustered separately to the other clonal types, when this approach of grouping was used (**Figure 8A**). This clustering was visible in the heatmap (**Figure 8B**), but not detectable in the PCoA (**Figure 7C**). Similarly, data from a study on Chinese potatoes found farmer's cultivars clustering separately (Wang *et al.* 2019).

No clustering based on the country of origin of the accessions was detectable (**Figure 7A**). A dendrogram limited to only NordGen accessions also did not reveal clustering according to origin (**Figure 9A**). However, some clustering

according to clonal type occurred within the accessions from NordGen (**Figure 9B**).

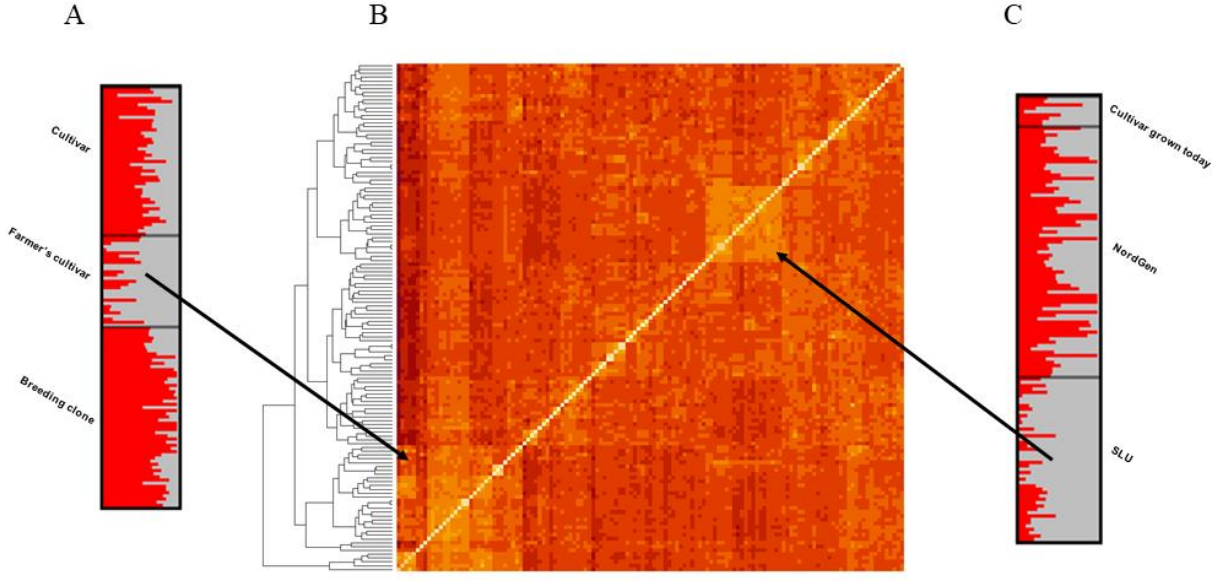


Figure 8. Population structure among the 133 accessions based on SNP markers (**A** and **C**) revealed by STRUCTURE and (**B**) heatmap with dendrogram showing the genetic diversity based on Nei's genetic distance. The two proposed subpopulations are separated by colour, grey or red. Three subpopulations were assumed: (**A**) cultivar, farmer's cultivar and breeding clone, and (**C**) cultivar grown today in Sweden, NordGen and SLU. The arrows to the heatmap **(B)** denotes the clusters corresponds to the subpopulations (majority of grey individuals in each assumed subpopulation) proposed by STRUCUTRE . (Source: Paper III)

Deploying a SNP chip for studying diversity may lead to ascertainment bias, as a panel of accessions were used when selecting the SNP markers included on the chip (Nielsen 2004; Albrechtsen et al. 2010; Heslot et al. 2013; Geibel et al. 2021). The limitation of accessions included in the chip development limits the heterozygosity of the loci and results in an underrepresentation of SNPs with extreme allele frequency. The SNP chip used in this study was developed using 569 unique genotypes selected from all over the world, with an emphasis on European cultivars (Vos et al. 2015). Some of the cultivars included in this study are represented among these European genotypes. These cultivars are often used as breeding parents in the SLU potato breeding program, and hence the genetic diversity of the SLU breeding germplasm should be captured by the array. However, only one of the accessions kept at

NordGen was included in the development of the chip. Hence some of the genetic diversity present in the NordGen accessions may be overlooked.

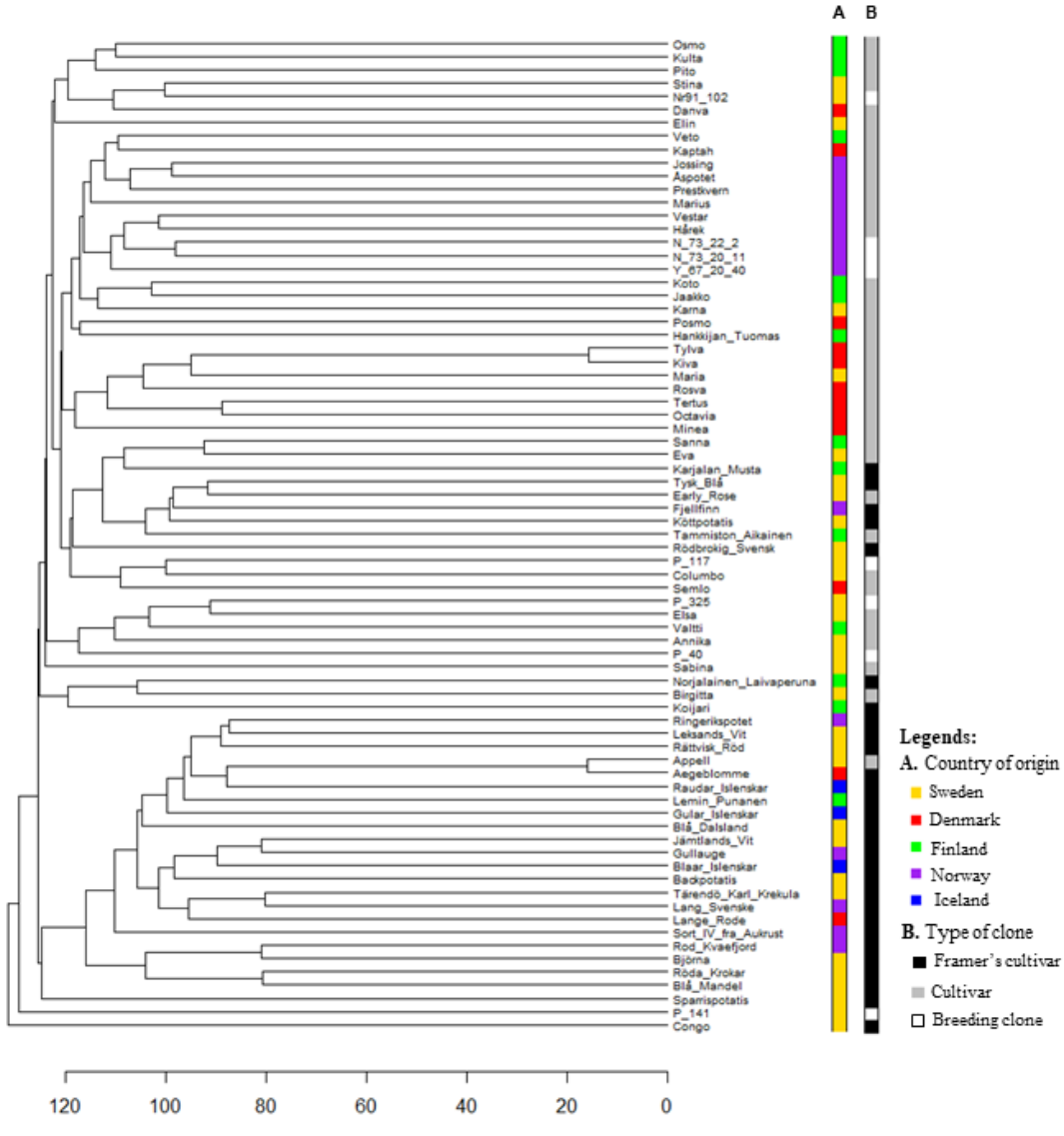


Figure 9. The Euclidian distance based on SNP markers among the 75 accessions from NordGen visualised as a dendrogram. Coloured groupings are based on (**A**) country of origin or (**B**) clonal type. (Source: Paper III)

In general, the genetic diversity among the 133 genotypes investigated was limited, which is expected since European cultivars originate from a very narrow genetic base (Glendinning 1983; Srivastava *et al.* 2016). The results from paper III suggest that for broadening the genetic base of the SLU breeding program, germplasm from genebanks outside the Nordic region should be sought out.

6. Conclusions

Applying genomic-led breeding methods, such as GS or MAS, may aid potato breeding by reducing the number of selection cycles, and by improving the selection accuracy in early clonal generations where the low number of replicates make phenotypic data unreliable for certain traits. The application of GS to a small potato breeding program was tested in the scope of this thesis. The results indicate that, although it is not possible to predict GEBVs using later clonal generations as training populations, predictions within and across full-sib families of unselected material may be possible. The application of this approach would limit the number of individuals to phenotype in the first clonal generation, thus reducing the size of the field experiment in T_1, with the possibility of saving time and money.

Though the application of MAS for potato breeding has been limited, there are successful examples. EB is a potato disease that in warm climates limits the potato yield. In Europe, the problem with EB may increase due to effects caused by climate change, such as increased temperatures and humidity. To limit the application of pesticides, development of resistant potato cultivars is central. We mapped resistance to EB in a tetraploid potato, and were able to find QTL independent from plant maturity. The mapping of these QTL may contribute towards finding molecular markers for MAS towards disease resistance in the future.

The study of the genetic diversity of breeding material and germplasm collections at genebanks is of interest to plant breeders when considering base broadening of breeding germplasm. Our results, based on molecular markers, suggest that the investigated potato germplasm from the breeding program at SLU and the accessions kept at NordGen are closely related. This

suggests that this particular germplasm is not useful for base broadening at the SLU breeding program, as the groups seem to have similar genetic backgrounds. However, a potential phenotypic diversity of the germplasm was not studied in the scope of this thesis, which may prove useful for introduction of specific traits such as resistance to pests or pathogens, or tuber and cooking qualities.

The limitations of available arable lands and the growth of human populations require increased crop yields, and an important tool for crop yield improvement is the development of new cultivars through plant breeding. Research on plant genetics, as well as to develop new breeding methods for future implementation, such as the research included in this thesis, may aid plant breeders in making informed decisions in their breeding programs.

7. Outlook

The discipline of plant breeding is ever evolving, and has benefited and will continue to benefit greatly from research on plant genetics. Genetic gain has been limited in potato compared to other crops, and Bradshaw (2017) identified the visual selection in early clonal generations to be one of the greatest bottlenecks in potato breeding. Genomic-led breeding, such as MAS, GS and identification of genetic structure in the breeding germplasm may aid potato breeders in making informed decisions for selection and for parental crossings, thus making potato breeding more efficient. Selections based on GEBVs for multiple traits in T_1 could potentially remove the need for phenotyping altogether in the first stage of potato breeding. However, the results from paper II suggest that advanced breeding material from the breeding program should not solely relied upon to train the GS models. More research on training population optimization is needed, but one suggestion could be a panel of advanced breeding clones and cultivars, often used as breeding parents in the program. For training population optimization, it will be vital to keep the genetic distance short to the clones where GEBVs will be predicted, while still including clones with a range of high to low phenotypic performance. This will require marker data as well as robust phenotypic data from a large number of accessions, but with the aims that this population will be able to predict many future breeding clones from within the same program, and with the possibility of updating the training population with more information each year from breeding clones from later T_i:s. The addition of high-throughput phenotypic methods may be valuable when scoring the phenotypic data for the training population, as the number of individuals might have to be large. Slater *et al.* (2016) suggested that the size of the training population to range between 500 to 4000 individuals, depending on heritability of the traits, but noted that with the inclusion of

breeding parents and relevant historic breeding material the size could be reduced.

Using MAS for potato breeding has been limited, possibly due to the complicated pattern of tetrasomic inheritance and additional levels of heterozygosity compared to diploid species. Hence, before implementing MAS in a breeding program, it is important to decide for which traits it should be applied. MAS will work efficiently when considering traits controlled by a small number of major QTL, which is often the case for host plant resistance to pathogens and pests. In paper I, QTL for resistance to EB were mapped. Linkage or genome-wide association analysis is the initial step towards developing markers for MAS. These findings will have to be validated in more environments and different populations, and further analysis applying transcriptomic methods to identify putative genes behind the QTL would be beneficial. Nevertheless, knowing the location of a resistance gene is also beneficial for other applicable plant breeding methods, namely gene modification or editing which also has a great potential to speed up the time of plant breeding.

To study the genomic-led breeding approaches in this thesis, tetraploid potato accessions have been used. However, with the recent advances in diploid potato breeding, there among the introduction of the *Sli* gene to allow for producing inbred lines (Eggers *et al.* 2021), the positive effects of applying these methods may speed up the process of potato breeding and hence increase the genetic gain for potato even further.

References

Agarwal, M., Shrivastava, N., Padh, H. (2008). Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Reports* 27, 617–631. DOI: 10.1007/s00299-008-0507-z

Albrechtsen, A., Cilius Nielsen, F., Nielsen, R. (2010). Ascertainment biases in SNP chips affect measures of population divergence. *Molecular Biology and Evolution* 27, 2534–2547. DOI: 10.1093/molbev/msq148

Amadeu, R.R., Ferrãro, L.F.V., de Bem Oliviera, I., Benevenuto, J., Endelman, J.B., Munoz, P.R. (2020). Impact of dominance effects on autotetraploid genomic prediction. *Crop Science* 60, 656–665. DOI: 10.1002/csc2.20075

Ames, M., Spooner, D.M. (2008). DNA from herbarium specimens settles a controversy about origins 373 of the European potato. *American Journal of Botany* 95, 252–257. DOI: 10.3732/ajb.95.2.252

Andersson, M., Turesson, H., Nicolia, A., Fält, A.-S., Samuelsson, M., Hofvander, P. (2017). Efficient targeted multiallelic mutagenesis in tetraploid potato (Solanum tuberosum) by transient CRISPR-Cas9 expression in protoplasts. *Plant Cell Reports* 36, 117–128. DOI: 10.1007/s00299-016-2062-3

Bernal-Galeano, V., Norton, G., Ellis, D. *et al.* (2020). Andean potato diversity conserved in the International Potato Center genebank helps develop agriculture in Uganda: the example of the variety 'Victoria'. *Food security* 12, 959–973. DOI: 10.1007/s12571-020-01037-8

Bernardo, R. (2020). Reinventing quantitative genetics for plant breeding: something old, something new, something borrowed, something BLUE. *Heredity*, 125, 375–385. DOI: 10.1038/s41437-020-0312-1

Boiteux, L.S., Reifschneider, F.J.B., Fonseca, M.E.N., Buso, J.A. (1995). Search for sources of early blight (*Alternaria solani*) field resistance not associated with vegetative late maturity in tetraploid potato germplasm. *Euphytica* 83, 63–70. DOI: 10.1007/BF01677862

Borlaug, N.E. (2002). Feeding a world of 10 billion people: the miracle ahead. *In Vitro Cellular & Developmental Biology – Plant*, 38, 221–228.

Bourke, P.M.; van Geest, G.; Voorrips, R.E. *et al.* (2018). Polymapr—linkage analysis and genetic map construction fromF1 populations of outcrossing

polyploids. *Bioinformatics* 34, 3496–3502. DOI: 10.1093/bioinformatics/bty371

Bradshaw, J.E. (2007). The canon of potato science: 4. Tetrasomic inheritance. *Potato Research* 50, 219–222. DOI: 10.1007/s11540-008-9041-1

Bradshaw, J.E. (2009). Potato breeding at the Scottish plant breeding station and the Scottish crop research institute: 1920–2008. *Potato Research* 52, 141–172. DOI: 10.1007/s11540-009-9126-5

Bradshaw, J.E. (2017). Review and analysis of limitations in way to improve conventional potato breeding. *Potato Research* 60, 171–193. DOI: 10.1007/s11540-017-9346-z

Bradshaw, J.E. (2021) *Potato breeding: theory and practise*. Springer Nature Switzerland AG, pp. 18–19, 21–23, 53–123, 125, 198–199, 203, 251–258, 367–396.

Bradshaw, J.E., Bryan, G.J., Ramsay, G. (2006). Genetic resources (including wild and cultivated *Solanum* species) and progress in their utilisation in potato breeding. *Potato Research* 49, 49–65. DOI: 10.1007/s11540-006-9002-5

Bradshaw, J.E., Hackett, C.A., Pande, B., Waugh, R., Bryan, G.J. (2008). QTL mapping of yield, agronomic and quality traits in tetraploid potato (*Solanum tuberosum* subsp *tuberosum*). *Theoretical -and Applied Genetics* 116, 193–211. DOI: 10.1007/s00122-007-0659-1

Bradshaw, J.E., Ramsay, G. (2005). Utilisation of Commonwealth Potato Collection in potato breeding. *Euphytica* 146, 9–19. DOI: 10.1007/s10681-005-3881-4

Bradshaw, J.E., Ramsay, G. (2009). Chapter 1 – Potato Origin and Production. In: Singh, J., Kaur, L. (eds) *Advances in Potato Chemistry and Technology*, Academic Press, pp. 1–26.

Branzell, C., Hellenäs, K.-E. (1999). Välj potatis av rätt sort: undvik höga halter av glykoalkaloider. *Vår föda* 3, 4–9.

Brigneti, G., Garcia-Mas, J., Baulcombe, D.C. (1997). Molecular mapping of the potato virus Y resistance gene Rysto in potato. Theoretical and Applied Genetics 94, 198–203.

Bryan, G., McLean, K., Bradshaw, J. et al. (2002). Mapping QTLs for resistance to the cyst nematode Globodera pallida derived from the wild potato species Solanum vernei. Theoretical and Applied Genetics 105, 68–77. DOI: 10.1007/s00122-002-0873-9

Burlingame B., Mouillé, B., Charrondière, R. (2009). Nutrients, bioactive non-nutrients and anti-nutrients in potatoes. *Journal of Food Composition and Analysis* 22, 494–502. DOI: 10.1016/j.jfca.2009.09.001

Burton, W.G. (1989). *The Potato*, 3rd edition Longman Scientific & Technical, New York, pp. 16–17, 21, 277–280, 384, 401–402.

Byrne, S., Meade, F., Mesiti, F., Griffin, D., Kennedy, C., Milbourne, D. (2020). Genome-wide association and genomic prediction for fry color in potato. *Agronomy* 10, 90. DOI: 10.3390/agronomy10010090

Cadman CH (1942) Autotetraploid inheritance in the potato: some new evidence. *Journal of Genetics* 44, 33–52.

Caruana, B.M., Pembelton L.W., Constable F., Rodoni B., Slater A. T. and Cogan N.O.I. 2019. Validation and genotyping by sequencing using transcriptomics for diversity and application of genomic selection in tetraploid potato. *Frontiers in Plant Science* 10, 670. DOI: 10.3389/fpls.2019.00670

Chujoy, E., Cabello, R. (2007). The canon of potato science: 29. True potato seed (TPS). *Potato Research* 50, 323–325. DOI: 10.1007/s11540-008-9058-5

Chung, B.N., Yoon, J.-Y., Palukaitis, P. (2013). Engineered resistance in potato against potato leafroll virus, potato virus A and potato virus Y. *Virus Genes* 47, 86–92. DOI: 10.1007/s11262-013-0904-4

Collard, B.C.Y., Mackill, D.J. (2008). Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. *Philosophical Transactions: Biological Sciences* 363(1491), 557–572. DOI: 10.1098/rstb.2007.2170

Crossa, J., de los Campos, G., Pérez, P. *et al.* (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics* 186(2), 713–724. DOI: 10.1534/genetics.110.118521

Daetwyler, H.D., Swan, A.A., van der Werf, J.H.J., Hayes, B.J. (2012). Accuracy of pedigree and genomic predictions of carcass and novel meat quality traits in multi-breed sheep data assessed by cross-validation. *Genetics Selection Evolution* 44, 33. DOI: 10.1186/1297-9686-44-33

de los Campos, G., Hickey, J.M., Pong-Wong, R., Daetwyler, H.D., Calus, M.P.L. (2013). Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics* 193, 327–345. DOI: 10.1534/genetics.112.143313

Desta, Z.A., Ortiz, R. (2014). Genomic selection: genome-wide prediction in plant improvement. Trends in Plant Science 19(9), 592–601. DOI: 10.1016/j.tplants.2014.05.006

Devaux, A., Kromann, P., Ortiz, O. (2014). Potatoes for sustainable global food security. *Potato Research* 57, 185–199. DOI: 10.1007/s11540-014-9265-1

Eggers, E.-J., van der Burgt, A., van Heusden, S.A.W. *et al.* (2021). *Nature Communications* 12, 4141. DOI: 10.1038/s41467-021-24267

El-Kharbotly, A., Leonards-Schippers, C., Huigen, D.J., *et al.* (1994). Segregation analysis and RFLP mapping of the R1 and R3 alleles conferring race-specific resistance to *Phytophthora infestans* in progeny of dihaploid potato parents. *Molecular and General Genetics MGG* 242, 749–754. DOI: 10.1007/BF00283432

Ellis, D., Chavez, O., Coombs, J. *et al.* (2018). Genetic identity in genebanks: application of the SolCAP 12K SNP array in fingerprinting and diversity analysis in the global in trust potato collection. *Genome* 61(7), 523-537. DOI: 10.1139/gen-2017-0201

Elshire, R.J., Glaubitz, J.C., Sun, Q. *et al.* (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6(5), e19379. DOI: 10.1371/journal.pone.0019379

Enciso-Rodriguez, F., Douches, D., Lopez-Cruz, M., Coombs, J., de los Campos, G. (2018). Genomic selection for late blight and common scab resistance in tetraploid potato (*Solanum tuberosum*). *G3: Genes, Genomes, Genetics* 8, 2471–2481. DOI: 10.1534/g3.118.200273

Endelman, J.B., Schmitz Carley, C.A., Bethke, P.C. *et al.* (2018). Genetic variance partitioning and genome-wide prediction with allele dosage information in autotetraploid potato. *Genetics* 209, 77–87. DOI: 10.1534/genetics.118.300685

Eriksson, D., Carlson-Nilsson, U., Ortíz, R., Andreasson, E. (2016). Overwiev and breeding strategies of table potato production in Sweden and the Fennoscandian region. *Potato Research* 59, 279–294. DOI: 10.1007/s11540-016-9328-6

Erjefält, L. (2001). The history of the potato in Sweden. In: Veteläinen, M., Bennvid, I.-B. (eds) Kartofler i Norden. En beskrivelse af gamle kartoffelsorter i Nordisk Genbank (The Nordic potato book), CAL-förlaget AB, Halmstad, pp. 56–62.

Evanno, G., Regnaut, S., Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14(8), 2611–2620. DOI: 10.1111/j.1365-294X.2005.02553.x

Evenson, R.E., Gollin, D. (2003). Assessing the impact of the green revolution, 1960 to 2000. *Science* 300(5620), 758–762.

FAO, IFAD, UNICEF, WFP and WHO (2021). The state of food security and nutrition in the world 2021. *Transforming food systems for food security,*

improved nutrition and affordable healthy diets for all. Rome, FAO. DOI: 10.4060/cb4474en

FAOSTAT Statistical Database, (2021). Food and Agriculture Organization of the United Nations, Rome. Available at: http://www.fao.org/faostat/ data retrieved 2021-09-29.

Forbes, G.A., Pérez, W., Andrade-Piedra, J. (2014). Field assessment of resistance in potato to *Phytophthora infestans*: International Cooperators Guide. (Lima, Peru: International Potato Center (CIP)), 8–17. DOI: 10.4160/9789290604402

Foster, S.J., Park, T.-H., Pel, M. *et al.* (2009). *Rpi-vnt1.1*, a *Tm-2*^{2} homologe from *Solanum venturii*, confers resistance to potato late blight. *Molecular Plant Microbe Interactions: MPMI* 22(5), 589-600. DOI: 10.1094/MPMI-22-5-0589

Gebhardt, C., Mugniery, D., Ritter, E., Salamini, F., Bonnel, E. (1993). Identification of RFLP markers closely linked to the H1 gene conferring resistance to *Globodera rostochiensis* in potato. *Theoretical and Applied Genetics* 85, 541–544. DOI: 10.1007/BF00220911

Geibel, J., Reimer, C., Weigend, S., Weigend, A., Pook, T., Simianer, H. (2021). How array design creates SNP ascertainment bias. *PLoS One* 16, e0245178. DOI: 10.1371/journal.pone.0245178

Glendinning, D.R. (1983). Potato introductions and breeding up to the early 20th century. *New Phytology* 94, 479–505.

Gopal, J., Gaur, P.C., Rana, M.S. (1992). Early generation selection for agronomic characters in a potato breeding programme. *Theoretical and Applied Genetics* 84, 709–713. DOI: 10.1007/BF00224173

Grandke, F., Ranganathan, S., van Bers, N., de Haan, J., Metzler, D. (2017). PERGOLA: fast and deterministic linkage mapping of polyploids. *BMC Bioinformatics*, 18, 12. DOI: 10.1186/s12859-016-1416-8

Grattapaglia, D. (2017) Status and perspectives of genomic selection in forest tree breeding. In: Varshney, R.K., Roorkiwal, M., Sorrells, M.E. (eds) Genomic selection for crop improvement. Springer, Cham, pp 199–249.

Gupta, P.K., Varshney, R.K. (2000). The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica*, 113, 163–185. DOI: 10.1023/A:1003910819967

Habier, D., Fernando, R.L., Dekkers, J.C.M. (2007). The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177, 2389–2397. DOI: 10.1534/genetics.107.081190

Habyarimana, E., Parisi, B., Mandolino, G. (2017). Genomic prediction for yields, processing and nutritional quality traits in cultivated potato (Solanum tuberosum L.). *Plant Breeding* 136, 245–252. DOI: 10.1111/pbr.12461

Hackett, C.A., Boskamp, B., Vogogias, A., Preedy, K.F., Milne, I., (2017). TetraploidSNPMap: Software for linkage analysis and QTL mapping in autotetraploid populations using SNP dosage data. *Journal of Heredity* 108(4), 438–442. DOI: 10.1093/jhered/esx022

Hackett, C.A., Bradshaw, J.E., McNicol, J.W. (2001). Interval mapping of quantitative trait loci in autotetraploid species. *Genetics* 159, 1819–1832. Pérez, P. and de los Campos, G. (2014). Genome-wide regression and prediction with the BGLR statistical package. *Genetics* 198, 483–495. DOI: 10.1534/genetics.114.164442

Hackett, C.A., Luo, Z.W. (2003). TetraploidMap: Construction of a linkage map in autotetraploid species. *Journal of Heredity* 94(4), 358–359. DOI: 10.1093/jhered/esg066

Hackett, C.A., Milne, I., Bradshaw, J.E., Luo, Z. (2007). Tetraploidmap for Windows: Linkage map construction and QTL mapping in autotetraploid species. *Journal of Heredity* 98(7), 727–729. DOI: 10.1093/jhered/esm086

Haverkort, A.J., Struik, P.C. (2015). Yield levels of potato: Recent achievements and future prospects. *Field Crops Research* 182, 76–85. DOI: 10.1016/j.fcr.2015.06.002

Hawkes, J.G., Francisco-Ortega, J. (1993). The early history of the potato in Europe. *Euphytica* 70, 1–421 7. DOI: 10.1007/BF00029633

Hayes, B.J., Bowman, P.J., Chamberlain, A.J., Goddard, M.E. (2009). Invited review: genomic selection in dairy cattle: progress and challenges. *Journal of Dairy Science* 92, 433–443. DOI: 10.3168/jds.2008-1646.

Haynes, K.G., Wilson, D. R., Kang, M. S. (1995). Genotype × Environment Interactions for Specific Gravity in Diploid Potatoes. *Crop Science* 35(4), 977–981. DOI: 10.2135/cropsci1995.0011183X003500040007x

Henfling, J.W. (1987). Late blight of potato. Vol. 4. International Potato Centre (CIP), Lima, Peru, pp. 6.

Hermundstad, S.A., Peloquin, S.J. (1987). Breeding at the 2x level and sexual polyploidization. *The production of new potato varieties: Technological advances* 197–210.

Heslot N, Rutkoski J, Poland J, Jannik J-L, Sorrells M.E. (2013) Impact of marker ascertainment bias on genomic selection accuracy and estimates of genetic diversity. *PLoS One* 8:e74612. DOI: 10.1371/journal.pone.0074612

Hijmans, R.J. (2001). Global distribution of the potato crop. *American Journal of Potato Research* 78, 403–412. DOI: 10.1007/BF02896371

Hijmans, R.J., Garrett, K.A., Huamán, Z., Zhang, D.P., Schreuder, M., Bonierbale, M. (2000). Assessing the geographical representativeness of genebank collections: the case of Bolivian wild potatoes. *Conservation Biology* 14(6), 1755–1765. DOI: 10.1111/j.1523-1739.2000.98543.x

Hijmans, R.J., Spooner, D.M. (2001). Geographic distribution of wild potato species. *American Journal of Botany* 88(11), 2101–2112. DOI: 10.2307/3558435

Hutten, R.C.B., Berloo, R. van (2001). An online potato pedigree database. URL: http://www.plantbreeding.wur.nl/PotatoPedigree/

Igarashi, T., Tsuyama, M., Ogawa, K., Koizumi, E., Sanetomo, R., Hosaka, K. (2019). Evaluation of Japanese potatoes using single nucleotide polymorphisms (SNPs). *Molecular Breeding* 39, 9. DOI: 10.1007/s11032-018-0917-8

Jansky, S.H., Charowski, A.O., Douches, D.S. *et al.* (2016). Reinventing potato as a diploid line-based crop. *Crop Science* 56 (4), 1412–1422. DOI: 10.2135/cropsci2015.12.0740

Jarquín, D., Kocak, K., Posadas, L. *et al.* (2014). Genotyping by sequencing for genomic prediction in a soybean breeding population. *BMC Genomics* 15, 740. DOI: 10.1186/1471-2164-15-740

Jo, K.-R., Kim, C.-J., Kim, S.-J. *et al.* (2014). Development of late blight resistant potatoes by cisgene stacking. *BMC Biotechnology* 14, 50. DOI: 10.1186/1472-6750-14-50

Kapsa, J.S., Osowski, J. (2012). Host-pathogen interaction between *Alternaria* species and *S. tuberosum* under different conditions. In: Thirteenth Euroblight workshop: St. Petersburg, Russia. *PPO-Special Report* 15, 107–112.

Kim, H., Grueneberg, A., Vazquez, A.I., Hsu, S., de los Campos, G. (2017). Will big data close the missing heritability gap? *Genetics* 207, 1135–1145. DOI: 10.1534/genetics.117.300271

Knight, T. A. (1807). On raising new and early varieties of the potato (*Solanum tuberosum*). *Trans. of the Horticultural Society of London* 1, 57–59.

Knol, E.F., Nielsen, B., Knap, P.W. (2016). Genomic selection in commercial pig breeding. *Animal Frontiers* 6(1), 15–22. DOI: 10.2527/af.2016-0003

Lachance, J., Tishkoff, S.A. (2013). SNP ascertainment bias in population genetic analyses: why it is important and how to correct it. *Bioessays* 35, 780–786. DOI: 10.1002/bies.201300014

Lander, E.S. (1996). The new genomics: global views of biology. *Science* 274(5287), 536–539. DOI: 10.1126/science.274.5287.536

Li, L., Tacke, E., Hoffbert, H.-R. *et al.* (2013). Validation of candidate gene markers for marker-assisted selection of potato cultivars with improved tuber quality. *Theoretical and Applied Genetics* 126, 1039–1052.

Li, X., van Eck, H.J., Rouppe van der Voort, J.N.A.M., Huigen, D.-J., Stam, P., Jacobsen, E. (1998). Autotetraploids and genetic mapping using common AFLP markers: the R2 allele conferring resistance to *Phytophthora infestans* mapped on potato chromosome 4. *Theoretical and Applied Genetics* 96, 1121–1128. DOI: 10.1007/s001220050847

Mackay G.R (2003) Potato breeding at SCRI during the last quarter of the 20th century. *Scottish Crop Research Institute Annual Report* 2001/2002 83–92.

Metzker, M.L. (2010). Sequencing technologies – the next generation. *Nature Reviews Genetics* 11, 31–46. DOI: 10.1038/nrg2626

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. DOI: 10.1093/genetics/157.4.1819

Milbourne, D., Pande, B., Bryan, G.J. (2007). Potato In: Kole, C (ed). Genome Mapping and Molecular Breeding in Plants, Volume 3 Pulses, Sugar and Tuber Crops, Springer-Verlag Berlin Heidelberg, pp. 208–210.

Mori, K., Asano, K., Tamiya, S., Nakao, T., Mori, M. (2015). Challenges of breeding potato cultivars grown in various environments and to meet different demands. *Breeding Science* 65, 3–16. DOI: 10.1270/jsbbs.65.3

Mosley AR, Chase RW (1993) Selecting cultivars and obtaining healthy seed lots. Potato Health Management. APS Press, St Paul, MN, pp 19-27.

Nadeem, M.A., Nawaz, M.A., Shahid, M.Q. *et al.* (2018). DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. *Biotechnology & Biotechnological Equipment* 32(2), 261–285. DOI: 10.1080/13102818.2017.1400401

Nakayasu, M., Akiyama, R., Lee, H.J., *et al.* (2018). Generation of α-solanine-free hairy roots of potato by CRISPR/Cas9 mediated genome editing of the St16DOX gene. *Plant Physiology and Biochemistry* 131, 70–77. DOI: 10.1016/j.plaphy.2018.04.026

Nielsen, R. (2004). Population genetic analysis of ascertained SNP data. *Human Genomics* 1, 218–224. DOI: 10.1186/1479-7364-1-3-218

Norell, R.J., Glaze, J.B., Chahine, M., Olsen, N. (2016). Measuring potato dry matter content on the farm. University of Idaho, CIS 1219.

Odilbekov, F., Selga, C., Ortiz, R., Chawade, A. & Liljeroth, E. (2020). QTL mapping for resistance to early blight in a tetraploid potato population. *Agronomy* 10, 728. DOI: 10.3390/agronomy10050728

Ortega, F., Lopez-Vizcon, C. (2012). Application of molecular marker-assisted selection (MAS) for disease resistance in a practical potato breeding programme. *Potato Research* 55, 1–13. DOI: 10.1007/s11540-011-9202-5

Ortiz R., Golmirzaie A. (2002). Hierarchical and factorial mating designs for quantitative genetic analysis in tetrasomic potato. *Theoretical and Applied Genetics* 104, 675–679. DOI: 10.1007/s001220100754

Ortiz, R. (2001). The state of the use of potato genetic diversity. In: Cooper, H.D., Spillane, C., Hodgkin, T. (eds) *Broadening the genetic base of crop production*. CABI Publishing, Wallingford, pp 181–200.

Ortiz, R., Crossa, J.F., Reslow, F., Pérez-Rodriguez, P., Cuevas, J.D. (submitted). Genome-based genotype × environment prediction enhances potato (*Solanum tuberosum* L.) improvement using pseudo-diploid and polysomic tetraploid modelling. Frontiers in Plant Science.

Ortiz, R., Simon, P., Jansky, S., Stelly, D. (2009). Ploidy manipulation of the gametophyte, endosperm and sporophyte in nature and for crop improvement: a tribute to Professor Stanley J. Peloquin (1921–2008). *Annals of Botany* 104, 795–807. DOI: 10.1093/aob/mcp207

Pajerowska-Mukhtar, K., Stich, B., Achenbach, U., Ballvora, A., Lübeck, J., Strahwald, J., *et al.* 614 (2009). Single nucleotide polymorphisms in the *Allene Oxide Synthase 2* gene are associated with 615 field resistance to late blight in populations of tetraploid potato cultivars. Genetics 181, 1115–1127. 616 DOI: 10.1534/genetics.108.094268

Pandey, J., Scheuring, D.C., Koym, J.W. *et al.* (2021). Genetic diversity and population structure of advanced clones selected over forty years by a potato breeding program in the USA. *Scientific Reports* 11, 8344. DOI: 10.1038/s41598-021-87284-x

Patterson, N., Price, A.L., Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genetics* 2(12), e190. DOI: 10.1371/journal.pgen.0020190

Piepho, H.-P., Koch, G. (2000). Codominant analysis of banding data from a dominant marker system by normal mixtures. *Genetics* 155(3), 1459–1468.

Pritchard, J.K., Stephens, M., Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics 155, 945–959.

Qaim, M. (2020). Role of new plant breeding technologies for food security and sustainable agricultural development. *Applied Economic Perspectives and Policy*, 42(2), 129–150. DOI: 10.1002/aepp.13044

R Core Team (2021). R: A language and environment for statistical computing. (Vienna, Austria: R Foundation for Statistical Computing). Available at https://www.R-project.org/

Rak, K., Bethke, P.C., Palta, J.P. QTL mapping of potato chip color and tuber traits within an autotetraploid family. *Molecular Breeding* 2017, 37, 15. DOI: 10.1007/s11032-017-0619-7

Resende, M.D.V., Resende Jr, M.F.R., Sansaloni, C.P. *et al.* (2012). Genomic selection for growth and wood quality in *Eucalyptus:* capturing the missing heritability and accelerating breeding for complex traits in forest trees. *New Phytologist* 194, 116–128. DOI: 10.1111/j.1469-8137.2011.04038.x

Rijk, B., van Ittersum, M., Withagen, J. (2013). Genetic progress in Dutch crop yields. *Field Crops Research* 149, 262–268. DOI: 10.1016/j.fcr.2013.05.008

Rosyara, U., De Jong, W.S., Douches, D.S., Endelman, J.B. (2016). Software for genome-wide association studies in autopolyploids and its application to potato. *The Plant Genome* 9, 2. DOI: 10.3835/plantgenome2015.08.0073

Rouppe van der Voort, J., van der Vossen, E., Bakker, E. *et al.* (2000). Two additive QTLs conferring broad-spectrum resistance in potato to *Globodera pallida* are localized on resistance gene clusters. *Theoretical and applied genetics* 101, 1122–1130.

Runno-Paurson, E., Loit, K., Hansen, M., Tein, B., Williams, I.H., Mänd, M. (2015). Early blight destroys potato foliage in the northern baltic region. *Acta Agriculturae Scandinavica Section B – Soil & Plant Science* 65(5), 422–432. DOI: 10.1080/09064710.2015.1017003

Ruttencutter, G., Haynes, F.L. & Moll, R.H. Estimation of narrow-sense heritability for specific gravity in diploid potatoes (Solanum tuberosum subsp.phureja andstenotomum). American Potato Journal 56, 447–452 (1979). DOI: 10.1007/BF02854064

Sambrook J., Williams, J., Sharp, P.A., Grodzicker, T. (1975). Physical mapping of temperature-sensitive mutations of adenoviruses. *Journal of Molecular Biology* 97, 369–390.

Schönhals, E.M., Ding, J., Ritter, E., Paulo, M. J., Cara, N., Tacke, E., et al. (2017). Physical mapping of QTL for tuber yield, starch content and starch yield in

tetraploid potato (*Solanum tuberosum* L.) by means of genome wide genotyping by sequencing and the 8.3 K SolCAP SNP array. *BMC Genomics* 18, 642. DOI: 10.1186/s12864-017-3979-9

Schumacher, C., Thümecke, S., Schilling, F. *et al.* (2021). Genome-wide aproach to identify quantitative trait loci for drought tolerance in tetraploid potato (*Solanum tuberosum* L.) *International Journal of Molecular Sciences* 22, 6123. DOI: 10.3390/ijms22116123

Selga, C., Reslow, F., Pérez-Rodíguez, P., Ortiz, R. (in press). The power of genomic estimated breeding values for selection when using a finite population size in genetic improvement of tetraploid potato. *G3 Genes|Genomes|Genetics.* DOI: 10.1093/g3journal/jkab362

Shendure, J., Balasubramanian, S., Church, G.M. *et al.* (2017). DNA sequencing at 40: past, present and future. *Nature* 550, 345–353. DOI: 10.1038/nature24286

Simmonds, N.W. (1997). A review of potato propagation by means of seed, as distinct from clonal propagation by tubers. *Potato Research* 40, 191–214.

Slater, A.T., Cogan, N.O.I., Forster, J.W., Hayes, B.J., Daetwyler, H.D. (2016). Improving genetic gain with genomic selection in autotetraploid potato. *The Plant Genome* 9, 3. DOI: 10.3835/plantgenome2016.02.0021

Solano, J., Acuña, I., Esnault, F., Brabant, P. (2014) Resistance to Phytophthora infestans in Solanum 639 tuberosum landraces in Southern Chile. *Tropical Plant Pathology* 39, 307–315. DOI: 640 10.1590/S1982-56762014000400005

Spooner, D.M., McLean, K., Ramsay, G., Waugh, R., Bryan, G.J. (2005). A single domestication for potato based on multilocus amplified fragment length polymorphism genotyping. *Proceedings of the National Academy of Sciences of the United States of America* 102(41), 14694–14699. DOI: 10.1073/pnas.0507400102

Srivastava, A., Bhardwaj, V., Singh, B.P., Khurana, S.M.P. (2016) Potato diversity and its genetic enhancement. In: Rajpal, V.R., Rao, S.R., Raina, S.N. (eds) Gene Pool Diversity and Crop Improvement, Sustainable Development and Biodiversity. Springer International Publishing, Switzerland, pp. 187–226. DOI: 10.1007/978-3-319-27096-8

Stewart HE, Bradshaw JE, Wastie RL (1994). Correlation between resistance to late blight in foliage and tubers in potato clones from parents of contrasting resistance. *Potato Research* 37, 429–434. DOI: 10.1007/BF02358357

Stich, B., Van Inghelandt, D. (2018). Prospects and potential uses of genomic prediction of key performance traits in tetraploid potato. *Frontiers in Plant Science* 9, 159. DOI: 10.3389/fspl.2018.00159

Sverrisdóttir E., Byrne, S., Sundmark, E.H.R. *et al.* (2017). Genomic prediction of starch content and chipping quality in tetraploid potato using genotyping-by-sequencing. *Theoretical and Applied Genetics* 130, 2091–2108. DOI: 10.1007/s00122-017-2944-y

Sverrisdóttir, E., Sundmark, E.H.R., Johnsen, H.Ø. *et al.* (2018). The value of expanding the training population to improve genomic selection models in tetraploid potato. *Frontiers in Plant Science*. 9, 1118. DOI: 10.3389/fpls.2018.01118

van der Vossen, E., Sikkema, A., Hekkert B.L. *et al.* (2003). An ancient R gene from the wild potato species *Solanum bulbocastanum* confers broad-spectrum resistance to *Phytophthora infestans* in cultivated potato and tomato. *The Plant Journal* 36(6), 867–882. DOI: 10.1046/j.1365-313X.2003.01934.x

van Eck, H.J., Jacobs, J.M.E., van den Berg, P.M.M.M., Stiekema, W.J., Jacobsen, E. (1994). The inheritance of anthocyanin pigmentation in potato (*Solanum tuberosum* L.) and mapping of tuber skin colour loci using RFLPs. *Heredity* 73, 410–421. DOI: 10.1038/hdy.1994.189

van Eck, H.J., Jacobs, J.M.E., van Dijk, J., Stiekema, W.J., Jacobsen, E. (1993). Identification and mapping of three flower colour loci of potato (*S. tuberosum* L.) by RFLP analysis. *Theoretical and Applied Genetics* 86, 295–300.

Vos, P. (2016). Development and application of a 20K SNP array in potato. PhD thesis, Wageningen University, Wageningen, NL, pp. 9-12.

Vos, P. G., Utidewilligen, J. G. A. M. L., Voorrips, R. E., Visser, R. G. F., van Eck, H. J. (2015). Development and analysis of a 20K SNP array for potato (*Solanum tuberosum*): an insight into the breeding history. *Theoretical and Applied Genetics* 128, 2387–2401. DOI: 10.1007/s00122-015-2593-y

Vos, P., Hogers, R., Bleeker, M. *et al.* (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* 23(21), 4407–4414. DOI: 10.1093/nar/23.21.4407

Wang, Y., Rashid, M.A.R., Li, X., et al. (2019). Collection and evaluation of genetic diversity and population structure of potato landraces and varieties in China. *Frontiers in Plant Science* 10, 139. DOI: 10.3389/fpls.2019.00139.

Welsh, J., McClelland, M. (1990). Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research* 18(24), 7213–7218. DOI: 10.1093/nar/18.24.7213

Wenzl, P., Carling, J., Kudrna, D. *et al.* (2004). Diversity arrays technology (DArT) for whole-genome profiling of barley. *Proceedings of the National Academy of Sciences* 101(26), 9915–9920. DOI: 10.1073/pnas.0401076101

Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A., Tingey, S.V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* 18(22), 6531–6535. DOI: 10.1093/nar/18.22.6531

Yang, W., Kang, X., Yang, Q., Lin, Y., Fang, M. (2013). Review on the development of genotyping methods for assessing farm animal diversity.

Yoshida, K., Schuenemann, V.J., Cano, L.M. *et al.* (2013). The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine. *eLife* 2, e00731. DOI: 10.7554/eLife.00731

Zabeau M., Vos P. (1993). Selective restriction fragment amplification: a general method for DNA fingerprinting. *EP Patent. publication. No.* 1993, 0534858:B2.

Zaheer, K., Akhtar, M.H. (2016). Potato production, usage, and nutrition – a review. *Critical Reviews in Food Science and Nutrition* 56(5), 711–721. DOI: 10.1080/10408398.2012.724479

Zhang, A., Wang, H., Beyene, Y., Semagn, K., Liu, Y., Cao, S., et al. (2017). Effect of trait heritability, training population size and marker density on genomic prediction accuracy estimation in 22 bi-parental tropical maize population. *Frontiers Plant Science* 8, 1916. DOI: 667 10.3389/fpls.2017.01916

Zhang, R. (2004). Genetic characterization and mapping of partial resistance to early blight in diploid potato. Ph.D. Thesis, The Pennsylvania State University, University Park, PA, USA.

Zhao, X., Jayarathna, S., Turesson, H. *et al.* (2021). Amylose starch with no detectable branching developed through DNA-free CRISPR-Cas9 mediated mutagenesis of two starch branching enzymes in potato. *Scientific Reports* 11, 4311. DOI: s41598-021-83462-z

Zhu, C., Gore, M., Buckler, E.S., Yu, J. (2008). Status and prospects of association mapping in plants. *The Plant Genome* 1, 1. DOI: 10.3835/plantgenome2008.02.0089

Popular science summary

The adhesive agent in the glue that holds together the pages of this book may very well be produced from potato starch. Adhesive agents are just one of a number of end uses of potato, which unsurprisingly to most of us, also includes some of the world's most popular foods such as curries, gratins, chips or crisps. This multitude of end use markets, together with the potato's complex genome, makes potato breeding a time consuming and difficult endeavour.

Potato breeding is the art and science of producing new and improved cultivars of potato with traits sought after by farmers and consumers. Genomic-led breeding methods are based on DNA markers and may aid plant breeders in making better-informed decisions for improvement of their breeding material and speed up the process of cultivar development. In this thesis, resistance to the important potato disease early blight is mapped to the potato genome, which in the future may help develop specific DNA markers to be used as an aid in selections towards potato cultivars with increased resistance to early blight. We have also studied the implementation of a breeding method called genomic selection, which has successfully been applied for breeding of some animals and crops, for potato breeding. The results show that although genomic selection may be used to predict the performance of potato plants, using it across cycles of selection in a breeding program may prove difficult. Lastly, we explored the genetic diversity of potato cultivars stored at the Nordic genebank. When the potatoes from the gene bank are compared to the breeding clones of the potato breeding program at SLU, it becomes apparent that the material is very closely related and that Nordic potatoes, just like European potatoes in general, seem to stem from a very narrow genetic base.

Populärvetenskaplig sammanfattning

En av ingredienserna i limmet som håller samman sidorna av denna bok kan mycket väl vara stärkelse utvunnet ur potatis. Stärkelse för limproduktion är bara en av flera slutgiltiga användningsmål för potatis, som föga förvånande för de flesta av oss, även inkluderar några av världens mest populära måltider som currys, gratänger, pommes frites och chips. Denna stora mängd av potentiella användningsmål, samt potatisens komplexa arvsmassa, bidrar till att potatisförädling är en tidskrävande och svår uppgift.

Potatisförädling är konsten och vetenskapen bakom framtagandet av nya, förbättrade potatissorter med egenskaper som är eftertraktade av odlare och konsumenter. Genomiskt baserade förädlingsmetoder, som baseras på användandet av DNA-markörer, kan hjälpa växtförädlare att ta bättre informerade beslut angående förbättring av förädlingsmaterial och skynda på framtagande av nya sorter. I denna avhandling presenteras den genetiska kartläggningen av resistens mot potatissjukdomen torrfläckssjuka, vilket i framtiden kan bidra till att utveckla specifika DNA-markörer som kan användas som hjälpmedel när urval av nya sorter med ökad resistens mot denna sjukdom ska tas fram. Vi har även studerat om en förädlingsmetod som kallas genomisk selektion, vilken tidigare framgångsrikt kunnat implementeras vid viss djuravel och förädling av andra växter, kan implementeras i potatisförädling. Våra resultat pekar mot att även om genomisk selektion kan användas för att förutspå potatisplantans egenskaper så är det svårt att använda den som selektionsmetod över flera förädlingsgenerationer. Slutligen har vi studerat den genetiska diversiteten bland potatissorter som förvaras hos den nordiska genbanken. När potatisen från genbanken jämförs med förädlingsmaterialet från SLUs potatisförädlingsprogram blir det uppenbart att materialen är mycket nära

besläktade och att de, precis som Europeiska potatissorter i allmänhet, verkar stamma från en mycket begränsad genetisk bakgrund.

Acknowledgements

The results in these studies demanded a lot of work, but thankfully I was not alone in creating them. Many people have helped me throughout my PhD studies, and I am looking forward to inviting you all to a drink and a meal, however I cannot resist to thank a few of them here as well.

First of all, my main supervisor Rodomiro Ortiz. You have allowed me to express myself freely and you have always showed that you believe in me and value my opinions. Your passion for science is contagious and I will greatly miss dropping by your office for spontaneous lessons and discussions. There has been no problem too small or too big to discuss with you, and solutions have always been close at hand. I am truly grateful for having you as my supervisor.

I was also given two extraordinary co-supervisors – Mariette Andersson and Aakash Chawade. Mariette, thank you for all of your support and always making me feel welcomed and included at the department. I have always felt blessed among PhD students to have you as my co-supervisor. Aakash, I am very happy that you agreed to join my supervisor team in a time of great need. You have treated me with patience and kindness when I came to Alnarp with my very limited experience in programming and omics.

I am truly grateful to Fredrik Reslow for everything you have helped me with. If it had not been for you, I would never had returned after my very first day in the field in 2016 when I was tasked to phenotype two thousand T_1 plants for late blight symptoms invisible to anyone's eyes except yours. Thank you for all of your help during these years, and for all the interesting discussions we have had. You have taught me more than you think!

Great thanks to all of my co-authors, Firuz Odilbekov, Erland Liljeroth, Paulino Pérez-Rodríguez, Ulrika Carlson-Nilsson and Pawel Chrominski, who have helped me with statistics, collecting data, writing, reading and re-reading manuscripts while being very friendly and encouraging. I am very happy to have had the opportunity to work with you.

I would also like to thanks Alf Ceplitis and Bo Gertsson at Lantmännen for taking me on as a mentee through the PlantLink mentorship program. It has been a privilege to have to such experienced plant breeders showing interest in my research from the very beginning of my PhD studies. Thank you for taking me under your wing and being so generous in showing me your work.

To my past and present office mates – Therése Bengtsson, Emelie Ivarson, Selvaraju Kanagarajan, Ibrahim Juma, Sjur Sandgrind, Jonas Skytte Af Sätra and Fluturë Novakazi – thank you for your support and for creating such a pleasant and welcoming office environment.

I have enjoyed my time as a PhD student in Alnarp a lot, but I have to admit that it is not only thanks to scoring such an interesting research project. My friends Anna-Lovisa, Antonio, Emilia, Maja, Rita, Sophie and several other co-workers has made this one of the most fun times of my life. You have offered me tonnes of support and cheered me on when I have needed it the most. Thank you for always providing me a shoulder to cry on at roughest of times, or a drink to cheer with at the easiest of times.

I would like to thank my family for all of the care they have shown me. Especially my mom Daina, without your help these last few months this book would not have been written.

Lastly, I would like to thank Alexander. Meeting you in the heart of Alnarp (Bobo's heart) has truly changed my life. You are my greatest cheerleader and the kindest and funniest person I have ever met. Thanks to you and Julian for all of your love and patience. You are the best out of all things that I will be bringing with me from my time in Alnarp!

Acta Universitatis agriculturae Sueciae
Doctoral Thesis No. 2021:80

Most cultivated potatoes are polyploids with tetrasomic inheritance, which makes breeding for new improved cultivars more complicated compared to diploid crops. In this thesis, we explore genomic-led breeding methods for tetraploid potato. These methods have the potential to speed up the breeding process and may help potato breeders to made better-informed crosses for germplasm improvement.

Catja Selga received her graduate education at the Department of Plant Breeding, Swedish University of Agricultural Sciences (SLU), Alnarp. She received her MSc in Biology from Lund University, Lund.

Acta Universitatis agriculturae Sueciae presents doctoral theses from the Swedish University of Agricultural Sciences (SLU).

SLU generates knowledge for the sustainable use of biological natural resources. Research, education, extension, as well as environmental monitoring and assessment are used to achieve this goal.

Online publication of thesis summary: https://pub.epsilon.slu.se

ISSN 1652-6880
ISBN (print version) 978-91-7760-837-0
ISBN (electronic version) 978-91-7760-838-7