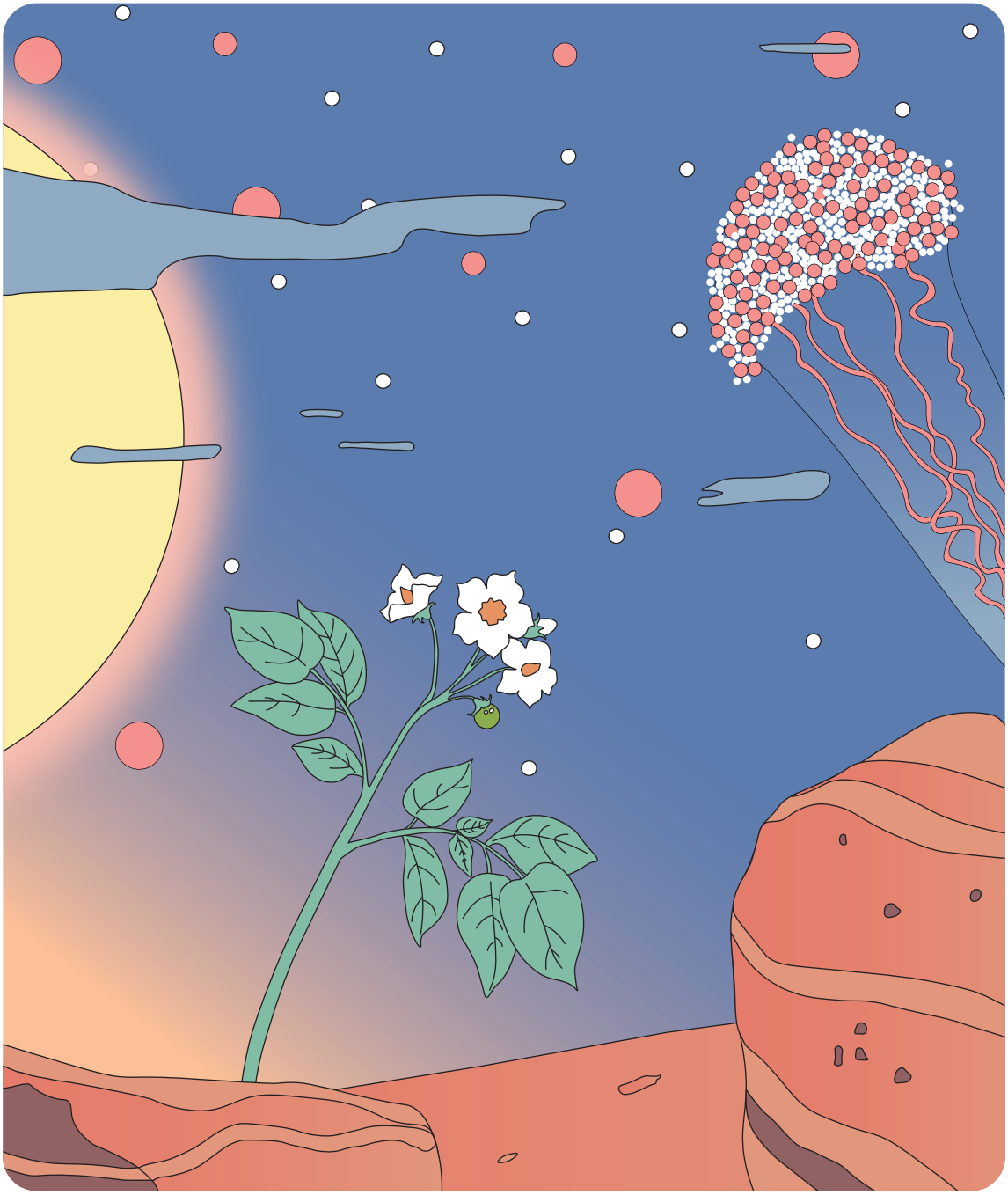

Natural variation in potato sexual reproduction facilitates breeding

Corentin Clot

Propositions

1. Transposons are potato breeders' best friends.
   (this thesis)

2. The value of interploidy breeding in potato will only diminish once diploid F_1 hybrid breeding has reached maturity.
   (this thesis)

3. Reflective solitary walks are an undervalued component of scientific training.

4. The game design principle "Fail Fast and Follow the Fun" is a good motto for scientific research.

5. Traditional folk tunes evolve like organisms do.

6. Dissociating meat consumption from masculinity is an environmental necessity.

7. Determinism is the only antidote against the toxicity of meritocracy.

Propositions belonging to the thesis, entitled
"Natural variation in potato sexual reproduction facilitates breeding"

Corentin Clot
Wageningen, 28 August 2023

Natural variation in potato sexual reproduction facilitates breeding

Corentin Clot

Thesis committee

Promotor

Prof. Dr. R.G.F Visser
Professor and chair of Plant Breeding
Wageningen University & Research

Co-promoter

Dr H.J. van Eck
Assistant professor, Plant Breeding
Wageningen University & Research

Other members

Prof. Dr. B.J. Zwaan, Wageningen University & Research, the Netherlands

Dr. H. Tai, Fredericton Research and Development Centre, Canada

Dr. E. Jongedijk, KWS Group, Einbeck, Germany

Dr. P. Lindhout, Solynta, Wageningen, the Netherlands

This research was conducted under the auspices of the Graduate School Experimental Plant Sciences.

Natural variation in potato sexual reproduction facilitates breeding

Corentin Clot

Thesis
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Monday 28 August 2023
at 1:30 p.m. in the Omnia Auditorium

Corentin Clot
Natural variation in potato sexual reproduction facilitates breeding, 182 pages

PhD thesis, Wageningen University, Wageningen, the Netherlands (2023)
With references, with summary in English

ISBN: 978-94-6447-752-8

DOI: 10.18174/633103

Table of contents

Chapter 1 **6**
General Introduction

Chapter 2 **20**
The origin and widespread occurrence of *Sli* based self-compatibility in potato

Chapter 3 **46**
High-density linkage map constructed from a skim sequenced diploid potato population reveals transmission distortion and QTLs for tuber yield and pollen shed

Chapter 4 **74**
Desynapsis in potato is caused by *StMSH4* mutant alleles and leads to either highly uniform unreduced pollen or sterility

Chapter 5 **94**
Identification of two mutant *JASON-RELATED* genes associated with unreduced pollen production in potato

Chapter 6 **122**
General discussion

References **143**

Summary **167**

Acknowledgements **171**

About the author **179**

Education statement **181**

Chapter 1

General introduction

Humans began to domesticate plants for food production around 12,000 years ago (Larson et al., 2014). This technological innovation, at the core of the Neolithic revolution, happened independently in many parts of the world. During this period, groups of hunter-gatherers formed sedentary agrarian societies which ultimately became the dominant form of human cultures (Olsson and Paik, 2016). Plant domestication is the result of both deliberate as well as unconscious artificial selection pressure set by early farmers on crop ancestors. This selection led to morphological and physiological modifications distinguishing crops from their wild relatives (Purugganan and Fuller, 2009). From Neolithic farmers to modern plant breeders, the incremental adaptation of cultivated plants to human needs has been a continuous process that is still ongoing. While the core principle of selecting superior individuals within a variable population has remained constant, the rediscovery of Mendelian genetics (Mendel, 1866) and the development of the infinitesimal model (Fisher, 1918) in the early 20th century has transformed the art of plant breeding into a science. In the meantime, the focus of plant breeders has expanded beyond simplifying cultivation and increasing yield to also include product quality and resistances to biotic and abiotic stressors. The latter two targets will be crucial to safeguard food production in the current context of climate change. Unfortunately, climate change is only one of the six planetary boundaries known to be exceeded by humanity. In fact, our food production systems contribute substantially to the transgressions of the biogeochemical flow (N and P), land-system change, biosphere integrity, novel entities, and green water boundaries (Persson et al., 2022; Steffen et al., 2015; Wang-Erlandsson et al., 2022). Hence, contemporary plant breeders must develop crop varieties that enable resilient food production while minimizing the impact of their cultivation on the biosphere.

To reach their goal, plant breeders may introgress valuable genes from crop wild relatives. However, already within the domesticated gene pool much can be achieved with the phenomena of transgressive segregation, whereby traits in the offspring fall outside of the range of the parents, and heterosis, whereby a hybrid outperforms its parents. Both phenomena are largely explained by the dispersal of favourable alleles between parents, which in the case of heterosis must also show directional dominance (Mackay et al., 2021). Thus, plant breeding boils down to reshuffling heritable variation into superior combinations. Following technological developments, the identification of superior allele combinations through phenotypical selection was supplemented by marker-assisted selection and more recently by genomic selection. On the other hand, the efficiency of reshuffling heritable variation was improved by gaining control over the reproductive biology of crops. An example of this is the F_1 hybrid system which facilitates the fixation of genetic gains through inbreeding and the exploitation of heterosis through outcrossing. F_1 hybrid breeding transformed maize production during

the previous century and promises to revolutionize potato breeding in the current one (Stokstad, 2019).

It is in this general context that this thesis was created. It aims to offer to breeders more control over the reproductive biology of potato with the hope that this will contribute to variety development with high yield in a variable climate while reducing the impact of potato production on the environment. In the remainder of this chapter, I will introduce the potato crop and highlight the limitations imposed by its reproductive biology on current breeding strategies. Then, I will introduce elements of angiosperm biology relevant to potato breeding with a focus on meiosis and mating systems. I will conclude by presenting the scope and outline of this thesis.

Potato

The breeding history of the potato crop

Although the origin of the cultivated potato (*Solanum tuberosum*) remains debated it is clear that potato was domesticated at least once in the Andes 8,000-10,000 years ago (Spooner et al., 2005a, 2005b; Sukhotu and Hosaka, 2006). The production of unreduced gametes by diploid potatoes (2n = 2x = 24) is arguably at the origin of the Andean tetraploid landraces (2n = 4x = 48) (Watanabe and Peloquin, 1989) which were introduced to the Canary Islands in 1567 (Hawkes and Francisco-Ortega, 1993). After their introduction to Europe, potatoes admixed with Chilean landraces which potentially contributed to their adaptation to long day conditions (Gutaker et al., 2019). By the end of the 18th century, potato had spread out from Europe and was cultivated in most of the world temperate climate zones. Potato breeding gained momentum in Europe and North America following the devastating late blight epidemics of the middle of the 19th century (Jansky and Spooner, 2018). This period coincided with the resurgence of Andean ancestry in historical European potato clones (Gutaker et al., 2019). To this day, biotic resistance is one of the main drivers of potato breeding and has resulted in a significant geneflow between potato wild-relatives and cultivars during the latter part of the 20th century (Vos et al., 2015). Overall, the potato genome has been shaped by introgression breeding, which further contributed to its already high level of heterozygosity (Hardigan et al., 2017; Hoopes et al., 2022). Today potato is the 3rd most important human food crop after wheat and rice and a staple food for 1.3 billion people worldwide (Stokstad, 2019). The potato market is divided in three segments: the fresh segment for direct consumption, the processing segment for French fries and crips and the industrial segment for starch and proteins extraction. For all three commercial segments, the production is dominated by highly heterozygous autotetraploid cultivars, vegetatively propagated through seed tubers.

Despite the economical and nutritional importance of this crop, genetic gains in potato have been limited compared to other major crops (Douches et al., 1996; Rijk et al., 2013); a fact highlighted by the lasting popularity of century old cultivars such as Bintje and Russet Burbank. This can be explained by 1) the fragmentation of the market which requires to breed for different and sometimes opposite ideotypes and by 2) the constrains set by potato reproductive biology on breeding strategies. Indeed, the combination of allogamy, heterozygosity and tetrasomic inheritance makes it challenging for breeders to reach fixation of favourable alleles and to remove deleterious ones. Although tetraploid potatoes can be selfed, this is uncommon in breeding programs due to the wish to combine traits from different parents and the fear of inbreeding depression. Therefore, conventional breeding relies on crossing heterozygous tetraploid parents to produce full-sib families. Because favourable alleles are not fixed in those parents, the resulting progeny will display a wide range of phenotypic segregants. Only a minor fraction of the offspring will display the desirable levels for quality traits and yield potential necessary to become a new variety or a progenitor clone. Importantly, the identification of those superior clones takes a considerable amount of time and space due to vegetative propagation. Indeed, the low multiplication rate of potato tubers combined with the destructive use of tubers for phenotyping and the plot size needed for reliable yield estimation led to a clonal selection cycle of 5-9 years (Gopal, 2015). With a single meiotic event per clonal selection cycle, the amount of allelic rearrangement over time is much lower for potato than for crops with an annual sexual reproduction. Consequently, some contemporary potato varieties are only seven meiosis away from Rough Purple Chili, a clone introduced to North America in 1851 and one of the founders of the contemporary potato germplasm (Love, 1999; van Berloo et al., 2007).

Alternative potato breeding approaches

Recognizing the limitation of conventional potato breeding, Chase (1963) proposed an alternative breeding approach based on ploidy manipulation. This so-called **analytical breeding** consists of three major steps: 1) ploidy reduction, 2) breeding and hybridisation with wild relatives at the diploid level and, 3) polyploidisation. To achieve step 1, dihaploids (2n = 2x = 24) can be routinely produced by pollinating tetraploid potatoes (2n = 4x = 48) with dihaploid inducer clones selected from diploid *S. tuberosum* group Andigenum (formerly *S. phureja*) (Hermsen and Verdenius, 1973). This is a laborious step and generally results in dihaploids exhibiting poor vigour and fertility due to inbreeding depression (Hutten et al., 1994b). During step 2, dihaploids can be crossed with wild relatives to improve their fertility and introgress traits of interest such as disease resistances. Recurrent selection and genetic analysis can then be performed more easily at the diploid level. Finally, during step 3, selected diploids are brought back to the tetraploid level through somatic genome doubling or sexual

polyploidisation. Since, increasing ploidy without adding new genetic variation does not lead to superior cultivars (Aversano et al., 2015; Uijtewaal et al., 1987a), sexual polyploidisation by means of unreduced gametes is preferred. Promising yields were observed in tetraploid progenies derived from unilateral (4x × 2x) and bilateral (2x × 2x) sexual polyploidisation (Hutten et al., 1994a; Kidane-Mariam and Peloquin, 1975; Mendiburu and Peloquin, 1971). Although many progenitor clones and several varieties such as Yukon Gold, Krantz, and Abbot were obtained through sexual polyploidisation, the analytical breeding approach has not been widely adopted by commercial potato breeders. This can be explained by the difficulty of obtaining fertile dihaploids, the elusive genetic basis of unreduced pollen production and the cost of developing and maintaining a diploid germplasm of agricultural value in addition to a tetraploid one (Jansky and Spooner, 2018). Other alternative breeding approaches such as somatic fusion of diploids (Wenzel et al., 1979) or the pursuit of apomixis (Hermsen, 1980) similarly failed to deliver.

In the late seventies, early eighties, the usage of varieties grown from potato botanical seeds, also known as true potato seeds (TPS), was advocated by the International Potato Center (CIP) to facilitate both the multiplication and transport of starting material, but also the management of diseases (Almekinders et al., 2009; Mendoza, 1985). This technology, (from now on termed **CIP-TPS**) was primarily targeted towards Global South countries where seed tubers are either imported at high costs or produced locally through informal systems which facilitate the accumulation of diseases (Gildemacher et al., 2009; Thomas-Sharma et al., 2016). CIP-TPS varieties were generally produced by intercrossing two tetraploid clones which resulted in a progeny lacking uniformity. Alternatively, CIP-TPS varieties could be produced by crossing a tetraploid female parent with a diploid male parent producing unreduced gametes. This promising strategy was reported to improve the progeny uniformity (Hutten et al., 1994a; Mendoza, 1985). Nonetheless, the limited agronomical knowledge of TPS cultivation and the lack of uniformity of TPS varieties when compared with seed tuber varieties limited the adoption of this technology.

Diploid F_1 hybrid breeding is another alternative potato breeding approach that has gained popularity in recent years. This approach relies on the development of diploid inbred lines and their subsequent hybridisation to produce uniform F_1 hybrid varieties. Those varieties can be propagated through TPS and benefit from aforementioned advantages inherent to this technology without suffering from a lack of uniformity. In addition, F_1 hybrid breeding enables the fixation of genetic gains and the removal of detrimental alleles during the inbreeding step. Inbreeding is enabled by the identification of a dominant S-locus inhibitor gene (*Sli*) conferring self-compatibility to the normally self-incompatible diploid potatoes (Hosaka and Hanneman, 1998a, 1998b). *Sli* was used to generate the first diploid potato inbred lines via repeated

selfing (Phumichai and Hosaka, 2006; Phumichai et al., 2006). The first proof of concept that F_1 hybrid breeding was feasible in potato came from the Netherlands a few years later (Lindhout et al., 2011). The potential of F_1 hybrid breeding was soon recognized in China (Li et al., 2013) and in North American (Jansky et al., 2016), starting a global effort to convert the highly heterozygous and tetraploid potato into an inbred-line-based diploid crop. Inbreeding depression is a major obstacle to this conversion and recent studies have highlighted the massive breeding and genomics efforts needed to overcome this problem in only a few breeding lines (Zhang et al., 2019; Zhang et al., 2021a).

In this thesis, I describe research that contributes ingredients to yet another breeding method, **Fixation-Restitution breeding**, which aims to solve the limitations of conventional potato breeding but also of F_1 hybrid breeding. To do so, this method simply combines Chase analytical breeding with self-compatible diploid clones. As in F_1 hybrid breeding, self-compatible diploids clones are used in rapid backcrossing schemes to accumulate favourable alleles. As in analytical breeding, those partial inbred lines can be hybridised to produce superior diploids which can then transfer their genetic make-up to the tetraploid level via sexual polyploidisation. Fixation-Restitution breeding, circumvent the need to reach the genome wide homozygosity, essential for F_1 hybrid uniformity, and thus limit the burden of inbreeding depression. It also bypasses the need to develop heterotic groups of diploid inbred lines by exploiting the current tetraploid elite germplasm in interploidy crosses.

Aspects of angiosperm reproductive biology relevant for potato breeding

Mating systems

Although most flowering plants are simultaneous hermaphrodites and could therefore self-fertilize, nearly half of them (including diploid potato) are allogamous due to the presence of self-incompatibility systems (Renner and Ricklefs, 1995). Self-incompatibility is an intraspecific reproductive barrier that prevent self-fertilization by allowing the pistil to discriminate between self- and non-self-pollen. This discrimination is generally mediated by a male and a female determinant encoded at a single genetic locus (S-locus) by multiallelic genes (Takayama and Isogai, 2005). In diploid potato, like in other gametophytic self-incompatible species from the *Solanaceae*, *Rosaceae*, and *Plantaginaceae* families, the female determinant encodes for cytotoxic S-RNases that inhibit pollen tube growth in the style (Hua et al., 2008; Lee et al., 1994; Murfett et al., 1994). The male determinant encodes for a combination of S-locus F-box proteins (SLF) that can mediate the ubiquitination of some S-RNases (Kubo et al., 2015; Sijacic et al., 2004). According to the collaborative non-self-

recognition model (Kubo et al., 2010), the combined action of the SLFs encoded by one S-haplotype results in the recognition and degradation of all S-RNases except of the one encoded by their own haplotype. Hence, self-incompatibility occurs when the S-haplotype of the pollen matches either of the two S-haplotypes of the style. This elegant model also provides an explanation for the phenomenon of competitive interaction that allows tetraploid potatoes to become self-compatible (Olsder and Hermsen, 1976). When tetraploid clones produce diploid heteroallelic pollen that expresses two complementary sets of S-locus F-box proteins (SLFs), all S-RNases are ubiquitinated, which breaks down gametophytic self-incompatibility.

Recently, self-compatible diploid potato clones were engineered by a knock-out of the gene encoding for the S-RNase (Enciso-Rodriguez et al., 2019; Ye et al., 2018). Furthermore, a self-compatible diploid clone bearing a naturally weakly expressed S-RNase was recently identified (Zhang et al., 2019). Other cases of self-compatibility among *S. tuberosum* dihaploids were reported in earlier studies (De Jong and Rowe, 1971; Olsder and Hermsen, 1976), but the genetic and mechanistic basis of their self-compatibility remained undetermined prior to this thesis. Despite those alternative options, the most widely used source of self-compatibility is the dominant gene *Sli* originally identified from a *S. chacoense* accession and located on chromosome *12* (Hosaka and Hanneman, 1998a, 1998b). The causal gene behind *Sli* was recently shown to encode for a pollen-expressed F-box protein functioning as a general S-RNase inhibitor (Eggers et al., 2021; Ma et al., 2021).

Meiosis

Meiosis is a specialized cell division found in sexually reproducing organisms that ultimately leads to the production of four haploid gametes out of a single diploid mother cell. This is achieved via two successive divisions, meiosis I and meiosis II, following a single S-phase. During meiosis II sister chromatids are separated in a division resembling a haploid mitosis (Mercier et al., 2015). In contrast, meiosis I is characterized by the formation of crossovers and the separation of homologous chromosomes (Zickler and Kleckner, 2015). Meiosis is also responsible for the reassortment of allelic variation over generations, making it a major source of genetic diversity upon which both natural and artificial selection can act (Mercier et al., 2015).

Meiotic Recombination

The process of meiotic recombination is initiated during prophase I by the formation of DNA double-strand breaks (DSBs). DSBs are induced by a complex of proteins that includes the highly conserved SPO11 (de Massy, 2013). DSBs occur in large numbers per mother cell (e.g. ~200 in *Arabidopsis thaliana* (Sun et al., 2012) and ~2000 in wheat (Gardiner et al., 2019)) and need to be repaired to preserve genome integrity. The repair mechanism starts with the resection of DNA strands resulting in 3’ single-

strand overhangs (Neale et al., 2005). One of those overhangs may then invades a sister or non-sister chromatid to form a D-loop (Lambing et al., 2017). Non-sister chromatid invasion leads to the formation of recombination intermediates that can be resolved either as crossovers (CO) or as non-crossovers (NCO) depending on the pathways involved. In addition, COs can be sub-divided into two classes each generated by their own pathways (Mercier et al., 2015). In the ZMM pathway (an acronym for the *Saccharomyces cerevisiae* proteins Zip1-4, Msh4-5 and Mer3), double Holliday junction intermediates are formed and resolved as class I COs. Class I COs represent 75-85% of COs in *A. thaliana* and are sensitive to interference, meaning that the occurrence of a CO inhibits the formation another CO nearby (Higgins et al., 2004; Mercier et al., 2005). In the second pathway, dependant on the MUS81 protein, unknown recombination intermediates are resolved to form class II COs. Class II COs account for the remain 15-25% of COs and are interference insensitive (Berchowitz et al., 2007). Finally, NCOs results from the dissolution of double Holliday junction and MSU81 recombination intermates, and from a third pathway named synthesis dependent strand annealing (SDSA). In the SDSA pathway, the strand invasion is rejected, and the damaged chromosome is repaired via DNA synthesis using the non-sister chromatid as template. Overall, only a minority of DSBs are repaired as COs leading to 1-3 COs per bivalent in most eukaryotes (Mercier et al., 2015). The presence of at least one CO per bivalent, known as obligate CO, is essential to guarantee that the connexion between homologous chromosomes is maintained until anaphase I. This CO provides a counterforce to the pole directed spindle forces ensuring the proper positioning and segregation of chromosomes during meiosis I (Page and Hawley, 2003).

The ability to manipulate meiotic recombination is a breeder's dream. On the one hand, increasing the frequency of meiotic recombination can speed up plant breeding. With several COs per chromosome, breeders could obtain in a single generation a combination of multi-recombinant haplotypes that would otherwise take several generations to create. On the other hand, decreasing the frequency of meiotic recombination would allow breeders to preserve previously selected allelic combinations. For example, the reverse breeding strategy relies on the reduction or the elimination of COs to obtained gametes with non-recombinant parental chromosomes from which doubled haploids can be regenerated (Calvo-Baltanás et al., 2020; Dirks et al., 2009). Similarly, apomeiosis, a phenotype in which meiosis is replace by a mitosis-like division, was engineering in *A. thaliana* and rice by eliminating recombination, modifying chromatid segregation and supressing the second meiotic division (d'Erfurth et al., 2009; Mieulet et al., 2016). This is regarded as an important step towards the introduction of apomixis (i.e. clonal reproduction through seeds) in crops, a feat which would revolutionize plant breeding and agriculture by allowing the fixation of heterosis (Conner and Ozias-Akins, 2017). In potato, an utopian breeding scheme replacing vegetative propagation with elements of apomixis was proposed to deliver uniform and

vigorous TPS varieties (Hermsen, 1980; Hermsen et al., 1985). This scheme relies on natural potato mutants with a reduced frequency of recombination, originally referred to as desynaptic (Jongedijk and Ramanna, 1989, 1988). Lack of crossovers results in a chaotic first meiotic division and the formation of unbalanced gametes. However, when combined with another meiotic alteration that restitutes the first meiotic division, these mutants can produce balanced unreduced gametes that preserve most of the parental heterozygosity (Jongedijk et al., 1991a, 1991b).

Ploidy Reduction

One of the crucial roles of meiosis is to maintain ploidy level across generations. This is achieved by halving the somatic number of chromosomes during gamete production. When fertilization occurs, the gametes fuse, restoring the original ploidy level in the offspring. However, the production of unreduced gametes with a somatic (2n) rather than a gametophytic (n) number of chromosomes has been observed in a wide variety of plant species (Harlan and de Wet, 1975; Kreiner et al., 2017a; Ramanna and Jacobsen, 2003). When those unreduced gametes participate in fertilization, polyploid offspring are produced in a process known as sexual polyploidisation (Palumbo et al., 2021). Unreduced gametes can result from a variety of cytological events that include pre- or post-meiotic genome doubling and several meiotic restitution mechanisms (de Storme and Geelen, 2013). Meiotic restitution mechanisms are caused by a combination of environmental and genetic factors (de Storme and Geelen, 2014, 2013). Depending on the mechanism at play, unreduced gametes can have a drastically different genetic composition. Unreduced gametes that are composed of homologous chromosomes preserve most of the parental heterozygosity and are genetically equivalent to First Division Restitution (FDR). In contrast, unreduced gametes that are composed of sister chromatids have a higher degree of homozygosity and are genetically equivalent to Second Division Restitution (SDR) (Ramanna and Jacobsen, 2003). In species that undergo a simultaneous cytokinesis, the most common mechanism of meiotic restitution is the mis-orientation of meiosis II spindles (Bretagnolle and Thompson, 1995). This mechanism brings the non-sister chromatids, initially separated in MI together, effectively equivalent to FDR. In *A. thaliana*, two genes, *AtPS1* (d'Erfurth et al., 2008) and *AtIAS* (de Storme and Geelen, 2011; Erilova et al., 2009), are known to specifically regulate the orientation of meiosis II spindles. Knocking out either of these genes results in the formation of FDR unreduced pollen. On the other hand, the production of SDR unreduced gametes is generally associated with a failure of entering meiosis II. This phenotype is observed in *A. thaliana* loss-of-function mutants of the cyclin TAM and of the protein OSD1 (d'Erfurth et al., 2010; d'Erfurth et al., 2009).

Unreduced gametes are valuable to plant breeders, who use them to link germplasm of different ploidy levels and to create new polyploids (Ramanna and Jacobsen 2003;

Younis et al. 2014). Specifically, disease resistances have been transferred from diploid wild-relatives of potato into tetraploid cultivars via unreduced gametes (Ortiz *et al.* 1994; Capo *et al.* 2002). From a heterosis standpoint, potato breeders prefer FDR unreduced gametes as they preserve ~80% of parental heterozygosity while SDR unreduced gametes only retain ~40% (Douches and Quiros, 1988a, 1988b; Jongedijk et al., 1991a; Peloquin et al., 2008). As a result, progenies derived from FDR unreduced gametes show a significant yield increase compared to those derived from SDR unreduced gametes (Hutten et al., 1994a; Kidane-Mariam and Peloquin, 1975; Mendiburu and Peloquin, 1977). In potato three meiotic restitution mechanism leading to unreduced gamete formations have been identified, one involving the mis-orientation of meiosis II spindles, and two involving premature cytokinesis (Mok and Peloquin, 1975a). Although all three mechanisms were initially described as independent monogenic and recessively inherited traits (Mok and Peloquin, 1975b), the validity of Mendelian interpretations of variable levels of unreduced gametes was questioned by Hermsen et al. (1985) showing that the thresholds used for phenotypic classification were ambiguous. Subsequent studies suggested a more complex genetic regulation of unreduced pollen production in potato (Dongyu et al., 1995; Ortiz and Peloquin, 1992).

Scope and outline of the thesis

The main objective of this thesis was to investigate the components required to enable the Fixation-Restitution breeding strategy. This strategy makes use of self-compatible diploid clones that can produce unreduced pollen to combine the ease of fixating genetic gains at the diploid level with the heterozygosity offered by tetraploidy. To achieve this goal, this thesis focuses on unravelling the genetic basis of self-compatibility, crossover reduction, and unreduced pollen production in potato.

In **Chapter 2**, we studied the inheritance of self-compatibility in two diploid potato bi-parental populations. Using a *k*-mer based bulk segregant analysis, we mapped self-compatibility on chromosome *12* in both populations. We further used those *k*-mers for haplotype mining and realised that the self-compatible haplotype we identified was indistinguishable from the *Sli* haplotype, originally described in *Solanum chacoense*. To our surprise, this haplotype was abundantly present among tetraploid *S. tuberosum* cultivars where it was hiding in plain sight due to the breakdown of gametophytic self-incompatibility at the tetraploid level. Finally, we developed a set of KASP markers enabling marker-assisted selection of self-compatibility.

In **Chapter 3**, we skim-sequenced a large bi-parental diploid potato population and used the resulting data to create high-density linkage maps. Those parental maps highlighted differences in male and female recombination rate and exposed structural variations, notably a 5.8 Mb inversion on chromosome *3*. We then focussed on transmission ratio distortion, a potential indicator of major deleterious mutations. We

identified 10 loci subject to transmission ratio distortion and four pairs of loci involved in non-independent assortment of alleles. Finally, we performed a QTL analysis for seedling tuber yield in pots and for pollen shed. Tuber yield in pots was essentially explained by a major-effect QTL colocalising with *StCDF1*. We also identified five minor-effect QTLs epistatic to *StCDF1* allelic combinations. For pollen shed, a crucial element of male fertility, we identified a total of seven QTLs including a major-effect locus located on chromosome *2*.

In **Chapter 4**, we investigated the inheritance of desynapsis, a meiotic aberration in which maintenance of homologous chromosome pairing is compromised due to a strong reduction in crossover numbers. This results in few bivalents, many univalent and unbalanced chromosome segregation during meiosis I. In this thesis we use the historical term desynapsis to describe this phenomenon, albeit the term chiasmata shortage may be less ambiguous. In modern literature desynapsis is more commonly used to refer to the withdrawal of the synaptonemal complex during diplotene which is part of the normal meiotic program (Dix et al., 1997; Guan et al., 2020). Exploiting the segregation of both desynapsis and FDR unreduced pollen within the population genotyped in Chapter 3, we used pollen microscopy to identify individuals with unstainable haploid and stainable unreduced pollen to classify synaptic and desynaptic individuals. Next, we ran a QTL analysis and identified a single causal locus on the short arm of chromosome *8*. Within this region, the most likely candidate gene was *StMSH4*, an essential component of the class I crossover pathway. We discovered that an exonic 7 bp insertion in *StMSH4* was associated with desynapsis in our mapping population. Mining the growing number of potato assemblies, we discovered another allele with a 3820 bp exonic insertion and confirmed that both alleles cannot complement each other. We subsequently showed that non-functional *StMSH4* alleles are present in many European cultivars and discuss the opportunities and limitations offered by desynapsis in the context of potato breeding.

In **Chapter 5**, we studied the mechanism of formation and the inheritance of unreduced pollen in three diploid potato bi-parental populations. In our three populations, the level of unreduced pollen production segregated quantitatively, and this variation was mainly due to genotypic effects. We identified three major-effect and two minor-effect QTLs regulating this trait. Two major-effect QTLs colocalised with genes showing homology to *AtJAS*, a key regulator of meiosis II spindles orientation in *A. thaliana*. Those candidate genes were consistent with the cytological observations of mis-oriented metaphase II plates detected in some of the parental clones of our populations. Finally, we identified sequence variants expected to affect the functionality of those candidate genes, specifically an exonic transposon insertion in the first candidate genes and an amino acid substitution within a highly conserved domain in the second one. Overall, this chapter sheds light on the natural variation

associated with unreduced pollen production and will facilitate interploidy breeding in potato.

In **Chapter 6**, I discussed the main findings of this thesis in the context of potato breeding. I exposed how the mutations identified in this thesis can be combined, in the context of Fixation-Restitution breeding, to improve potato breeding efficiency. I argued that this can be achieved without relying on fully homozygous inbred lines and thus circumvent the challenges of inbreeding depression. I also reflected on the forward genetic strategy employed throughout this thesis and proposed that adopting a "translational genomics" approach could accelerate the identification of natural mutations affecting potato sexual reproduction.

Chapter 2

The origin and widespread occurrence of *Sli* based self-compatibility in potato

Authors
Corentin R. Clot[1], Clara Polzer[1,3], Charlotte Prodhomme[1,4], Cees Schuit[2], Christel J.M. Englen[1], Ronald C.B. Hutten[1], Herman J. van Eck[1]

Affiliations
[1] Plant Breeding, Wageningen University and Research, Po Box 386, 6700 AJ Wageningen, Netherlands
[2] Bejo Zaden B.V., Trambaan 1, 1749 CZ Warmenhuizen, The Netherlands
[3] Current address: Aardevo B.V., Johannes Postweg 8, 8308 PB Nagele, The Netherlands
[4] Current address: La Fédération Nationale des Producteurs de Plants de Pomme de Terre (FN3PT), Agrocampus Ouest, UMR IGEPP, F-29260 Ploudaniel, France

Published in Theoretical and Applied Genetics (2020)
https://doi.org/10.1007/s00122-020-03627-8

Abstract

Self-compatible (SC) diploid potatoes allow innovative potato breeding. Therefore, the *Sli* gene, originally described in *S. chacoense*, has received much attention. In elite *S. tuberosum* diploids, spontaneous berry set is occasionally observed. We aimed to map SC from *S. tuberosum* origin. Two full-sib mapping populations from non-inbred diploids were used. Bulks were composed based on both pollen tube growth and berry set upon selfing. After DNA sequencing of the parents and bulks we generated *k*-mer tables. Set algebra and depth filtering was used to identify bulk-specific *k*-mers. Coupling and repulsion phase *k*-mers, transmitted from the SC parent, mapped in both populations to the distal end of chromosome *12*. Intersection between the *k*-mers from both populations, in coupling phase with SC, exposed a shared haplotype of approximately 1.5 Mb. Subsequently we screened read archives of potatoes and wild relatives for *k*-mers specific to this haplotype. The well-known SC clones US-W4 and RH89-039-16, but surprisingly, also *S. chacoense* clone M6 were positives. Hence, the *S. tuberosum* source of SC seems identical to *Sli*. Furthermore, the candidate region drastically reduced to 333 kb. Haplotype specific KASP markers were designed and validated on a panel of diploid clones including another renown SC dihaploid G254. Interestingly, *k*-mers specific to the SC haplotype were common in tetraploid varieties. Pedigree information suggests that the SC haplotype was introduced into tetraploid varieties via the founder 'Rough Purple Chili'. We show that *Sli* is surprisingly widespread and indigenous to the cultivated gene pool of potato.

Introduction

Despite potato being a major food crop, genetic gains for yield has been insignificant over the last century (Douches et al., 1996) while for maize yield increases of 1% per year have been achieved (Duvick, 2005). This lack of progress in breeding is embodied by century old varieties, such as Russet Burbank and Bintje, still widely grown today (Lindhout et al., 2011). Sexual reproduction of potato is characterized by polyploidy, tetrasomic inheritance of multiple alleles, and inbreeding depression upon selfing. Tetraploids can be selfed, but in conventional breeding this is uncommon. Hence, conventional potato breeding by crossing two highly heterozygous tetraploid clones is a numbers game, and the probability to find progeny with many combinations of beneficial alleles is small. Likewise, selection against mutational load in the germplasm is hardly feasible because deleterious recessive alleles are rarely exposed. Furthermore, clonal selection takes five to nine years (Gopal, 2015) due to the destructive use of tubers for phenotyping and the modest tuber reproduction rate. More importantly, with a single meiotic event per clonal selection cycle, the amount of allelic rearrangement over time is much lower than for crops with annual sexual reproduction (Jansky and Spooner, 2018). To overcome the limitations of conventional breeding, alternatives breeding schemes have been proposed. Analytical breeding, proposed by Chase (1963) involves ploidy reduction and breeding at the diploid level, followed by resynthesis of tetraploids. Recently, F_1 hybrid breeding at the diploid level has been proposed (Jansky et al., 2016; Lindhout et al., 2011), where inbred lines allow fixation of genetic gains and removal of unfavourable alleles. This requires self-compatible diploid germplasm tolerating inbreeding depression. However, while tetraploid potatoes are self-compatible, at the diploid level an S-RNase based gametophytic self-incompatibility system prevents inbreeding (McClure et al., 2011; Nettancourt, 1997).

S-RNase-based self-incompatibility (SI) is found among the *Rosaceae*, *Plantaginaceae* and the *Solanaceae* (McClure et al., 2011; Nettancourt, 1997). It is controlled by a single, highly polymorphic *S*-locus, encoding for a style specific S-RNase gene and several pollen-specific *S*-locus F-box (SLF) genes (Kao and Tsukamoto, 2004). Molecularly, SI is due to the cytotoxic effect of S-RNases inhibiting pollen tube growth in the style (Lee et al., 1994; Murfett et al., 1994). On the other hand, self-compatibility (SC) result from the SLF mediated ubiquitination of style S-RNases (Kubo et al., 2015; Sijacic et al., 2004). As predicted by the collaborative non self-recognition model (Kubo et al., 2010), each SLF allelic variant can specifically mediate ubiquitination of some of its non-self S-RNases and the combined action of the SLFs of one *S*-haplotype mediates degradation of all except its own S-RNases (Sun et al., 2018). Hence, SI occurs when the *S*-haplotype of the pollen matches either of the two *S*-haplotypes of the style. In potato, the *S*-locus is located on chromosome *1* (Gebhardt et al., 1991; Jacobs et al., 1995) within a region of low recombination,

consistent with the hypothesis that SI can only be maintained in the absence of recombination between S-RNases and SLFs (Kubo et al., 2015). Tetraploids are SC because their diploid pollen is heteroallelic. The expression of two different sets of SLFs enables mutual weakening or competitive interaction (Kubo et al., 2010) resulting in the ubiquitination of all S-RNases.

SC diploid clones can be engineered by the introduction of an extra *SLF* gene to induce competitive interaction, or by knock-out of the S-RNase (Enciso-Rodriguez et al., 2019; Ye et al., 2018) or other essential genes such as *HT* (Kondo et al., 2002). Alternatively, SC can be introgressed from natural mutants since SC clones have been spotted among several SI relatives of potato (Cipar et al., 1964). For instance Zhang et al. (2019) recently described the *S. stenotomum* landrace Huasa Amarilla (a.k.a. C151 or CIP 705468) as SC. The most well-known source conferring SC is the dominant *Sli* (*S*-locus inhibitor) gene originally identified in chc 525-3, a clone of the typically self-incompatible species *S. chacoense* (Hosaka and Hanneman, 1998a). Subsequent hybridisation between chc 525-3 and *S. phureja* was used to introduce SC in cultivated potato (Phumichai et al., 2006). Seven generations of inbreeding of another *Sli* bearing *S. chacoense* clone lead to the selection of the highly homozygous clone M6 (previously known as chc 523-3) (Jansky et al., 2014). SC is also found in *S. tuberosum* germplasm. So far, three clones have been described in the literature: US-W4 (de Jong and Rowe, 1971) and G254 (Olsder and Hermsen, 1976), dihaploids extracted from Minn. 20-20-34 and Gineke respectively, and the diploid clone RH89-039-16 (Peterson et al., 2016). Interestingly, in three independent mapping studies the distal end of chromosome *12* was associated with self-compatibility, or self-fertility: 1) the *Sli* gene of chc 525-3 (Hosaka and Hanneman, 1998b); 2) fruit set upon selfing in clone RH89-039-16 (Peterson et al., 2016); and 3) a major QTL for SC using clone CD-320-20 (Gardner et al., 2019). Clone CD-320-20 is derived from US-W4 (personal communication Dr. Walter S. De Jong, Cornell). Innovation of potato breeding would benefit from broadening the gene pool of self-compatible germplasm with good agronomical traits. In our diploid breeding program, we recognised plants with spontaneous berry set which could represent new sources of SC. Therefore, it seems prudent to characterize these additional SC sources from potentially different genetic backgrounds.

As whole genome sequencing (WGS) becomes more affordable, mapping by DNA sequencing of bulks has been applied successfully to identify the causal loci of various traits of interest (Schneeberger et al., 2009; Tribhuvan et al., 2018; D. Wu et al., 2018). The recently developed Comparative Subsequence Sets Analysis (CoSSA), based on sub-reads (*k*-mers) and set algebra, has proven its efficiency and robustness to identify haplotype specific SNPs linked to a potato wart resistance gene (Prodhomme et al., 2019). The use of pedigree information allows to confirm haplotype specificity of

sequence variants via clones identical-by-descent (IBD) and to trace back the origin of specific haplotypes. In this study, we used the CoSSA workflow to map SC in two segregating diploid potato populations. Subsequently, we developed haplotype specific KASP markers for marker assisted breeding. Finally, sequencing data was used to compare various SC clones and to traced back the origin of SC with the help of pedigree information.

Materials and Methods

Plant material

Two diploid *S. tuberosum* mapping populations were developed: IVP16-587 and IVP17-618, both segregating for SC (pedigree information shown in Suppl. Fig. 1). During previous growing seasons, spontaneous berry set was observed on IVP11-389-22, the female parent of population IVP16-587 and on IVP06-158-2 the male parent of population IVP17-618. Both mapping populations were sown in July 2018, and 100 vigorous seedlings per population were transplanted in five litre pots in an open ground greenhouse compartment with drip irrigation and wet pad-and-fan evaporative cooling system.

Phenotypic observations

Phenotypic data were collected on pollen fertility, female fertility, pollen tube growth in the style, berry set and seed set. Pollen fertility was examined by light microscopic estimating the percentage of stainable pollen using 0.2% acid fuchsine (1% fuchsine w/v glacial acetic acid 12.5%, 25% water, 67.5% glycerol). During a six weeks flowering period, every other day, up to three flowers per plant were selfed, until at least five up to 13 flowers were selfed. Other flowers were used to evaluate male and female fertility with testcrosses using unrelated diploids with known male and female fertility to distinguish infertility from self-incompatibility. The number of berries was recorded after eight weeks. Seed set in ripened fruits was examined to rule out parthenocarpic development. The amount of seed per berry was visually estimated on an arbitrary scale from 1 (few) to 3 (many) by cutting three berries per plant.

Pollen tube growth in the style and pollen tube arrest was observed under UV microscope with a DAPI filter (excitation filter 365 nm; dichroic mirror 425 nm; emission filter 470 nm). In three replications styles were collected 48 hours after selfing. The fixation and staining protocol followed Peterson et al. (2016). Images were taken with the Zeiss Axiomcam 305 colour camera to record the amounts and growth of pollen tubes in the style near the stigma and near the ovary (Fig. 1). This was scored on an arbitrary scale from 0 to 3 as follows: no pollen tube, few pollen tubes (less than 20), some pollen tubes, many pollen tubes. Plants with few pollen tubes near the

stigma, or inconsistent scores between biological replicates were excluded from evaluation.

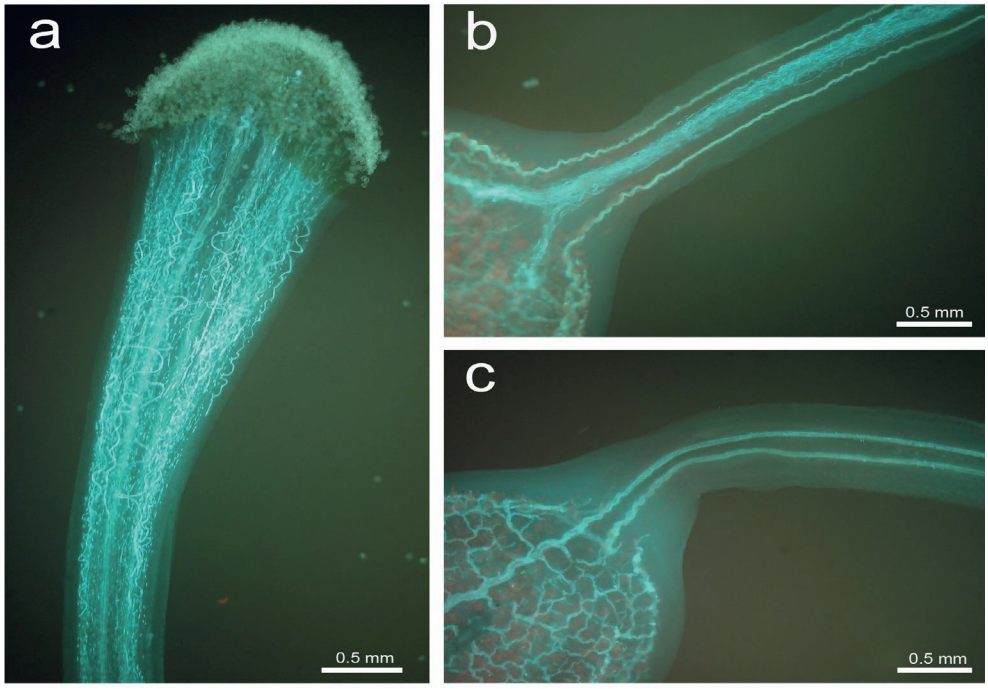

Figure 1:Pollen tube growth in style. **a)** Pollen tube growth from stigma to style. **b)** Ovary end of a self-compatible clone with pollen tubes reaching the ovary. **c)** Ovary end of a self-incompatible clone without pollen tubes reaching the ovary.

Kendall's Tau correlation (τ) was calculated for the phenotypic observations described above. Plants were regarded as SC when a substantial amount of pollen tubes reached the ovaries and resulting in more than 75% berry set. SI was concluded when all (or occasionally when almost all) tubes showed growth arrest in the style and none reached the ovaries. This resulted in bulking 10 SC and 11 SI plants for population IVP16-587, and 12 SC and eight SI plants for population IVP17-618.

Genotyping

To identify the SC locus we followed the Bulked Segregant Analysis method (BSA; Michelmore et al., 1991). Vigorous offspring yielding berries with seeds and without stylar pollen tube arrest were selected and pooled in the SC bulk. Offspring with stylar arrest of pollen tube growth, berry drop upon selfing and berry set upon crossing with fertile testers were pooled in the SI bulk. This results in four DNA samples per mapping population indicated as PSC and PSI representing the two parental lines and BSC and BSI representing the bulks. The samples BSC and BSI were created by pooling equal amounts of young leaf tissue. DNA was extracted with DNeasy 96 Plant Kit (QIAGEN,

Venlo, NL) and from each sample, 25Gb of clean sequences (100 bp paired-end small insert) were ordered from BGI (Copenhagen, DK) according to manufacturer's conditions. The sequencing data generated is available from the European Nucleotide Archive (ENA) under the BioProject ID PRJEB36551.

Read archives of SC diploid clones RH89-039-16, M6 and US-W4, were retrieved from the ENA (BioProject ID PRJEB2504, PRJNA362370 and PRJNA356643 respectively). Pair-ends runs ERR033768, ERR033769, ERR033770, ERR033771, ERR033772 and ERR033773 were used for RH89-039-16; SRR5264017, SRR5264021 and SRR5264022 were used for M6; and SRR5090798 for US-W4. The read archive of (Hardigan et al. 2017), describing DNA sequence diversity of wild tuber-bearing Solanum species, landraces and American varieties, was retrieved from BioProject PRJNA378971. The *Solanum phureja* DM1-3 (DM) pseudomolecule DM v4.03 was retrieved from Spud DB (http://solanaceae.plantbiology.msu.edu/).

Bulk specific *k*-mers were identified with the CoSSA strategy (Prodhomme *et al.* 2019; https://github.com/cprodhom/CoSSA-workflows). In short, sequence reads were quality trimmed and adapter sequence where removed using Trimmomatic (v0.32) (LEADING:3, TRAILING:3, *SLI*DINGWINDOW:4:15 and MINLEN:70) (Bolger et al., 2014). The Glistmaker program of the GenomeTester4 toolkit (Kaplinski et al., 2015) was used to build *k*-mer tables for each sample in this study. A *k*-mer size of 31 nucleotides was selected, compromising between sequence uniqueness and sequence correctness. *K*-mers observed with a frequency of one were removed from the dataset as they were likely caused by sequencing errors. The GlistCompare program of GenomeTester4 was then used to perform set operations such as union, intersection or differences.

K-mers in coupling and repulsion phase with SC were obtained using set operations, as illustrated in Fig. 2 and the resulting volumes of *k*-mers are reported in Suppl. Table 1. Hereafter we use the A\B and A∩B notation to indicate the set difference (elements specific to A) and the intersection of elements shared between the sets A and B.

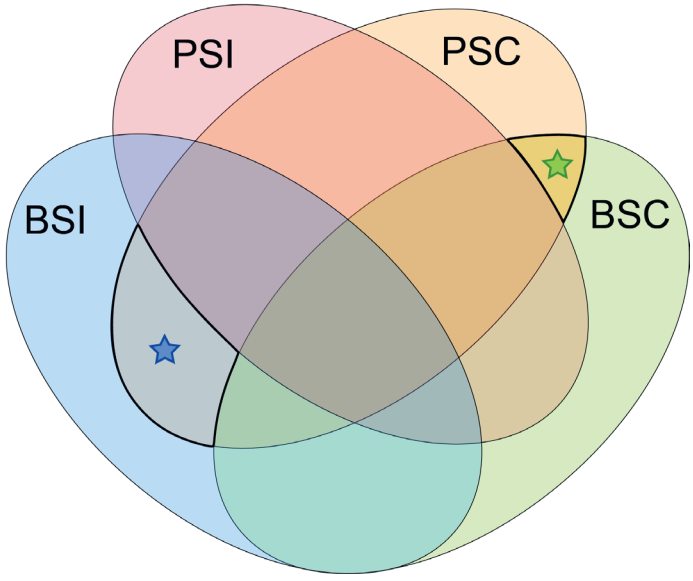

Figure 2: Set operations shown as Venn diagram. Intersections of *k*-mers from the SC (orange oval) and SI parent (red oval), and the SC (green oval) and SI bulk (blue oval) result in identification of *k*-mers associated with SC in coupling phase (green star) or re repulsion phase (blue star).

K-mers in coupling phase with a dominant SC locus must be specific to the SC-bulk and transmitted from the SC-parent. They are obtained from the intersection between SC-bulk specific and SC-parent specific *k*-mers. Using set algebra notation, those operations can be summarized as (BSC\BSI)⋂(PSC\PSI)). *K*-mers in repulsion phase with the SC locus must be specific to the SI-bulk and transmitted from the SC-parent. Assuming that the haplotype in repulsion phase with the SC locus in the SC-parent is not shared with the SI-parent, the *k*-mers linked in repulsion phase are obtained via the operation (BSI\BSC)⋂(PSC\PSI).

Tables of *k*-mer spectra (reported in Suppl. Table 2) were mapped to DM v4.03 using BWA aln (v0.7.12) allowing 3 mismatches. The number of *k*-mers per 1 Mb bins, counted using bedtools (v2.25) (Quinlan and Hall, 2010), were plotted using Matplotlib (Hunter, 2007).

K-mer analysis were performed independently for mapping populations IVP16-587 and IVP17-618. The intersection between *k*-mers in coupling phase with the SC locus of IVP16-587 and IVP17-618 (IVP16/17SCcoup) to test if the two mapping populations shared the same SC haplotype. The resulting table was depth refined, mapped and plotted as described above.

Comparison with reads archives

To investigate the origin of the SC haplotype, we used this IVP16/17SCcoup *k*-mer table to assess the presence of the SC haplotype specific *k*-mers in well-known SC diploid clones M6, RH89-039-16, US-W4 and C151, and in a panel of potato varieties, landraces and wild relatives (Hardigan et al., 2017). To do so we computed the intersection between our IVP16/17SCcoup *k*-mers and the *k*-mers generated from these read archives (Suppl. Table 3). The resulting *k*-mers tables were depth refined and mapped as described above. The number of *k*-mers per bins of 50kb was counted and plotted using seaborn (Waskom et al., 2014). Alternatively, mapped *k*-mers were directly visualised using Integrative Genome Viewer v.2.5.3 (IGV) (Robinson et al., 2011). Finally, the online potato pedigree database (van Berloo et al., 2007) was used for the familial interpretation of our data.

Estimation of genetic distances

Genetic distances between clones were estimated using the Mash algorithm version 2.2 (Ondov et al., 2016). Mash distances resemble Jaccard index between *k*-mers sets and correlate well with the average nucleotide identity. Trimmed reads were used as input with default sketches parameters and 31-mers. The distance matrix obtained was used in a Principal Coordinate Analysis (PCoA) and plotted with Past version 3.26 (Hammer et al., 2001).

Results

Analysis of SC related phenotypic data and bulk composition

Out of the 100 plants grown in populations IVP16-587 and IVP17-618, 57 and 47 plants were excluded from analysis, respectively, because of poor vigour or because less than nine flowers could be selfed. Phenotypic data on berry set, amounts of seed set, pollen viability and amounts of pollen tube growth were collected on the remaining plants (Suppl. Table 4).

Kendall's Tau correlations between pollen stainability, amount of pollen tubes in the style near the stigma, berry set, and seed set were weak and τ never exceeded 31%. In contrast, the amount of pollen tubes in the style near the ovary correlated strongly with berry set (τ = 0.64 and 0.69 in population IVP16-587 and IVP17-618, respectively). Berry set correlated reasonably well with the amounts of seed per berry (τ = 0.60 and 0.51). This is consistent with the observation that parthenocarpic fruit development was not frequent and typically observed only when flower drop (no berry set) was already high.

Among the 43 remaining plants of IVP16-587, 15 plants did not set any berry and were considered SI, while 13 plants with more than 75% of berry set were considered SC.

Hence fifteen plants were inconclusive based on berry set alone. Among the 53 remaining plants in IVP17-618, 14 plants did not set any berry and were considered SI, while 11 plants with more than 75% of berry set were considered SC, leaving 28 plants inconclusive based on berry set alone. For both populations, we find no reason to reject a 1:1 Mendelian segregation for SC or SI, when inconclusive plants are ignored. We used the assumption that SC inherits as a monogenic dominant trait, in our subsequent strategy to bulk samples and to perform and analyse *k*-mer operations.

In addition to this first classification based on berry set, we used pollen tube arrest as an additional criterion, as well as a threshold of at least nine self-pollinated flowers. Pollen tube arrest – recorded as a strong decrease of the amount of pollen tubes from stigma to ovaries – resulted in berry drop. Likewise, unaffected amounts of pollen tubes near the ovaries resulted in berry set. Finally, our interpretation of reproduction traits resulted in the following bulks: BSC N=10 and, BSI N=11 plants for IVP16-587; and BSC N=12 and, BSI N=8 plants for IVP17-618.

Comparative Subsequence Sets Analysis (CoSSA) to identify haplotype specific SNPs associated with Self Compatibility

DNA sequencing of the SC and SI parents and SC and SI bulks (PSC, PSI, BSC, BSI) of both populations yielded between 27 and 29 Gb of sequence data for each sample (reported in EMS2). Considering a haploid genome of 844 Mb, this represents a read depth of ~34x per diploid parent or 17x per haploid genome. Assuming a uniform sequence coverage, the expected depth of 31-mers per haploid genome is ~12x and agreed well with observed *k*-mer spectra (reported in Suppl. Table 2). Hence, tables of *k*-mers associated in coupling and repulsion phase with SC can be refined by selecting for a depth between 6x and 23x. We followed CoSSA (Prodhomme et al, 2019) to identify haplotype specific SNPs via set algebra.

In population IVP16-587, out of the 62,128,547 SC bulk specific *k*-mers (BSC\BSI), 58% were transmitted from the SC parent only ([BSC\BSI]⋂[PSC\PSI]). After depth filtering we kept 20,972,831 unique *k*-mers assumed to be linked in coupling phase with a dominant SC locus. Similarly, 81,326,597 *k*-mers were unique to the SI bulk (BSI\BSC) out of which 40% were inherited from PSC only ([BSI\BSC]⋂[PSC\PSI]), giving after depth filtering 17,935,863 unique *k*-mers assumed to be associated in repulsion phase with SC. From the sets of *k*-mers associated in repulsion and coupling phase with SC, 92% and 84% could be mapped to DM v4.03 respectively. In Fig. 3 the volumes of unique *k*-mers are plotted to the potato chromosomes in 1 Mb intervals.

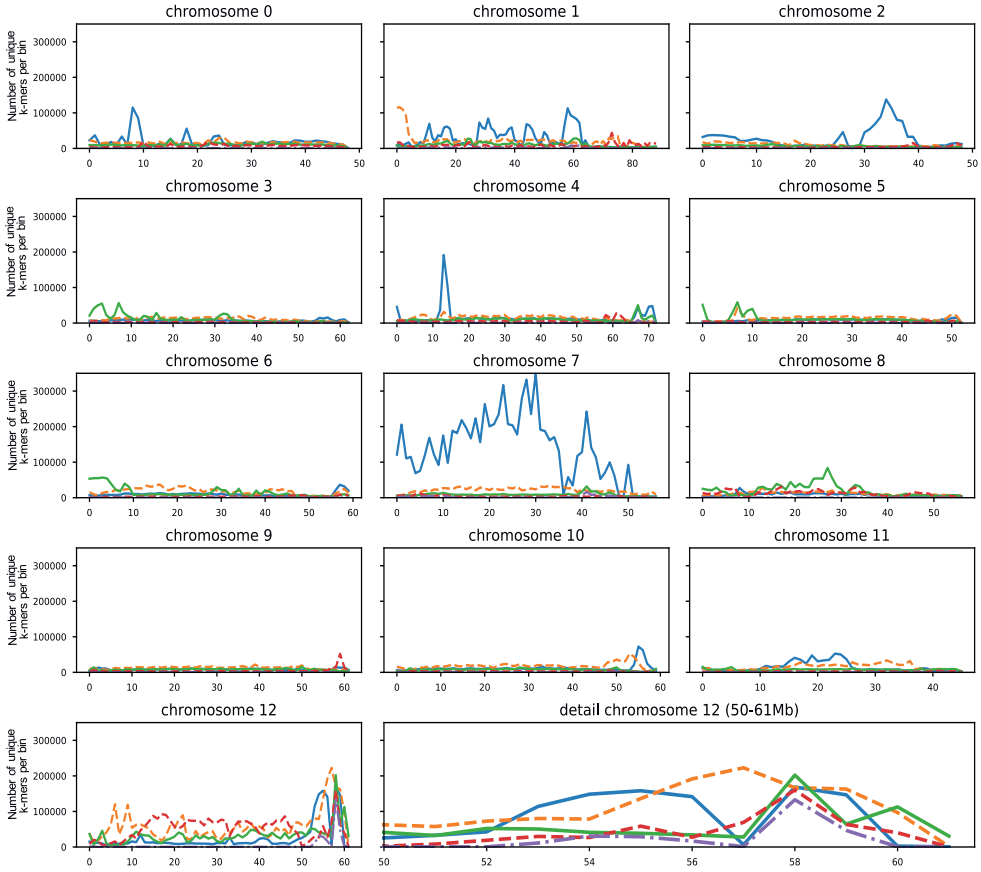

Figure 3: Self-compatibility mapped to the distal end of chromosome 12 using density graphs of unique *k*-mers mapped in 1 Mb bins of the reference genome DM v4.03. *K*-mers associated in coupling phase (solid blue line) or in repulsion phase with SC in population IVP16-587 (dashed orange line). *K*-mers in coupling phase (solid green line) or repulsion phase (dashed red line) with SC in IVP17-618. Intersection of coupling phase linked *k*-mers shared by IVP16-587 and IVP17-618 (dashed-dotted purple line).

Strikingly, 39.8% of the *k*-mers associated in coupling phase with SC mapped on a series of overlapping peaks spanning most of chromosome 7, from the proximal end to 48 Mb. Besides of chromosome *7*, the highest peak was identified on the distal end of chromosome *12*, from 53Mb to 60Mb, where 4.6% of the *k*-mers associated in coupling phase with SC mapped. Several smaller peaks of *k*-mers could be observed in other genomic regions, notably on chromosomes 0, 1, 2 and 4. An elevated frequency of *k*-mers linked in repulsion phase with SC was observed on the entire chromosome *12*. The highest peak was located between 55 and 61 Mbp where 6.5% of the *k*-mers mapped. Additionally, a smaller peak was observed on the first 4 Mb of chromosome *1*.

In population IVP17-618, out of the 87,512,765 SC bulk specific *k*-mers, 31% were inherited from PSC only. After depth filtering, 11,289,401 unique *k*-mers remained, which are assumed to be linked in coupling phase with SC. Similarly, 53,170,315 *k*-mers were unique to the SI bulk out of which 30% were inherited from PSC only. After depth filtering, we kept 7,326,761 unique *k*-mers, assumed to be linked in repulsion phase with SC. From the sets of *k*-mers associated either in coupling or repulsion phase with SC, 86% and 92% could be mapped to DM v4.03 respectively, also shown in Fig. 3.

Again chromosome *12* shows high frequencies of *k*-mers in coupling phase with a peak between 58 and 61 Mbp where 3.9% of the *k*-mers mapped. Several smaller peaks were observed on chromosomes *3*, *4*, *5*, *6* and *8*. Similarly, the high frequencies of *k*-mers associated in repulsion phase with SC were observed across chromosome *12*. The highest peak was identified on chromosome *12* between 58 and 61 Mbp where 4.3% of the *k*-mers mapped. Additionally, smaller peaks could be observed on chromosomes *1*, *8* and *9*.

Overall, the CoSSA suggest various chromosomal regions potentially associated with SC. Interestingly, when comparing the distribution of *k*-mer peaks assumed to be linked in coupling phase with SC with those in repulsion phase, only the peaks on chromosome *12* overlapped. Moreover, the distal end of chromosome *12* was supported by coupling and repulsion phase *k*-mer peaks in both populations. This suggests that both populations may share the same haplotype involved in SC. To verify this hypothesis an intersection was made between the *k*-mers in coupling phase with SC of population IVP16-587 and IVP17-618. After depth filtering, a set of 905,192 unique *k*-mers assumed to be in coupling phase the SC locus in both populations was obtained. 84% of those *k*-mers could be mapped to potato reference genome. The only peak observed was located on the distal end of chromosome *12* were 29.8% of the *k*-mers mapped (Fig. 3). Visualization of the same set of *k*-mers in IVG allowed a more precise identification of the haplotype boundary shared by the two SC bulks, which ran from 58.03 Mb to 59.36 Mb (Suppl. Fig. 2a).

Read archives expose the widespread presence of the haplotype conferring Self-Compatibility

The identification of a locus involved in SC at the distal end of chromosome *12* is reminiscent to the position of the *Sli* locus as described by Hosaka and Hanneman (1998b), as well as the locus involved in self-fertility proposed by Peterson et al. (2016). Therefore we compared our haplotype specific *k*-mers with *k*-mers generated from publicly available read archives of the diploid clones M6 (representing the *Sli* locus) and RH89-039-16 used by Peterson et al. (2016). This comparative study also included other well-known diploid SC clones: US-W4 and C151, as well as a panel of

varieties, landraces, and wild tuber-bearing *Solanum* species sequenced by Hardigan et al., (2017). The *k*-mer volumes of the intersection between *k*-mers in coupling phase with SC shared by population IVP16-587 and IVP17-618 (IVP16/17 SC coupling) and these public read archives are visualized in Fig. 4a.

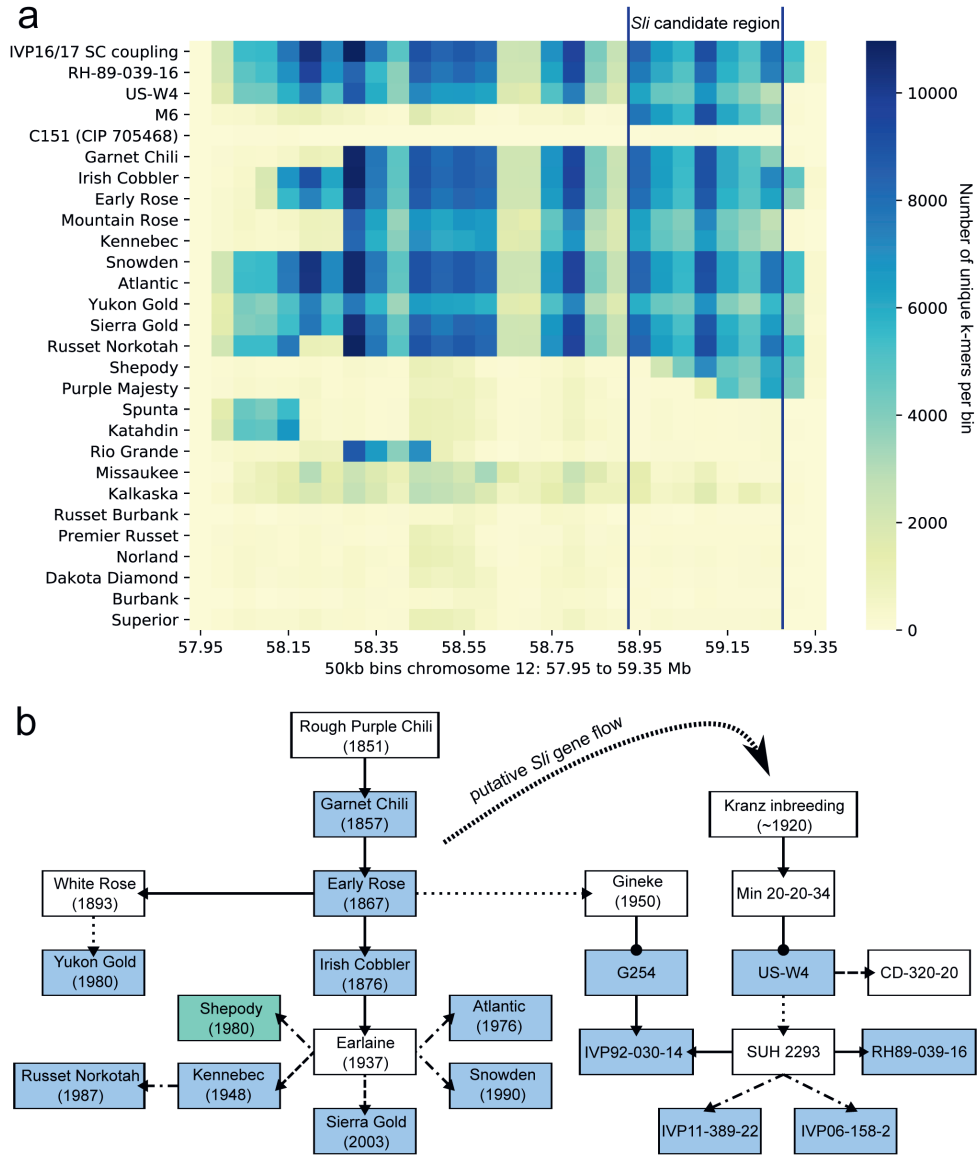

Figure 4: **a)** Detection of high densities of specific *k*-mers associated with SC (mapped to 50 kb bins of DM v4.03; columns) across various potato genotypes (rows) suggests the identification of one common haplotype involved in SC at a distal locus on chromosome *12*. Row 4 represents the *Sli* locus of M6. This suggests that one haplotype identical-by-state to *Sli* is also present in potato varieties. Genomic sequences from well-known SC diploids and varieties (Hardigan et al., 2017) were retrieved from public archives. **b)**

Pedigree relationships between genotypes positive for the SC haplotype. Year of market release is shown for varieties. The first (putative) SC ancestor of each clone was annotated together with the generation at which it was found. Blue boxes: clone bearing a complete SC candidate region; teal box: clone bearing an incomplete SC candidate region; white boxes: clones not tested in this study. Relatives are connected with arrows. First generation relatives—solid arrows; 2nd generation relatives—dashed arrows; 3rd generation relatives—dashed-dotted arrows; 6th generation relative—dotted arrow. Line with dot-node: dihaploidisation.

Each row in Fig. 4a shows a plot of the density of SC-associated *k*-mers in windows of 50kb, as identified in each specific potato sample. A clear differentiation can be made between genotypes where the intersection identified high or low densities of shared *k*-mers. We observed that the SC clones RH89-039-16, M6 and US-W4, display high shared *k*-mer densities. This observation supports identity-by-state (IBS) of the *Sli* locus of M6 with the SC haplotype of RH89-039-16 and US-W4, as well as with the one of the SC parents of our mapping populations: IVP11-389-22 and IVP06-158-2. Interestingly, the SC clone C151 does not share *k*-mers with this SC haplotype which is in agreement with the different mechanism conferring SC in C151 (Zhang et al., 2019).

Fig. 4a shows the physical length of the SC (sub-)haplotypes as well as the positions of recombination breakpoints. While our SC associated *k*-mers identified a long SC haplotype in RH89-039-16, *k*-mers from bin 59.35-59.40 Mb were absent in US-W4 and M6 haplotype. Interestingly, most of them were found in PI498359, a *S. kurtzianum* accession (Suppl. Fig. 3). More importantly, the *k*-mers mapping form 58Mb to 58.95Mb were virtually absent in M6, suggesting a recombination breakpoint and allowing us to limit the *Sli* candidate region to seven bins of 50 kb, located from 58.95 Mb to 59.30 Mb. Direct visualisation of *k*-mer mapped using IGV allowed us to identify recombination breakpoints at 58.945 Mb and 59.278 Mb pinpointing *Sli* to a candidate region of 333 kb (Suppl. Fig. 2b).

It was unexpected to recognize high densities of SC-associated *k*-mers in several tetraploid varieties, while no significant amounts of SC-associated *k*-mers were found in the landraces and wild species (Suppl. Fig. 3) sequenced by Hardigan et al. (2017). A long SC haplotype was found in Garnet Chili, Irish Cobbler, Early Rose, Kennebec, Russet Norkotah, Sierra Gold, Yukon Gold, Snowden, Atlantic and Mountain Rose. The SC-associated *k*-mers found in Katahdin and Spunta as well as Rio Grande only mapped to bins outside of the *Sli* candidate region (58Mb to 58.2Mb and 58.3Mb to 58.55Mb respectively). Interestingly, SC-associated *k*-mers of Shepody and Purple Majesty only mapped from 59Mb to 59.4Mb and 59.1Mb to 59.4Mb respectively, suggesting a shorter haplotype due to recombination events within the SC haplotype in those two clones.

When sequencing depth allows it, an estimation of *Sli* dosage can be deduced by comparing the *k*-mer spectra of *k*-mers mapping to the *Sli* candidate region with the

total *k*-mer spectrum of a sample. The peak of haplotype specific *Sli k*-mers can either coincide with the simplex peak of all *k*-mers or be shifted toward higher frequency. This shift represents a higher allele dosage of *Sli*. The data presented in Suppl. Table 5 suggests that possibly Garnet Chili, Irish Cobbler, Early Rose and certainly Russet Norkotah are duplex for the *Sli* allele. Elevated allele dosages are not unexpected given the prevalence of the *Sli* locus and its distal position facilitating double reduction.

The widespread observation of a region involved in SC identical-by-state in old North American varieties and contemporary diploid progenitor clones prompted the question of how these potato genotypes relate to each other. A summary of the relatedness of varieties with the SC haplotype is shown in Fig. 4b based on potato pedigree information (van Berloo et al., 2007). The SC haplotype seems to have spread among tetraploid varieties via the frequent use of Early Rose, Irish Cobbler and potentially Earlaine as parents. Because all positive genotypes are closely related, we conclude that the SC haplotype is identical-by-descent. The variety Rough Purple Chili, introduced into the USA in 1851 by Goodrich (1863), is the origin of the SC haplotype. Fig. 4b also shows that all SC diploids have US-W4 as common ancestor. Subsequently, the SC haplotype spread in the breeding stock of Wageningen University via the use of the (now lost) clone SUH 2293. US-W4 is a dihaploid extracted from Minn. 20-20-34 a tetraploid progenitor with an unknown pedigree from the breeding program of Frank Krantz in the 1920s. Therefore it is reasonable to assume that Minn. 20-20-34 is also derived from Rough Purple Chili.

Remarkably, high densities of SC-associated *k*-mers, signature of the SC haplotype identified in this study, were also identified in clone *S. chacoense* M6. A shared haplotype between a wild and cultivated potato clones can be explained by introgression or coancestry. To test these hypothesis the genetic distance was estimated between the clones used in this study and those sequenced by Hardigan et al. (2017) using Mash (Ondov et al., 2016). As expected, *S. chacoense* accession PI-275139 is most closely related clone to M6. However, while *S. chacoense* PI-275139 is genetically close to *S. berthaultii, S. infundibuliforme, S. boliviense* and various *S. brevicaule* accessions; M6 shares a low Mash distance with several *S. tuberosum* tetraploid varieties (Table 1). Moreover, when the pairwise distance matrix of all clones was used in a PCoA (Fig. 5), M6 clustered between varieties and landraces and a cluster of wild species. In addition, M6 seems to cluster with *S. candolleanum* sp. *multidissectum* (PI 210044) and *S. tuberosum* sp. *stenotomum* (PI 195204), both described as hybrids between cultivated landraces and wild species (Hardigan et al., 2017). These results suggest that M6 is not a pure *S. chacoense*.

Table 1: Top 10 genetically nearest clones of S. chacoense (PI 275139) and M6 based on genetic distance estimates using Mash software.

S. chacoense **(PI 275139)**	**Mash distance**	**M6**	**Mash distance**
M6	0.0235827	*S. chacoense* 2x	0.0235827
S. berthaultii 2x (PI 458365)	0.0254441	Dakota Diamond tuberosum 4x	0.0249416
S. infundibuliforme 2x (PI 458324)	0.0281594	Snowden tuberosum 4x	0.0257861
S. boliviense 2x (PI 545964)	0.0283536	Premier Russet tuberosum 4x	0.0259593
S. brevicaule-sparsipilum 2x (PI 473385)	0.0286483	Rio Grande Russet tuberosum 4x	0.0261340
S. brevicaule-gourlayi 2x (PI 473065)	0.0286483	Sierra Gold tuberosum 4x	0.0269388
S. brevicaule-brevicaule 2x (PI 498112)	0.0289472	Mountain Rose tuberosum 4x	0.0269388
S. brevicaule-spegazzinii 2x (PI 472978)	0.0299747	Burbank tuberosum 4x	0.0270302
S. verrucosum 2x (PI 275260)	0.0302927	Yukon Gold tuberosum 4x	0.0271220
S. brevicaule-leptophyes 2x (PI 545987)	0.0302927	Irish Cobbler tuberosum 4x	0.0271220

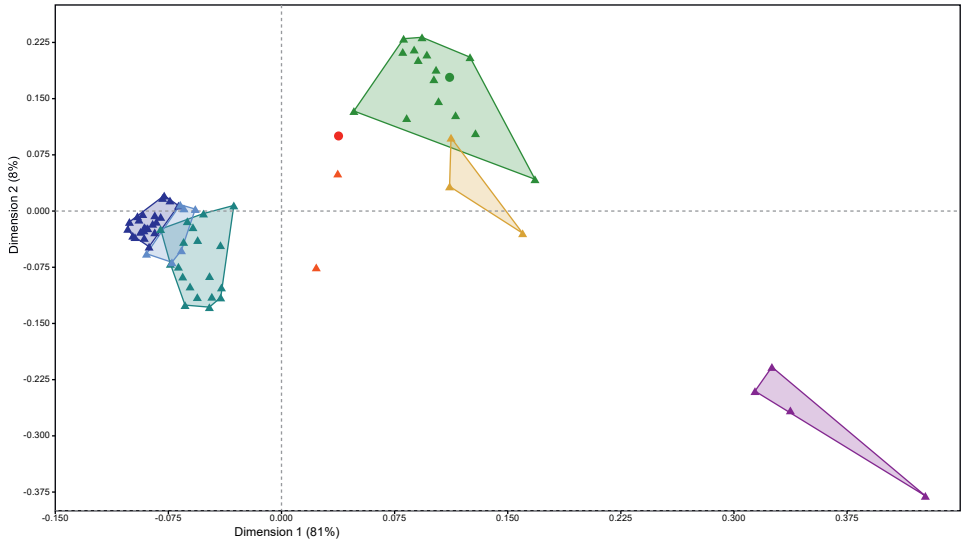

Figure 5: Principal coordinate analysis (PCoA) between the *Solanum* clones based on Mash distance. Symbol colours are based on clusters identified before by Hardigan et al. (2017; Fig. 1, data set 1). Wild species outgroup (purple triangles), wild species (green triangles), wild subgroup diverging from the cultivated lineage after most other species (gold triangles), landraces (teal triangles), tetraploid varieties (navy triangles); with the addition of hybrid between wild species and landraces (orange triangles) and diploid *S. tuberosum* (cornflower blue triangles). M6 is represented by a red dot and *S. chacoense* (PI 275139) by a green dot.

Haplotype specificity of *k*-mers

Only for IVP11-389-22 and IVP06-1582, the SC mapping parents of this study, it is assumed that all SC-associated *k*-mers are haplotype specific and thus linked in coupling phase. However, in the panel sequenced by Hardigan et al. (2017), *k*-mers can originate from homologous chromosomes and represent a blend of haplotypes. If pedigree information is true and Early Rose is a direct descendant of Garnet Chili, then *k*-mers mapping from bins 58.10 to 58.25 Mb and in bin 59.30 Mb, absent in Garnet Chili, must originate from the pollinator chromosome(s). Subsequent recombination must have brought those *k*-mers in coupling phase in IVP11-389-22 and IVP06-158-2, the parents of our mapping populations. Similarly, *k*-mers from Russet Norkotah which could be mapped from 58 Mb to 58.15 Mb and 58.3 to 59.3 Mb might be on different homologues. Two recombination events in bins 58.20 and 58.25 Mb is a less probable explanation of this *k*-mer pattern.

A striking feature shown in Fig. 4 is the similarity of the pattern of the pseudo-colours which reflect variations in *k*-mer densities among rows positive for the SC haplotype. These variations reflect the genomic landscape of the reference genome where unique or low-complexity DNA enabled mapping of many or few *k*-mers respectively. These patterns contain no information on IBD. For instance, the low frequency of *k*-mers in mapping between 58.65 Mb and 58.75 Mb as well as between 58.85 Mb and 58.95 Mb can be explained the 5000 bp gaps between DM scaffolds PSGC0003DMB000000034 and PSGC0003DMB000000750 as well as between and PSGC0003DMB000000750 and PSGC0003DMB000000114. Overall, IBD is concluded on the basis of the congruence between the rows, but not on the shade of the colour itself.

Validation of CoSSA results and development of MAS tools

To verify our *in silico* result biologically, haplotype specific KASP markers were designed both to verify the SC haplotype on chromosome *12* as well as to disprove large *k*-mers identified in coupling phase with SC during CoSSA of IVP16-587 and IVP17-618. We tested those markers on a set of 93 diploid clones hereafter referred to as the validation panel. The validation panel was composed of the bulked and parental clones of our two mapping populations, a subset of two parental lines and five progenies from a full sib-population IVP16-560 and 37 distantly related diploids with well-known SC or SI phenotype or tested for seed set upon self-pollination. The results are illustrated in Suppl. Table 6.

No association was observed with SC for the 27 KASP markers at the 10 population specific *k*-mers peaks beyond chromosome *12*, confirming the tentative conclusion that those other peaks observed were false positive. All 18 KASP markers designed within the SC candidate region on chromosome *12* allowed a perfect prediction of SC

in populations IVP16-587 and IVP17-618, albeit few missing values. Hence, CoSSA results could be confirmed with KASP markers.

Another full sib-population IVP16-560, studied for self-fertility in prior unpublished work, showed a seven self-infertile vs. 83 self-fertile segregation, severely deviating from the expected 1:1 ratio. The five IVP16-560 offspring clones tested with KASP (three infertile and two fertile clones) are heterozygous for all 18 KASP markers, and the SC parent IVP92-030-14 is homozygous. Marker data explain the lack of segregation for SI/SC in IVP16-560, because the SC parent IVP92-030-14 is homozygous for the SC haplotype and thus generated a uniformly heterozygous SC progeny. The seven self-infertile descendants observed in IVP16-560 did not set berries, not because of SI, but rather because of poor fertility.

KASP genotyping of the rest of the validation panel, composed of unrelated clones from our diploid collection, showed a correlation between the KASP markers and SC phenotype for 30 out of 40 clones. Nine clones positive for SC specific KASP alleles did not set berries due to lack of viable seeds after selfing or due to lack of fertility. One clone yielded 25 seeds after selfing despite the absence of SC specific KASP alleles. Contamination during pollination or DNA sampling/extraction errors may explain this contradiction.

Interestingly, the validation panel included RH89-039-16, G254 and two SC clones raised from *S. chacoense* seedlings: 07-2N100-1-5 and 07-2N104-2 (kindly provided by Dr. Kazuyoshi Hosaka). KASP results for these four clones indicated the presence of the SC haplotype, and confirms IBS for *Sli* from *S. chacoense* and the SC haplotype observed in *S. tuberosum* clones. The SC of the dihaploid clone G254 (Hermsen, 1978), extracted from the tetraploid variety Gineke can now be explained, because Gineke is a descendant of Early Rose (Fig. 5), so IBD is confirmed between the SC haplotype of G254 and Rough Purple Chili.

Discussion

Pollen tube growth is most indicative for SC

Spontaneous berry set is rarely observed in diploid potato, but the lack of successful sexual reproduction upon selfing has many reasons apart from self-incompatibility, such as infertility of male or female gametes, lack of gamete fusion, embryo abortion, etc. (reviewed by Johnson et al., 2019) Vegetative crops are subjected to low selection pressure for fertility traits, and only a few tetraploid varieties are known for a high level of spontaneous berry set such as Cara, Désirée and Gineke (see: https://europotato.org/characters/view/Berries). In diploids, deleterious recessive mutations are more easily exposed and thus ploidy reduction will aggravate fertility problems. Nevertheless, it was commonly assumed that SC diploids are rare and SI the

major reason for the inability to self-fertilize diploids. Initially, our phenotypic observations were also focussed on berry set, because this simple visual key allows classifying descendants in a segregating population easily, provided that one parent is renowned for spontaneous berries. However, berry set upon selfing is an observation related to self-fertility rather than self-compatibility (Peterson et al., 2016). To avoid mapping loci involved in fertility or parthenocarpy, we performed comprehensive phenotyping of many reproductions related traits. We found low to moderate correlations between these traits and the ability to set berries. The highest correlation was observed between berry set and amount of pollen tube reaching the ovaries. Gametophytic self-incompatibility is caused by S-RNase activity in stylar tissue and therefore stylar observations should allow to distinguish SI/SC better than indirect or confounding traits. However, in our experimental conditions, even the pollen tube arrest phenotype shows quantitative variation. Based on current experience, we recommend recording both pollen tube growth and berry set to unambiguously classify SC and SI plants.

Peterson et al. (2016) also started research on SC because several descendants of DM × RH89-039-16 set berries upon selfing. Surprisingly, segregation for pollen tube arrest could not be observed in their F_1 population. Peterson et al. (2016) focused on phenotyping self-fertility upon inbreeding, and concluded that the south of chromosome *12* was responsible for fruit set in self-fertile plants in the F_1 generation, whereas other genomic regions contributed to the ability of S3 plants to set fruit after self-pollination. In this study, we proved that diploid clone RH89-039-16, bears *Sli* at a position which is in agreement with their map position at the distal end of chromosome *12*. Hence, we argue Peterson et al. (2016) mapped self-compatiblity in the F_1 and identified a valuable QTL associated with self-fertility and/or inbreeding tolerance in the S3.

The benefits of *k*-mer based analyses

Several reasons support the Comparative Subsequence Sets Analysis of DNA sequences (Prodhomme et al., 2019). In the first place because this approach does not rely on a reference genome for variant calling. Previous mapping studies have speculated on the molecular cause of SC. Olsder and Hermsen (1976) originally favoured a translocation of the *S*-locus, which resulted in competitive interaction. Subsequent results by Thompson et al. (1991) suggested a translocation of the pollen part of the S-locus only, now known as SLFs genes (Kao and Tsukamoto, 2004). Alternatively, an independent inhibitor allele active in the pollen only was also proposed (Hosaka and Hanneman, 1998a; Thompson et al., 1991). Translocation events may result in unreliable read mapping and also complicate the interpretation of mapping studies due to pseudo linkage (Farré et al., 2011). As a result, reference free sequence comparison method was favoured to map the locus involved in SC.

A second reason to employ CoSSA is computational simplicity. Demanding bioinformatics steps such as read mapping and variant calling could be omitted. Instead, set algebra allows fast and stringent comparisons between samples to retain small subsets of *k*-mers.

Thirdly, the *k*-mers identified by CoSSA (Prodhomme et al., 2019) represent haplotype specific SNPs, associated in coupling phase or repulsion phase with SC. Initially, the data from mapping population IVP16-587 identified a large number of *k*-mers that mapped to chromosome *7*. This peak could be ruled out because *k*-mers associated in repulsion phase with SC only confirmed the peak on chromosome *12*. This a great advantage of diploid organisms, where the accuracy of CoSSA greatly benefits from the use of repulsion phased *k*-mers. This aspect was not mentioned in the original paper, where tetraploids were used (Prodhomme et al., 2019). Haplotype specificity of markers (*k*-mers), initially obtained by BSA, can be improved by subtracting *k*-mers from any other potato clone without the trait-specific haplotype.

Marker assisted breeding

The haplotype specific *k*-mers obtained with CoSSA were easily converted into KASP marker assays following recommendations by Prodhomme et al. (2019). This allowed validation of the candidate region on the distal end of chromosome *12* and to disprove the false positive *k*-mer peaks, in particular the peak on chromosome *7*. The KASP markers are available for breeders and researchers to apply marker assisted selection of SC. This will improve the efficiency of breeding programs, because time-consuming phenotyping of SC and distinguishing it from self-fertility can now be circumvented. Markers will be particularly beneficial for the selection of progenitors with the SC haplotype, including tetraploid progenitors used for the extraction of SC diploids, during the early stage of diploid breeding programs. Once breeding with SC diploids makes progress, the value of marker assisted selection will reduce. Indeed, selfing SC clones will result in 100% SC offspring, because only the pollen tubes bearing the *Sli* gene will be able to reach the ovary and produce offspring. This will result in a rapid fixation of *Sli* in homozygous state.

Is SC associated with lethality?

Interestingly, both Hermsen (1978) and Hosaka and Hanneman (1998a) reported a recessive lethal allele linked to *Sli* inducing skewed segregation for the SC. Similarly, distorted segregation of markers was observed on the distal end of chromosome *12* in RH89-039-16 derived populations (Endelman and Jansky, 2016; Peterson et al., 2016; van Os et al., 2006; Zhang et al., 2019). Recently, Endelman et al. (2019) suggested that the lethal allele linked to *Sli* was also present in DM in the interval 57.2-57.8 Mb of chromosome *12*. The position of this region, about 1 Mb south of the *Sli* candidate region identified in this study, is consistent with the hypothesis stating that *Sli* is linked

to a lethal allele. In our KASP validation panel, we identified seven putative *Sli* homozygous clones. The observation of homozygosity is not only based on KASP markers. One of the homozygous clones, IVP92-030-14, who received one *Sli* allele from G254 and the other from US-W4 (Fig. 5), only produced SC offspring. This confirms that KASP marker data are reliable predictors of the genotype at the *Sli* locus and the resulting phenotype. Hemizygosity at the *Sli* locus can be excluded as well. Therefore, it seems that the lethal allele described above has recombined from *Sli* and is not a limitation to reach homozygosity in this region. This is in agreement with Marand et al. (2019) who did not identify recalcitrant heterozygosity in the distal end of M6 chromosome *12*.

Allele mining from read archives

Increasingly, the FAIR (Wilkinson et al., 2016) availability of nucleotide read archives allows allele mining (van de Weg et al., 2004). Allele mining implies the identification of alleles present in the gene pool where IBD is utilised to express the identity of specific alleles in terms of alleles of founding cultivars. With set algebra, this study takes allele mining to a next level, because we directly evaluated the intersection between our haplotype specific *k*-mers with *k*-mers from potato varieties sequenced before. Furthermore, this study demonstrates the scalability of this strategy because it circumvents read mapping and variant calling. The large dataset generated by (Hardigan et al., 2017) appeared to be very useful for mining of the *Sli* allele. From the nucleotide archive of clone M6 we could identify a recombinant haplotype and narrow the candidate region of the *Sli* locus from 1.5 Mb to 333kb. Interestingly, recombination breakpoints within the SC candidate region were also identified in the haplotype of Shepody and Purple Majesty. Fertile dihaploids derived from those varieties could be used for recombinant analysis to fine map the *Sli* locus. With the increasing amounts of sequencing data submitted to online repositories, these *k*-mer based set operations will gain in relevance and power.

Rough Purple Chili as founder clone of SC germplasm

Rough Purple Chili, identified as the source of *Sli* in this study, was described by Hawkes (1979) as one of the four introductions event establishing the genetic foundation of all European and North American varieties. Rough Purple Chili can be found in the pedigree of nearly all American (Plaisted and Hoopes, 1989) and European varieties (Hawkes, 1979). Descendants of Rough Purple Chili, such as Garnet Chili, Early Rose and Irish Cobbler are recognized as major contributing ancestors (Love, 1999). As they were positive for *Sli*, it is clear that these varieties facilitated the spread of *Sli* into the American gene pool. Bradshaw *et al.* (Bradshaw et al., 2006) mentioned that Early Rose was also widely used by European breeders. For instance, Early Rose was the female parent of Abundance and Epicure, from the Scottish Plant Breeding Station. Interestingly, the dihaploid B16 induced from the clone Black 4495, bred in the

same station, was described as self-compatible by Olsder and Hermsen (1976). In the meantime, Early Rose was also used as female parent in Germany to produce Imperator, an ancestor of Gineke, from which the SC dihaploid G254 (Olsder and Hermsen, 1976), found positive for *Sli* KASP markers in this study, is derived. Hence, the *Sli* locus may be surprisingly common in both American and European germplasm.

Thoughts on the origin of *Sli*

The identification of a SC haplotype, indistinguishable between historical *S. tuberosum* cultivars and *S. chacoense* clones is not easily understood. Interspecific gene flow between cultivated and wild potato has been extensively reported (Celis et al., 2004; Hardigan et al., 2017; Rabinowitz et al., 1990; Scurrah et al., 2008) and interploidy genetic exchange is facilitated by the production of unreduced gamete in diploid potato (Watanabe and Peloquin, 1989). *Sli* may be another example of geneflow between wild and cultivated potato germplasm. Based on our data, the direction and timeframe of a putative introgression of *Sli* remain speculative. Hybridisation may have occurred between the Chilean *S. tuberosum* ancestor of Rough Purple Chili and a wild *S. chacoense*. Alternatively, a more recent hybridisation may have happened between SC *S. tuberosum* dihaploids and the ancestor of M6.

A final example of the utility of *k*-mers was offered by the distance estimation software Mash (Ondov et al., 2016). Because all *k*-mer tables were already generated this software allowed to explore the question on the origin of the *Sli* locus in *S. chaocense* clone M6. Despite being an estimation, the genetic distances calculated by Mash with *k*-mers agreed very well with the classification by Hardigan et al. (2017) who used read mapping and variant calling. The wild outgroup, the wild potato species, the landraces and tetraploid varieties are properly separated by Mash distance. Hardigan et al. (2017) also identified a wild subgroup diverging from the cultivated lineage, composed of three Peruvian species (*S. medians, S. megistacrolobum, and S. raphanifolium*). Based on Mash distance, the separation of this sub-group from the wild species cluster is not clear. Interestingly, the hybrids *S. candolleanum* sp. *multidissectum* (PI 210044) and *S. tuberosum* sp. *stenotomum* (PI 195204) are clustering with M6 between landraces and wild species in this study (Suppl. Fig. 4). This was also reported by Hardigan et al. (2017) in Supplementary Figure 6.

Remarkably, M6 share a close Mash distance with Dakota Diamond and Snowden. Those cultivars have 1/8 and 1/16 *S. chacoense* ancestry, respectively. However, in Fig. 5 Dakota Diamond and Snowden clearly cluster with the rest of the tetraploid varieties. On the other end, M6 is localized between wild species and tetraploid varieties, suggesting that despite its putative hybrid status, M6 present a much larger proportion of *S. chacoense* ancestry than those clones. Recently genomic data on structural variation was used to compare several landraces and M6 (Kyriakidou et al., 2020). The

position of M6, clustering with the *S. tuberosum* landrace CIP 705053 once more suggests that an ancestor of M6 hybridised with *S. tuberosum*. Although the *Sli* haplotype was not observed in wild species and landraces in the panel of Hardigan et al. (2017), and hybridisation with *S. tuberosum* may explain SC in M6 more easily, we cannot exclude a second origin of *Sli* from *S. chacoense*.

Candidate genes

Complex biological mechanisms such as gametophytic self-compatibility are usually controlled by many different genes, coding for catalytic proteins of a metabolic route or transcription factors for a signalling cascade. Because mutations in each component could be anticipated, it is difficult to speculate on the identity of the *S*-locus inhibitor gene(s) encode by the *Sli* locus. For example, in *S. lycopersicum*, breakdown of self-incompatibility is associated with a loss of function mutation in both *S-RNase* and *HT* genes (Kondo et al., 2002). No obvious candidate inhibitor genes or *SLF*s were found in our candidate region. However, the annotation of the DM reference genome reveals the presence of a Self-incompatibility S1 family protein (PGSC0003DMG400016877 at 58.968 Mb) and a cluster of four F-box proteins (PGSC0003DMG400016863, PGSC0003DMG400016862, PGSC0003DMG400016861 and PGSC0003DMG400046496 at 59.024-59.057 Mb) that could be proposed as candidate genes.

Implications for diploid potato breeding

This study shows that *Sli* is widespread and that many clones can be used to breed SC diploid potato, and removes the genetic bottleneck that could be experienced by breeders in the early days of diploid inbreeding. We also show that many sources of SC could be unified into a single locus known as *Sli*. Clone C151 described as SC due to a low expression level of an *S*-allele (Zhang et al., 2019) is an alternative to *Sli* based SC. Additionally, two artificial SC clones were recently generated by mutation of the S-RNase gene with CRISPR-Cas9 (Enciso-Rodriguez et al., 2019; Ye et al., 2018). Both authors emphasize the necessity to engineer SC mutants because introgression of *Sli* in the cultivated germplasm would result in linkage drag of undesirable *S. chacoense* traits such as long stolons or high tuber glycoalkaloid content. Because the *Sli* locus appears to be indigenous in cultivated potato germplasm, the use of genetically engineered plants can be circumvented.

Zhang et al. (2019) proposed to introduce SC into *S. tuberosum* by crossing dihaploids from selected tetraploid varieties with M6 and other wild sources to introduce SC. Considering the widespread of the SC haplotypes among the tetraploid varieties, the one selected for dihaploids induction should include *Sli*. This suggests that the diploid SC gene pool can be expanded easily without the need for subsequent *Sli* introgression.

With this study, the challenges in diploid potato breeding will shift. The need for SC germplasm with agronomical value can be satisfied by screening genomes for specific *k*-mers or with the use of KASP markers. The next challenge is to develop material which remains highly fertile upon inbreeding. Male fertility appears as the next limitation to overcome and was already identified as an impediment to breeding in the early 20th century (Krantz, 1924). Tetraploids cultivars with renowned spontaneous berry set should be excellent candidates for the induction of fertile dihaploids, but fertile and SC dihaploids such as G254 can be expected only if their tetraploid parents are *Sli*-bearing descendants from Rough Purple Chili.

Author contribution statement

CRC, CP, RH and HJvE conceived and designed the experiment and interpreted the results. RH and CE provided plant materials and advised on phenotypic data collection, CRC and CP collected and analysed phenotypic data in combination with sequence data. CRC traced the origin of *Sli*. CS generated KASP data. CRC and CP wrote the manuscript with the help of HJvE who edited the final version. All authors read and approved the manuscript.

Acknowledgements

We thank the members of the public-private partnership "A new method for potato breeding: the 'Fixation-Restitution' approach" and "DIFFUGAT: Diploid Inbreds For Fixation, and Unreduced GAmetes for Tetraploidy" (Averis Seeds B.V., Bejo Zaden B.V., Danespo A/S, Germicopa, Den Hartigh B.V., C. Meijer B.V., and Teagasc) for providing their support. We are grateful to Danny Esselink, Martijn Kaauwen and Chengcheng Cai for help on computational analysis. We are thankful to Ea Riis and Christan Bachem for their constructive criticism of the manuscript. Our colleagues of Unifarm are acknowledged for plant care in the greenhouse.

Supporting information

Supplementary files are available at: https://doi.org/10.1007/s00122-020-03627-8 under the name Electronic Supplementary Material (ESM).

Suppl. Fig. 1 (ESM1): Pedigrees of population IVP16-587 and IVP17-618.

Suppl. Table 1 (ESM2): Overview of DNA sequencing and *k*-mer data sets from IVP16-587 and IVP17-618 used for set operations.

Suppl. Table 2 (ESM3): Frequency spectra of *k*-mers across all samples used in this study.

Suppl. Table 3 (ESM4): Overview of *k*-mer data sets from the read archives used for set operations.

Suppl. Table 4 (ESM5): Overview of phenotypic data for reproduction related traits in IVP16-587 and IVP17-618.

Suppl. Fig. 2 (ESM6): Boundaries of the SC haplotype. IVG viewer was used to show map positions of SC specific *k*-mers on DM v4.03.

Suppl. Fig. 3 (ESM7): Heatmap displaying the density of SC specific *k*-mers on the distal end of chromosome 12 for all Hardigan et al. (2017) samples.

Suppl. Table 5 (ESM8): Allele dosage estimation of Sli in diploids and tetraploids.

Suppl. Table 6 (ESM9): Conversion of *k*-mers into KASP markers and validation of KASP marker predictions using a panel of diploid clones with known self-compatibility/fertility phenotype.

Suppl. Fig. 10 (ESM10): Dendrogram of clones used in this study based on neighbour-joining clustering of Mash distance. Colour used is based on Fig. 5.

Chapter 3

High-density linkage map constructed from a skim sequenced diploid potato population reveals transmission distortion and QTLs for tuber yield and pollen shed

Authors

Corentin R. Clot[1], Xulan Wang[1], Joey Koopman[1], Alejandro Thérèse Navarro[1], Johan Bucher[1], Richard G.F. Visser[1], Richard Finkers[1,2], Herman J. van Eck[1]

Affiliations

[1] Plant Breeding, Wageningen University and Research, Po Box 386, 6700 AJ Wageningen, Netherlands

[2] GenNovation B.V., Agro Business Park 10, 6708 PW Wageningen, The Netherlands

Published in Potato Research (2023)

https://doi.org/10.1007/s11540-023-09627-7

Abstract

The reinvention of potato, from a tetraploid clonal crop into a diploid seed-based hybrid crop, requires insight in the mutational load, recombination landscape and the genetic basis of fertility. Genomics based breeding and QTL discovery rely on efficient genotyping strategies such as skim sequencing, to gather genotypic information. The application of skim sequencing to full-sib population of non-inbred parents remains challenging. Here, we report on an R implementation of the OutcrossSeq pipeline for diploids. We applied this pipeline to a large diploid skim sequenced potato population. We used the resulting bin-markers for the construction of high-density parent specific linkage maps, highlighting variation in parental recombination rate and structural variations. We subsequently explored transmission ratio distortion and non-independent assortment of alleles, indicative of large-effect deleterious mutations. Finally, we identified QTLs for seedling tuber yield in pots and pollen shed. This study showcases the range of genetic analyses, from marker inference, identification of transmission ratio distortion, linkage map construction to QTL mapping, resulting in new insights that contribute to breeding diploid potato.

Introduction

Potato (*Solanum tuberosum*) breeding programs are undergoing a drastic transformation from the classical tetraploid system and clonal propagation to diploid F_1 hybrid breeding system (Jansky et al., 2016; Lindhout et al., 2011). The generation of diploid inbred lines is expected to bolster genetic gains in potato by circumventing the complexity of tetrasomic inheritance and facilitating the fixation of beneficial alleles. The original obstacle of self-incompatibility at the diploid level was recently overcome with the identification of the widespread *Sli* gene (Clot et al., 2020; Eggers et al., 2021; Ma et al., 2021) or by using dysfunctional *S-RNase* alleles (Enciso-Rodriguez et al., 2019; Zhang et al., 2019). The second obstacle encountered in this transformation is inbreeding depression. Inbreeding at the diploid level results in the loss of vigour and fertility within a few generations due to the exposure of recessive deleterious mutations. Evaluation of 5377 dihaploids, extracted from tetraploid varieties and progenitor clones, illustrates the problem of mutational load in potato germplasm. Only 60% of these dihaploids set tubers, the majority show problems with flowering and fertility, and only 4% displayed good pollen stainability (Hutten et al., 1995). When the focus on clonal reproduction is replaced by sexual reproduction, fertility issues will prompt for new research projects to understand the genetic factors involved in flowering (Seibert et al., 2020), anther development (Endelman and Jansky, 2016), pollen shed and pollen viability. Despite some effort to identify QTLs regulating self-fertility (Peterson et al., 2016; Phumichai and Hosaka, 2006) or pollen stainability (Zhang et al., 2019) the genetic basis of fertility in potato remains elusive. In diploid offspring, large-effect deleterious mutations will result in gametic or zygotic selection and can be identified by studying transmission ratio distortion (TRD) (Zhang et al., 2019). The removal of large-effect deleterious alleles identified via TRD is central to the strategy of hybrid potato genome design proposed by (Zhang et al., 2021). This type of genetic analysis relies on genotyping large segregating populations in a cost-effective way. With the low cost of next-generation sequencing, genotyping-by-sequencing (GBS) strategies are an interesting option. GBS strategies can include a step of genome complexity reduction using restriction enzymes (Elshire et al., 2011), capture baits (Uitdewilligen et al., 2013), or amplicon sequencing like potatoMASH (Leyva-Pérez et al., 2022). Owing to the development of multiplex sequencing (Cronn et al., 2008), methods based on whole genome sequencing at low depth, also known as skim sequencing, have become increasingly popular. However, the low read depth makes robust genotyping challenging for full-sib populations from non-inbred parents. The recently released OutcrossSeq pipeline (Chen et al., 2021) is addressing this issue by local clustering of incomplete sequence variant data to infer genetic markers.

In this study, we implemented the core idea of the OutcrossSeq pipeline for diploids in the R package OutcrossSeqDiploidR and applied it to a skim sequenced diploid potato population of 1536 individuals. We explored the minimal read depth necessary for marker inference and used the resulting markers for the construction of high-density parent specific linkage maps, highlighting variation in parental recombination rate and structural variations on chromosome *3* and *10*. Subsequently, we detected nine gametic and one zygotic TRD, as well as four pairs of loci with non-independent assortment of alleles. We identified a major QTL for tuber yield from seedlings in pots co-localizing with the position of St*CDF1* and 5 minor effect QTLs epistatic to *StCDF1* allele combinations. Finally, we identified seven QTLs regulating pollen shed with the largest one located on chromosome *2*.

Materials and Methods

Plant materials

A population of 1536 diploid full-sib seedlings, descending from a cross between two heterozygous potato clones C (USW5337.3) and E (77.2102.37), was sown the 3rd of July 2020. Before transplanting, about one quarter of the seedlings with a crumpled phenotype (Jongedijk et al., 1990) was discarded from the mapping population. A few crumpled seedlings were kept for further analysis of the phenotype. Plants were raised in 19 cm pots in a greenhouse at ambient temperatures (~18 °C) and under natural daylight (~16-13 h). This population is a backcross (clone C is the female parent of E) with mixed ancestry of *Solanum tuberosum* Group Tuberosum and Phureja, and *S. vernei*.

DNA extraction, library preparation and sequencing

Young leaf material from the 1536 CE-XW full-sibs were collected on ice in 96 deep well plates and stored at -20 °C. Genomic DNA extraction was performed at VHL Genetics (Wageningen, NL) and DNA concentration was normalized to 12.5 ng/µl. Sequencing libraries were produced using RipTide DNA library prep kit (iGenomX, Carlsbad, CA) following manufacturer instructions. Briefly, individual samples were labelled in sixteen 96-well plates using well and plate specific barcoded random primers. Once labelled, samples in each plate were pooled together and converted into a NGS library in one single tube. Libraries of each plate were pooled together and sequenced by Limes Innovations B.V (https://limes-innovations.com) with 150bp paired end reads using Illumina iSeq 100. The parental clones were sequenced at BaseClear (Leiden, NL) with 100bp paired-ended reads using Illumina Hiseq2000. The sequencing data generated are available from the European Nucleotide Archive (ENA) under the BioProject ID PRJEB56778.

Demultiplexing, quality check and alignment

DNA sequences were demultiplexed to individual descendants using the plate and position barcodes with fgbio version 1.4.0 (https://fulcrumgenomics.github.io/fgbio/). Sequence reads of parents and offspring were quality trimmed using fastp version 0.19.5 (Chen et al., 2018) dropping reads with a complexity lower than 20% and a length lower than 70bp (option -l 70 -y 20 -5 --cut_front_window_size 1 -3 --cut_tail_window_size 1). Trimmed reads were aligned to the potato reference genome DM v6.1 using BWA-MEM algorithm v.0.7.17 (Li and Durbin, 2009) with default parameters. Alignment summary statistics were extracted with qualimap v.2.2.2 (Okonechnikov et al., 2015).

Variant calling

Variant calling was performed in the parental clones using bcftools v.1.13 (Danecek et al., 2021) mpileup and call functions and filtered based on quality and depth (option -e '%QUAL<20 || FORMAT/DP>40 || FORMAT/DP<10). Parental bcf files were merged and indels removed using bcftools view to create a target SNP file with bcftools query (option -f '%CHROM\t%POS\t%REF,%ALT\n'). This target file was used to constrain variant calling in each offspring sample to these target coordinates and alleles using bcftools call -T and -C options. Variant calling in the offspring was parallelised by chromosome. Multiallelic sites were split into multiple rows using bcftools norm -m and the resulting chromosomal bcf files were converted into chromosomal variant matrices, filtering out variants with a missing rate above 0.5 using VCFtools v0.1.16 (Danecek et al., 2011) (options --max-missing 0.50 --012).

Clustering of adjacent sequence variants into robust marker genotypes with OutcrossSeqDiploidR

The R package OutcrossSeqDiploidR (https://github.com/ccrclot/OutcrossSeqDiploidR) was used to cluster offspring, which co-inherited the same heterozygous and private sequence variants of either parent in windows of 0.1 Mb. Based on cluster membership, genotypes were inferred resulting in potentially 7312 intervals of 0.1Mb for each parent hereafter referred to as 'bin-markers'. Our R package follows the core idea of the OutcrossSeq pipeline for diploids (Chen et al., 2021), relying on a reference genome to locally infer genotypes from a population kinship matrix. However, in this implementation, female and male variants are clustered independently to estimate uniparental genotypes rather than the integrated genotyping information produced by the original pipeline. In addition, a quality check function was implemented allowing users to flag bin-markers with aberrant recombination or transmission ratio values. Recombination percentages above 50% allowed to bring all (initially randomly assigned) bin-markers alleles from the same parental chromosome into linkage phase. All steps are summarized in Fig. 1 and in the package vignette. Firstly, the function

recodeParentMatList was used to 1) recode variant dosage into presence absence data, since artefactual homozygous absence could be derived from low coverage information on a heterozygous site, and 2) extract heterozygous variant privates to each parent (1x0 and 0x1) into separate female and male specific binary matrix. The following steps of OutcrossSeqR pipeline were applied independently to chromosomal female and male matrices. The function makeBins with parameter bin_size = 100000 was used to group variants in physical bins of 100kb. Within each bin, the Euclidian distance between individuals was calculated based on presence absence of variants with the function getDist. The resulting kinship matrix was then used to cluster the individuals using the ward.D2 method with the function clusterPop. The function getHap with parameter k = 2 was used to extract, from the clustering data, two groups of individuals per bin-marker, corresponding to the two alleles that could be inherited form one parent. The function qualCheck with default parameters was then used to flag noisy bin-markers displaying 1) high recombination rates with adjacent bins ($0.3 < r < 0.7$) and 2) deviations exceeding 0.15 points from the local transmission ratio, calculated with five flanking markers at either side. Flagged bin-markers were removed. Bin-markers data with $r > 0.5$ were swapped to phase adjacent bin-markers with the function phaseHap. Phased bin-markers were visualized, and phasing was manually improved, when necessary, with the function manualPhasing before being formatted for Smooth Descent (Thérèse Navarro et al., 2023) with the function formatSD.

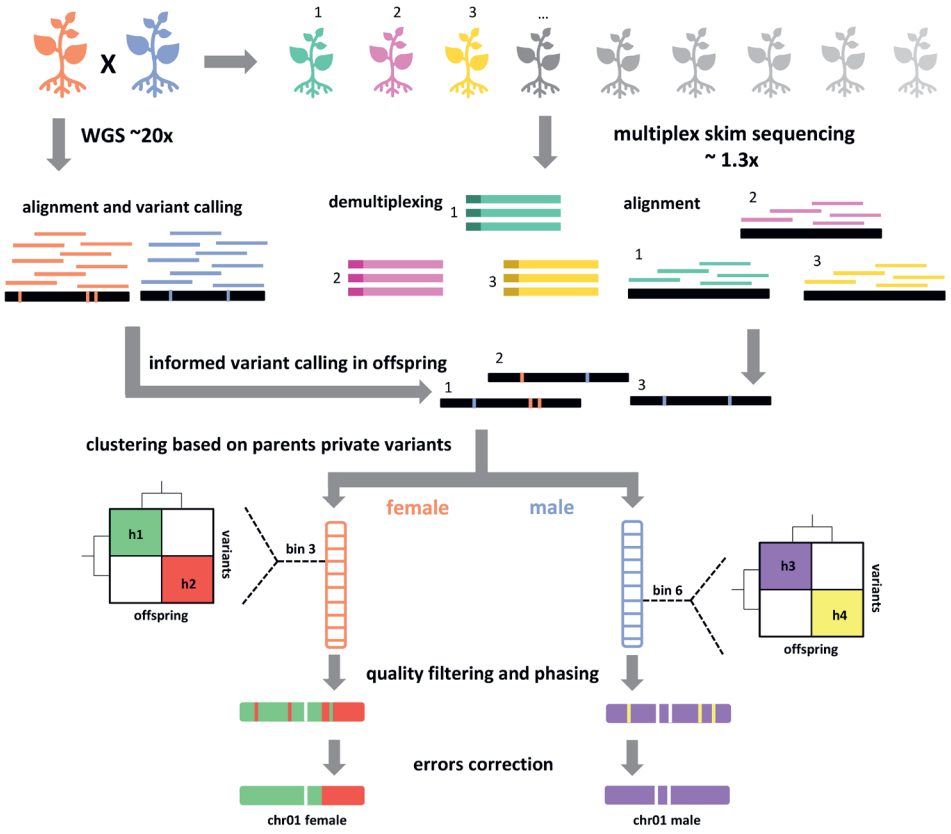

Figure 1: Graphical summary of the OutcrossSeqDiploidR pipeline

Correcting genotyping errors with Smooth Descent

The segregation data of bin-markers obtained with OutcrossSeqDiploidR were corrected for putatively erroneous data points with the algorithm Smooth Decent following the package vignette (Thérèse Navarro et al., 2023). Briefly, Smooth Decent makes use of identity-by-descent probabilities, in our case based on the physical order of bin-markers, to detect putative genotyping errors and impute the most probable genotype given the data points of flanking markers. A total of seven iterative rounds of data improvement with Smooth Descent were performed. The first five rounds were based on a prediction interval of 1 Mb and the two final rounds on a prediction interval of 5 and 10 Mb.

Linkage map construction

The cleaned bin-marker data were used to generate independent male and female linkage maps with polymapR version 1.1.2 (Bourke et al., 2018) following the package vignette. Recombination frequencies were calculated for all pairs of bin-markers. Bin-

markers were assigned to 12 chromosomal linkage groups of either parent, based on their physical positions on DM v6.1 reference genome. Next, bin-markers were ordered using MDSmap_from_list, a wrapper function around the estimate.map function from MDSMap (Preedy and Hackett, 2016). During the mapping process, 47 outlying bin-markers with a high nearest-neighbour fit score or an abnormal position in the principal curve analysis were removed.

Chromosomal recombination rate and Marey map

The chromosomal recombination rate (cM/Mb) was calculated using the genetic length of a linkage map and the physical length of the corresponding chromosomes of DM v6.1 assembly. Similarly, Marey maps (Chakravarti, 1991) were derived from the comparison between the physical and genetic position of each marker. Finally, the positions of CENH3-binding domains on DM v6.1 were obtained from(Pham et al., 2020).

Transmission ratio distortion and non-independent assortment of alleles

Male and female bin-markers were tested for significant deviation from the expected mendelian 1:1 ratio via a χ^2 test with 1 degree of freedom and a p-value for significance set at 0.01. Following the method used by Dukić and Bomblies (2022), we corrected for multiple testing in 23 chromosome arms and utilized a significance threshold of $\alpha = 0.01/23 \approx 0.0004$ ($\chi^2 = 12.4$, df = 1). A region with transmission ratio distortion (TRD) was considered biologically meaningful when markers displayed significant TRD over at least 2 Mb. Using the same threshold for significance, TRD due to zygotic rather than gametic selection were detected with a χ^2 test based on a 2×2 contingency table where the observed frequencies of one of the four possible haplotype combinations were compared with the expected frequencies based on random assortment of the observed parental haplotypes frequencies.

During the construction of the linkage map we detected associations of bin-markers alleles located on different linkage groups. We decided to explore those deviations from independent assortment of alleles between all marker pairs using the G-test option implemented in the function linkage of ploymapR. Like the zygotic χ^2 test, this option calculates the G^2 statistic for independence based on a 2x2 contingency. Instead of returning a G^2 value, this function returned a LOD score of independence and defined as $\frac{G^2}{\log(100)}$. We considered that bin-markers pairs with a LOD of independence > 5 were displaying biologically significant deviations from independent assortment.

Phenotyping seedling yield in pots and pollen shed

Phenotypic data on flowering and pollen shed was collected from the seventh to the tenth week of the growing season. The ability to produce at least one flower reaching anthesis was recorded and resulted in a binary classification. A pollen sample (extracted with an electric toothbrush from a freshly open flower at anthesis) was collected for each flowering seedling. If the pollen extraction of the first collected flower failed, the plant was resampled. When only a few pollen grains could be extracted after resampling of four flowers on two different days the plant was classified as a poor pollen shedder, resulting in a binary classification for pollen shed (poor vs. good). After a growing cycle of 18 weeks, three quarter of the population had senesced and tuber harvest started. Over a period of two weeks, tubers larger than 5 mm were harvested in one paper bag per descendant. The fresh weight in gram was measured with an electronic scale and used as yield per pot estimate.

QTL mapping

QTL mapping was performed using the package polyqtlR version 0.0.6 (Bourke et al., 2021). The function singleMarkerRegression was used to fit an additive model at each marker position returning the $-\log_{10}$ p-value of model fit per marker. The significance thresholds for QTL detection were determined via permutation tests on the phenotypic values with N = 1000 cycles and $\alpha = 0.05$. QTL discovery was based on the separate maternal and paternal maps, but at significant QTL positions we analysed the phenotypic effects of the combined male and female haplotypes. This was done with a one-way ANOVA with multi-comparison Tukey's HSD post hoc test ($\alpha = 0.05$). To estimate the effects size of QTLs and their potential interactions, all significant QTLs, and their interactions if significant, were used as factor in (generalized) linear models predicting for the phenotype of interest. While a linear regression was used for tuber yield, we used a logistic regression for the binary trait pollen shed and estimated effect size with odds ratio.

Results

Analysis of sequencing data

The sequencing of the libraries comprising the 1536 seedlings yielded 3,364 gigabases (Gb) which after demultiplexing resulted in 2,910 Gb. After trimming and quality filtering we obtained 19,766,020,170 usable reads corresponding to a total of 2,562 Gb, which corresponds to an average of 1.7 Gb per seedling. A proportion of 93.6% of reads could be mapped to DM v6.1 generating a median coverage per individual of 1.34x, or 0.67x per haploid genome (Suppl. Fig. 1). A total of 18 Gb and 22.6 Gb were obtained for the parental clone C and E. After trimming and quality filtering we obtained 149,677,752 and 182,224,166 usable reads corresponding to a total of 14.8 Gb and

18 Gb. A proportion of 98.7% and 98.6% of reads could be mapped to DM v6.1 generating a coverage of 18.72x for clone C and 22.47x for clone E or 9.36x and 11.24x per haploid genome. After variant calling and quality filtering in the parents, we obtained a total of 12,983,013 variants which were subsequently called in the offspring. Only, a subset of 1,889,749 variant could be observed in more than 50% of the offspring. Knowing that the low read-depth will yield erroneous dosage information, we focused on heterozygous parent-private variant (1x0 and 0x1). A total of 554,806 and 902,714 were heterozygous and private to clone C and E respectively. The distribution of heterozygous private variants was not uniform across parents and chromosomes (Fig. 2a). For example, on chromosome *7* the female parent C showed a 5-fold higher number of heterozygous parent-private variants than the male parent E. For chromosomes *3, 6 and 11* the reverse was observed, where the male parent was more polymorphic (Fig. 2b). Regions where parents differ in variant density tend to coincide with the location of the pericentromeric heterochromatin (Suppl. Fig. 2). We concluded that linkage map construction based on segregating sequence variants was not possible. The low sequencing depth disallows dosage estimates to distinguish between homozygous or heterozygous offspring. The presence of heterozygous parent-private variants can be scored, but their absence can be *bona fide* (due to inheritance) or the result of low coverage information. This motivated us to use the presence/absence information of all co-inherited heterozygous parent-private variants from a 0.1Mb interval to infer the genotype score of this interval as a so-called bin-marker. As a result, we generated sparser but more robust genotyping data.

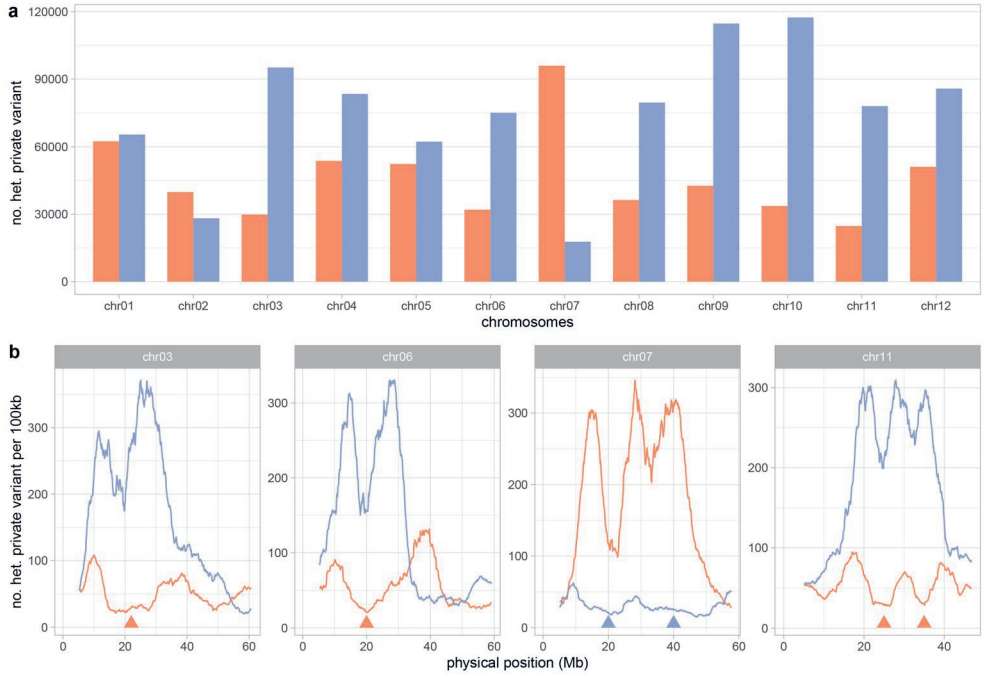

Figure 2: Distribution of heterozygous parent-private sequence variant. a) Total number of parent-private variant per chromosome per parent. B) 5 Mb moving average of the number of heterozygous parent-private variants across bins of 100 kb. Clone C data are displayed in orange and clone E in blue. The triangles indicate regions with a low density of heterozygous parent-private variants disallowing the inference of bin-markers.

Inference of bin-markers with OutcrossSeqDiploidR

The DM v6.1 reference genome was subdivided in 7312 bins of 100 kb to infer bin-markers. The parent-private variants within these bins were used to cluster descendants and to infer the segregating alleles of each bin-marker. Regions with a low density of parent-private variants, in combination with the low sequencing depth, lacked information to infer bin-marker genotypes. This resulted in larger chromosomal intervals without genetic markers. The regions with low density of parent-private variants are plotted in Fig. 2b, and the impact on bin-marker imputation and thus genetic map coverage is shown in Table 1 and Fig. 3. We arrived at 7145 female and 7174 male bin-markers, which had at least one heterozygous parent-private variant. After all subsequent cleaning steps (see Materials and methods), we obtained 4935 female and 4746 female bin-markers with genetically and biologically sound characteristics. Although this represented a ~34% reduction of markers, this predominantly affected marker coverage at pericentromeric regions.

Smooth Descent improved genotyping data

Smooth Descent detected and imputed 1,075,417 erroneous datapoints, corresponding to 7.2% our bin-marker dataset. Data inspection showed that Smooth Descent was unable to clean noise in several plants because of a too unfavourable signal to noise ratio due to low sequencing depths. To identify a threshold to reject such plants we compared sequence coverage with the amount of noise. The expected number of true genetic recombination events per plant is low. One erroneous data point suggests two extra recombination events relative to the flanking markers. When counting all 'recombination events' per plant from a noisy dataset we reached values, shown in Suppl. Fig. 3, so much inflated by error (exceeding 500), that we can use the number of 'recombination events' to indicate noise. Unsurprisingly, the amount of recombination (noise) was not randomly distributed across individuals but increased with decreasing genome coverage. Smooth Descent corrected more erroneous datapoints in plants with low coverage, but this ability reached a tipping point and started to decrease around an average diploid coverage of 0.3x (Suppl. Fig. 3). The 75 individuals with a coverage below 0.3x were removed. The remaining 1461 individuals were used in subsequent analyses.

High density linkage maps

After filtering 47 outlying bin-markers during the mapping process, a total of 4894 male and 4740 female bin-markers segregating across 1461 individuals could be mapped on 24 linkage groups. Those linkage groups corresponded to the 12 paternal and 12 maternal chromosomes (Table 1). On the female side, the average number of bin-markers per linkage group was 408 ranging from 257 for chromosome *11* to 549 for chromosome *1*. On the male side the average of 395 bin-markers per linkage group was comparable, with a minimum of 99 bin-markers for chromosome *7* and a maximum of 523 for chromosome *9*. The marker density and gaps (up to 16.9 cM) in the genetic maps corresponded to chromosomal regions with a low density of parent-private variants (Fig. 2b). In addition, extreme Transmission Ration Distortion (TRD) also hampered to infer bin-markers, resulting in gaps on female and male chromosome *1* of 6.4 and 8.9 cM, respectively. Nonetheless, both parental maps were of high density with an average interval between bin-markers of 0.3 cM on the female and 0.2 cM on the male maps.

Table 1: Overview of genetic lengths and marker densities per linkage group

	female (C)				male (E)			
LG	**Total no. markers**	**Map length (cM)**	**Average distance (cM)**	**Max gap (cM)**	**Total no. markers**	**Map length (cM)**	**Average distance (cM)**	**Max gap (cM)**
chr01	549	133.2	0.3	9.8	194	106.0	0.6	6.4
chr02	289	93.3	0.4	7.8	344	59.5	0.2	4.1
chr03	332	107.1	0.3	5.6	474	72.5	0.2	9.1
chr04	552	109.6	0.3	4.9	540	69.3	0.2	1.9
chr05	402	92.3	0.3	8.3	426	57.7	0.2	2.9
chr06	275	87.2	0.3	5.3	390	63.3	0.2	6.6
chr07	399	87.9	0.2	9.8	99	68.5	0.7	16.9
chr08	454	80.7	0.2	4.2	410	69.7	0.2	7.3
chr09	525	97.1	0.3	4.4	523	65.4	0.1	2.9
chr10	365	81.2	0.3	10.6	493	47.7	0.1	3.5
chr11	257	81.3	0.3	2.9	404	49.8	0.1	1.3
chr12	495	107.3	0.4	5.0	443	54.3	0.1	7.0
all	**4894**	**1158.2**	**0.3**	**10.6**	**4740**	**783.8**	**0.2**	**16.9**

Structural variation and chromosomal recombination rate variation

Comparing the physical bin-marker positions, using DM v6.1 reference genome, with the genetic marker positions on parent specific linkage maps, we could estimate an average genome wide recombination rate of 1.58 cM/Mb for the female clone C and of 1.07 cM/Mb for the male clone E, and this genome-wide difference was also true for every chromosome (Table 2). In both parental maps, the lowest chromosomal recombination rate per Mb was observed on chromosome *10* and the highest one on the acrocentric chromosome *2* with values ranging from 1.33 cM/Mb to 2.02 cM/Mb for clone C and, 0.78 to 1.29 for clone E. As expected, the recombination rate was far from uniform along each chromosome with virtually no recombination in pericentromeric regions as evidenced by the horizontal sections around the positions of CENH3-binding domains, indicated with triangles, on the Marey maps (Fig. 2). In addition, two euchromatic regions without recombination were observed in clone C. On the south arm of chromosome *3* a segment of 5.8 Mb from 42.9 to 48.7 Mb was found, and 6.3 Mb segment from 52.8 Mb to 59.1 on chromosome *10*. In clone E, this chromosome *3* segment was inverted relatively to DM suggesting the presence of a homozygous inversion. In clone C this inversion must be heterozygous, explaining the suppression of recombination. In addition, the north arms of chromosomes *3* and *6* of clone E displayed a strong reduction of recombination.

Table 2: Estimated recombination rates for each linkage group

		female (C)		male (E)	
LG	**Physical length**	**Genetic length**	**Rec. Rate (cM/Mb)**	**Genetic length**	**Rec. Rate (cM/Mb)**
chr01	88.6	133.2	1.50	106.0	1.20
chr02	46.1	93.3	2.02	59.5	1.29
chr03	60.7	107.1	1.76	72.5	1.19
chr04	69.2	109.6	1.58	69.3	1.00
chr05	55.6	92.3	1.66	57.7	1.04
chr06	59.1	87.2	1.47	63.3	1.07
chr07	57.6	87.9	1.52	68.5	1.19
chr08	59.2	80.7	1.36	69.7	1.18
chr09	67.6	97.1	1.44	65.4	0.97
chr10	61.0	81.2	1.33	47.7	0.78
chr11	46.8	81.3	1.74	49.8	1.06
chr12	59.7	107.3	1.80	54.3	0.91
all	**731.3**	**1158.2**	**1.58**	**783.8**	**1.07**

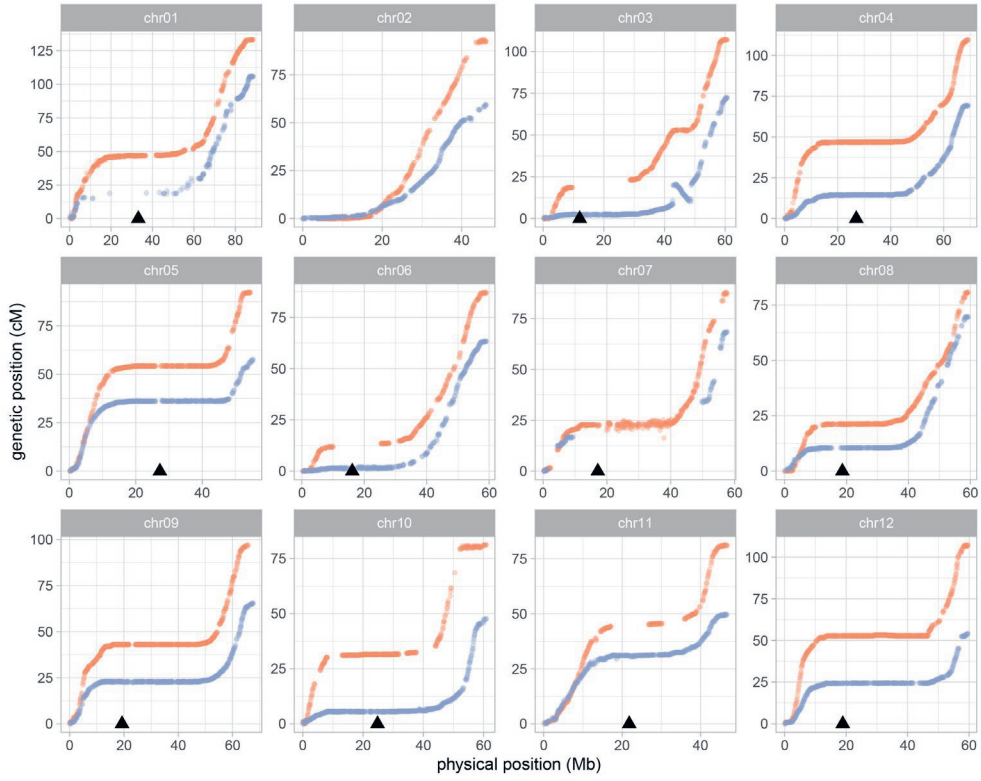

Figure 3: Marey maps in which genetic distance (cM) is plotted over physical distance (Mb). Clone C data are displayed in orange and clone E in blue. Triangles indicate the position of CENH3-binding domains on DM v6.1.

Transmission ratio distortion

We detected significant gametic TRD (α = 0.0004) at loci located on eight different chromosomes (Fig. 4). On the female map four loci displayed significant TRD. The locus with the largest deviation (~1:6) was identified on chromosome *1* (distortion clone C chromosome 1, abbreviated as *DC1*), at 69.65 Mb. Loci with milder TRD were identified on chromosomes *4* (*DC4* at 58.65 Mb) and *12* (*DC12* at 9.35 Mb). A locus with weak but significant TRD was also detected in the pericentromeric region of chromosome *5*. On the male map, six loci displayed significant TRD. Again, the locus with the largest deviation was detected on chromosome *1* (*DE1*), but its localisation in or near the pericentromeric heterochromatin is inaccurate due to the inability to infer bin-markers. On chromosome *11* another extreme TRD locus (~9:1) was mapped at 6.55 Mb (*DE11*). Except for the last 10 Mb the entire chromosome *8* showed mild TDR. This could be due to the presence of two TRD loci at either side of the pericentromeric region as suggested by the mild local maxima observed on this chromosome (*DE8a* at 4.45 Mb and *DE8b* at 44.55 Mb). Finally, significant but weak TRD regions were identified on chromosome 2 (32.05 to 34.45 Mb), the short arm of chromosome 4 and the peri-centromeric regions of chromosome 9. The loci with strong TDR coincide with previously described loci such as the *S*-locus and *ar1* for *DE1* and *DC1* on chromosome *1* and *la1* for chromosome *8* (Jacobs et al. 1995; Zhang et al. 2019; see discussion). We then investigated if those TRD could be caused by zygotic rather than gametic selection. Only one case of zygotic selection was observed on chromosome *1* at locus *DZ1* (distortion zygotic chromosome *1*) located at 69.65Mb. *DZ1* co-locates with *DC1* at 69.65Mb on the female map, but the localisation of the male counterpart is blurry because the gametic selection in the pericentromeric region (*DE1*) adds to the zygotic selection around 69.65Mb (Suppl. Fig. 4). We postulate that the near absence of one male by female allele combination is the result of selection against about one quarter of the seedlings with the crumpled phenotype (Suppl. Text 1), which were excluded beforehand from analysis.

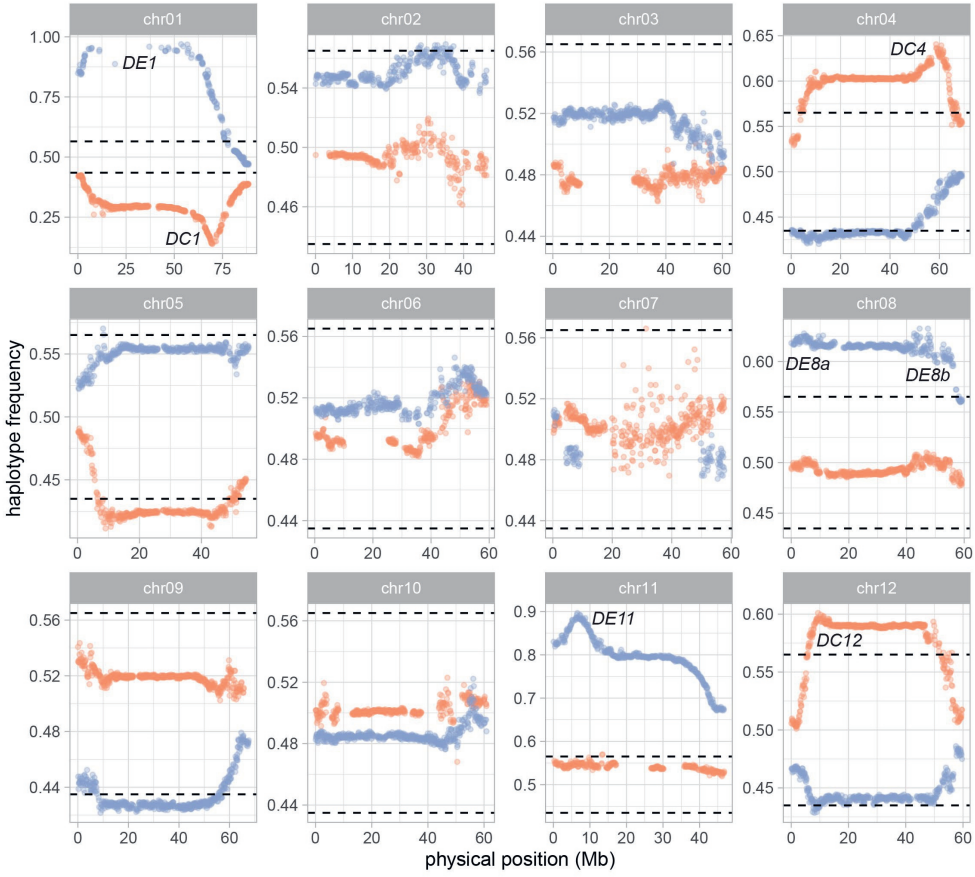

Figure 4: Segregation ratio for haplotype h1 in C and h3 in E plotted over physical distances (Mb) for each chromosome. Thresholds of significance (black dashed line) were determined with χ^2 test with $\alpha = 0.0004$. For all panels, clone C data are displayed in orange and clone E in blue.

Non-independent assortment of alleles

Following the detection of a zygotic TRD, resulting in non-independent assortment of alleles at one locus, we investigated deviations from independent assortment between all bin-marker pairs and observed biologically significant deviations (LOD of independence >5) on three female and five male chromosomes. As expected, male and female bin-marker alleles around 69.6 Mb to 70.7 Mb on chromosome *1* displayed extreme deviation from independent assortment due to the zygotic TRD *DZ1* (Fig. 5b). A mild deviation from independent assortment was identified between marker pairs from female chromosomes *4* and *12* around the position of *DC4* and *DC12* (Fig. 5c). Another association of comparable significance was identified between most of male chromosome *3* (from the low recombining north arm to the south arm inversion) and the pericentromere of chromosome *9* already associated with *DE9* (Fig. 5d). Finally, two regions near *DE11* on male chromosome *11* showed a highly significant deviation

from independent assortment of alleles: one from 5.3 Mb to 6.6 Mb interacting with most of male chromosome *3* and the other from 7.6 Mb to 8.2 Mb interacting with the *DE8a* region on male chromosome *8* (Fig. 5e). While close, the LOD-2 confidence interval of those two regions on chromosome *11* did not overlap suggesting the presence of two separate, albeit linked loci showing these associations.

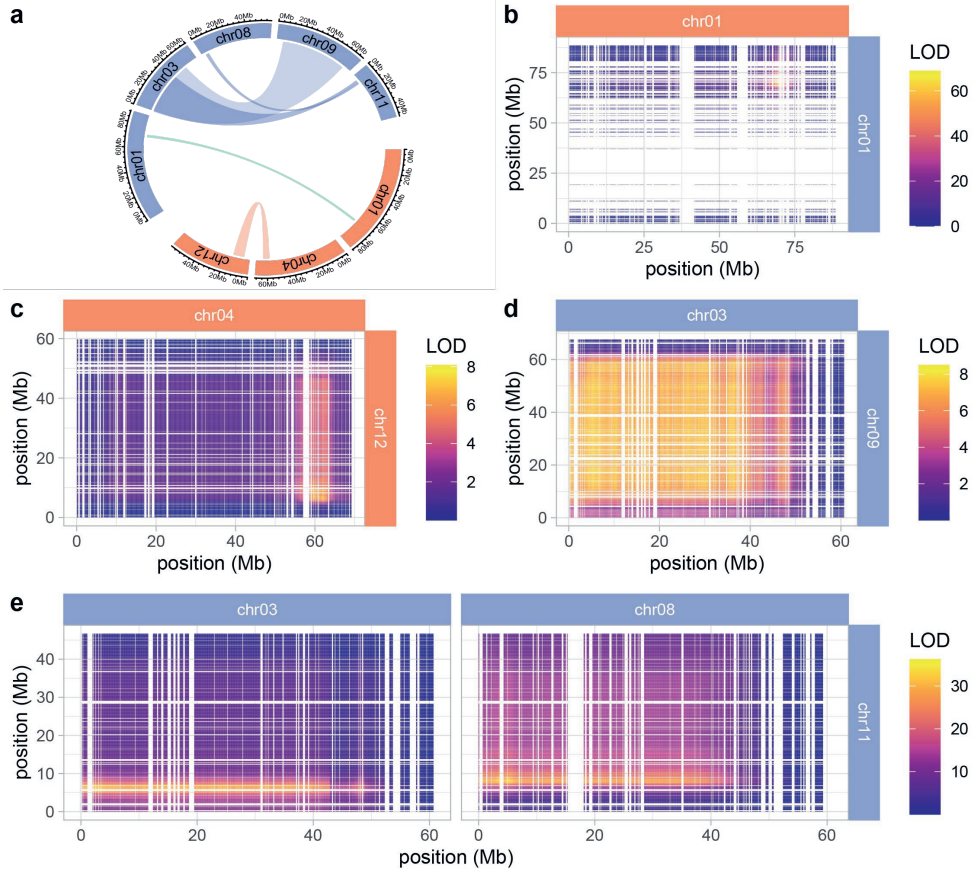

Figure 5: a) Overview of pairs of loci showing non-independent assortment of alleles. Ribbons link non-independently associated chromosomal region (LOD-2 confidence interval). Pair-wise non-independent assortments occurring between parents are depicted in green the one occurring within parent are depicted in orange for C and in blue for E. **b-e)** Heatmap displaying the LOD of independence of chromosomes with biological significant non-independent assortment of alleles (LOD >5) against physical position.

QTL mapping for seedling yield in pots

Phenotypic data for tuber yield in pots, defined as fresh weight of tubers larger than 5 mm, were collected on 1534 clones after a growing cycle of 18 weeks. Tuber yield ranged from 0 g (no tubers above 5 mm) to 213.1 g with a mean value of 69.1 g (Fig. 6a). Three significant QTLs were identified for tuber yield (TY): the minor QTL *TYE8* (QTL for tuber yield clone E chromosome *8*) and the major QTLs *TYC5* and *TYE5* (Table 3, Fig. 6b).

The LOD-1.5 confidence interval around the top marker of *TYC5* and *TYE5* overlap with the position of the candidate gene *StCDF1.* In this population, alleles *StCDF1.1* and *StCDF1.2* are known to segregate from clone C and alleles *StCDF1.1* and *StCDF1.3* from clone E (Kloosterman et al., 2013). Interestingly, the significance for QTL discovery differs vastly between parents (*TYE5* -$\log_{10}$(*p-value*)=118.7; TYC5 -$\log_{10}$(*p-value*)=23.7), indicating a larger phenotypic effect on tuber yield by the *StCDF1.3* allele. This observation is better understood when the effects on tuber yield of the four *StCDF1* haplotype combinations are compared (Fig. 6c). Both *StCDF1.2* and *StDCF1.3* haplotypes contributed to TY, but with unequal effect (Tukey's HSD post hoc test α = 005). The large-effect *StCDF1.3* haplotype was dominant over the minor-effect *StCDF1.2* haplotype. The deviation from additivity was highly significant (*p*<0.001). Overall, in this population 44.7% of variation in tuber yield can be explained by variation in haplotype combination at the *StCDF1* locus.

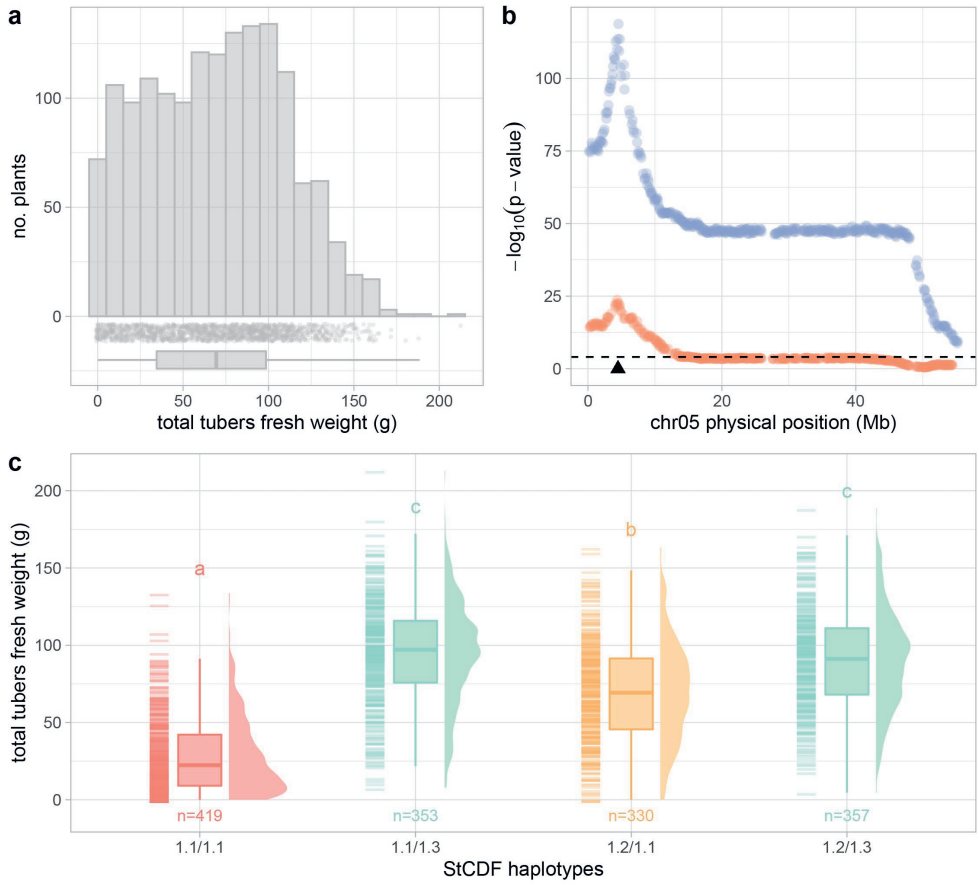

Figure 6: a) Distribution of tuber yield (TY) (n=1459). **b)** Localisation of significant QTLs for tuber yield on the male and female map of chromosome *5*. The X axis represents physical position (Mb), the Y axis represents -$\log_{10}$(*p*-value) and threshold of significance is indicated by the black dashed line. Clone C data

are displayed in orange and clone E in blue. The triangle indicates the position of *StCDF1* on DM v6.1. **c)** Raincloud plots illustrating the effect of *StCDF1* haplotype combinations on TY. The different letters and colours indicate haplotype combinations with significantly different tuber yield (Tukey's HSD post hoc test $\alpha = 0.05$).

To cancel the overwhelming effect of the *StCDF1* locus, we performed QTL discovery for tuber yield within three subpopulations with equal maturity, corresponding to cohorts of plants homozygous for *StCDF1.1*, the *StCDF1.1* / *StCDF1.2 heterozygotes*, and offspring with the *StCDF1.3* allele. Now five minor-effect QTLs surfaced, two in the low yielding *StCDF1.1* homozygotes, one in the medium yielding 1.2/1.1 cohort and two in the high yielding cohort with *StCDF1.3* (Table 3; Suppl. Fig. 5). Interestingly, each sub-population resulted in new QTL positions. For example, QTL *TYC1* explained 5% of tuber yield variation in the 1.2/1.1 cohort but did not account for a significant difference in the other two cohorts. Similarly, clone C chromosome *7* bears on each side of its centromere QTLs *TYC7a* and *TYC7b* specific to the 1.1/1.1 and -/1.3 cohort respectively. These observations showed that tuber yield was not only controlled by *StCDF1* allelic combinations, but that these allelic combinations also epistatically controlled how minor QTLs can exert their effects.

Table 3: Summary of QTLs identified for yield in pots in the entire population and in maturity sub-populations.

QTL id	panel	chr	parent	LOD score	R2adj	top bin-marker (Mb)	LOD-1.5 CI (Mb)
TYC5	entire offspring	chr05	C	23.7	0.07	4.35	4 - 4.8
TYE5	entire offspring	chr05	E	118.7	0.31	4.55	4.4 - 4.6
TYE8a	entire offspring	chr08	E	4.6	0.01	40.85	0 - 45.5
TYE2	1.1/1.1 cohort	chr02	E	4.2	0.04	34.75	31.5 - 40.9
TYC7a	1.1/1.1 cohort	chr07	C	4.3	0.04	43.25	12.4 - 48.7
TYC1	1.2/1.1 cohort	chr01	C	5.0	0.05	3.25	0.7 - 4.7
TYC7b	-/1.3 cohort	chr07	C	4.9	0.03	29.95	5.6 - 42.5
TYE8b	-/1.3 cohort	chr08	E	4.9	0.03	42.15	11.4 - 45.7

QTL mapping for pollen shed

During the growing season most plants flowered profusely, except 57 offspring plants that did not flower or dropped their buds. Using the ability to produce flowers as binary trait, we could not identify any significant QTL regulating flower production. Out of the remaining 1479 flowering plants, 134 were classified as poor pollen producers and 1345 as good pollen producers. Using this binary classification as a phenotype we identified seven significant QTLs regulating pollen shed (Fig. 6a, Table 4). *PSE2* (QTL pollen shed clone E chromosome *2*), a QTL with a LOD score of 14.6, was identified on clone E chromosome *2* located at around 37.65 Mb. QTLs with LOD scores ranging

from 4.5 to 7.7 were identified on C chromosomes *1*, *5*, *10*, and *11* and on E chromosomes *1* and *9* and are detailed in Table 4. The associations between pollen shed and segregating marker alleles are presented as contingency tables visualised as mosaic plots in Fig. 6b. Holding the effect of other QTLs constant, the odds of being a low pollen shedder were 8.1 times higher for the plants bearing the detrimental allele at *PSE2*. We therefore considered *PSE2* as a major QTL controlling pollen shed in this population.

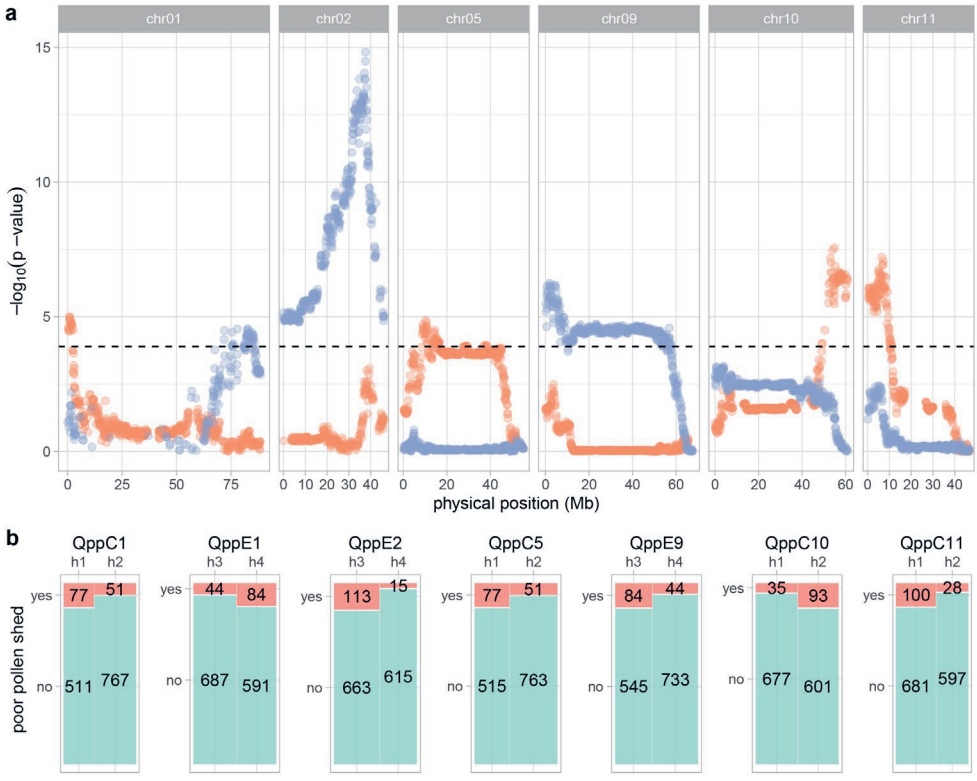

Figure 7: a) Location of significant QTLs for pollen shed (n=1406). The X axis represents physical position (Mb), the Y axis represents -$\log_{10}$(p-value) and threshold of significance is indicated by the black dashed line. Clone C data are displayed in orange and clone E in blue. b) Mosaic plots illustrate the effect of the different haplotypes on the poor pollen shedder's classification at QTL positions. Counts of poor pollen shedders are displayed in red and counts of good pollen shedders in green.

Table 4: Summary of QTL identified for pollen shed.

QTL id	chr	parent	LOD score	Odds Ratio	top bin-marker (Mb)	LOD-1.5 CI (Mb)
PSC1	chr01	C	5.0	2.2	0.85	0.2 - 2.7
PSE1	chr01	E	4.5	2.9	82.75	67.5 - 86.8
PSE2	chr02	E	14.6	8.1	37.65	33.6 - 38.1
PSC5	chr05	C	4.9	2.6	10.25	4.1 - 45.2
PSE9	chr09	E	6.6	2.7	1.95	0.1 - 9.0
PSC10	chr10	C	7.7	2.6	54.85	52.3 - 60.9
PSC11	chr11	C	7.3	3.0	6.55	0.4 - 9.1

Discussion

OutcrossSeqDiploidR bin-markers inference

This study explored the construction of linkage maps of skim sequenced offspring at a very low depth. This resulted in elevated missing values at SNP loci which prohibited linkage analysis. However, grouping of parent-private SNPs in windows of 0.1 Mb allowed clustering of descendants to impute bin-markers. Performance of OutcrossSeqDiploidR was hampered at specific regions with either high or low sequence divergence. High sequence divergence of the parents relative to the DM reference genome results in failure of read mapping. Low sequence divergence among the parental alleles resulting in less than ~30 parent-private variant per bin of 100kb, also excluded accurate imputation of bin-markers. Such regions were observed on maternal (clone C) chromosomes *3*, *6*, and *11* and on paternal (clone E) chromosome *7*. Bin-markers from such incongruent or nearly invariant regions were recognised as inconsistent and removed by subsequent quality filters. The loss of one third of the potential number of bin-markers may have little impact on our results, because most markers were lost in the non-recombining heterochromatic regions and potentially in truly homozygous regions. However, we occasionally noticed a small detrimental effect on the precision of QTL mapping when bin-markers were lost in regions of absolute TRD. Contrary to the original OutcrossSeq pipeline, we did not choose to replace the flagged bin-makers with values of adjacent bin-markers, so it remains clear to users where genotyping failed.

In this study we explored the minimal depth of skim sequencing to infer genetically consistent bin-markers of 0.1Mb. We used the sum of recombination events between physically adjacent marker-bins as a metric to monitor the performance of Smooth Descent to correct spurious datapoints. We determined that our R package could not generate genetically consistent bin-markers data when the diploid genome coverage was below 0.3x. This agrees with the test results of the diploid outcrossing module of

the original OutcrossSeq pipeline which also showed increasing error rates ranging from 1.29% to 14.05% for decreasing coverage ranging from 4x to 0.5x. While agreeing with Chen et al. (2021) who recommend a minimum coverage of 1x to ensure sufficient genotyping power, we show that in a diploid potato population of 1536 individuals, a coverage lower than 1x but higher than 0.3x can still be used for 100 kb bin-marker inference by OutcrossSeqDiploiR, in combination with the algorithm Smooth Decent to correct genotyping errors.

Comparison of genetic and physical maps

Bin-markers were used to construct separate maternal and paternal linkage maps. The quality and specific features of these maps were evaluated by comparing the genetic and physical bin-marker position with Marey maps. This allowed us to visualize an approximately 5.8 Mb paracentric inversion on the long arm of E chromosome *3* (42.9 - 48.7 Mb) relative to the DM reference genome. While parent E is homozygous for this inversion, parent C must be heterozygous, because the same ~5.8 Mb interval did not show recombination events. The same inversion was recently reported in a diploid potato pan-genome study and validated by chromatin interaction (Tang et al., 2022b). This inversion appears to be rather common and was observed in 12 out of 20 *S. tuberosum* diploid landraces. A second region without recombination of 6.3 Mb length, observed on the linkage map of clone C of chromosome *10*, could be identical to the inversion reported on chromosome *10* from 52.7 to 59.1Mb in *S. tuberosum* Group Andigenum clone PG6244 (i.e. CIP 703509). Genetic map length, reflecting the recombination frequency, of the maternal clone C is always higher than the one of the paternal clone E. This agrees with Anithakumari et al. (2010) who reported genetic maps length of 1012.4 cM for clone C and 774.6 for clone E. We cannot determine if this difference in recombination rate is due to heterochiasmy or other factors because the reciprocal cross never succeeded.

Transmission ratio distortion

Loci subject to TRD have been reported in almost every mapping study in diploid potato (Manrique-Carpintero et al., 2016). The relevance of their identification to avoid inbreeding depression was recently highlighted (Zhang et al., 2019; Zhang et al., 2021). We reported three female, six male loci with gametic TDR and one locus showing zygotic selection. Here, the underlying causes will be discussed. TRD locus *DE1*, located in the pericentromeric region of the paternal chromosome *1* is due to the S-locus involved in gametophytic self-incompatibility (Gebhardt et al., 1991; Jacobs et al., 1995). The zygotic TRD at locus *DZ1* centred around at 69.65 Mb on chromosome *1* in our population coincides with the *ar1* locus and candidate gene StSIEL identified in clone E (i.e. PG2662) selfing population by Zhang et al. (2019). Their phenotypic description of *ar1* mutants, abnormal rooting and shoot differentiation, partially agrees with the phenotype crumpled, as previously reported in CxE population by Jongedijk et

al. (1990) and detailed in Suppl. Text 1. More generally, all of the paternal TRD observed in our population coincide with TRD presented by Zhang et al. (2019) in E selfing population. For instance, the gametic TRD at locus *DE8a* is identical to the locus *la1*. Here we show that in addition to the previously reported zygotic selection against homozygous *la1,* the *la1*/*DE8a* region is also subject to gametic selection. Furthermore, the strong gametic TRD at locus *DE11* is identical to an unnamed but equally strong zygotic TRD reported by the same authors. Those potentially conflicting observation on the gametic or zygotic stage of TRD could be reconciled assuming that clone C is homozygous for the allele causing zygotic TRD. In addition, most of the TRD loci identified in maternal clone C overlap with previously identified TRD in selfing population of *S. tuberosum* Group Phureja clones E86-69, C10-20 and PG6359 (Zhang et al., 2021). However, overlapping TRD regions identified in different clones can be due to a shared locus or to independent loci located in the same region. The TRD at locus *DC12* overlaps with a TRD region observed in clone PG6359 selfings. The TRD at locus *DC4* overlaps with TRD regions observed in clone E86-69 and C10-20 selfings that could not be linked with a visible phenotypic defect. In addition to TRD we explored non-independent assortment of alleles between pairs of loci and to our knowledge are the first to report their presence in a potato population. Most of the associated pairs of loci overlap with regions displaying TRD such as *DE8a* and *DE11* or *DC4* and *DC12*. Interestingly, this was not the case for the low recombining male chromosome *3* involved in non-independent assortments of alleles with both chromosome *9* and *11* without harbouring TRD. Non-independent assortment of alleles between pairs of unliked loci can be explained by the presence of co-adapted gene complexes (Clegg et al., 1972). Alternatively, such associations can be found in interspecific crosses where they can be caused by Bateson–Dobzhansky–Muller incompatibilities (Bateson, 1909; Dobzhansky, 1936; Muller, 1942) resulting in selection against heterospecific allele combinations in hybrids. Some of the associations observed in our population could be due to such incompatibilities since the parental clones of our population present a mixed ancestry of *S. tuberosum* Group Tuberosum and Phureja, and *S. vernei*. Overall, the identification of unliked loci having alleles that positively or negatively associate with each other adds a new layer of complexity to the understanding of inbreeding depression in potato. When designing the genome of inbred potato, as proposed by Zhang et al., (2021), one should consider that complementation and interaction are not only allelic phenomena but also inter-locus ones, acting between different genomic regions.

Seedling tuber yield in pots

The identification of the *StCDF1* locus as the major-effect QTL regulating yield in pots in our population is consistent with previous mapping studies (Manrique-Carpintero et al., 2015; Marand et al., 2019) and with the molecular function of *StCDF1*, which has

been described as the master regulator of potato maturity (Kloosterman et al., 2013). Maturity is observed by comparing the status of above-ground plant development (foliage discolouration, prostrated stems, end of apical sprout and flowering activity) to the maturity status of reference varieties. The maturity values predict the duration of the life cycle of field grown cultivars, which varies between 80 and 140 days. Maturity also predicts the dependency on daylength to trigger tuber development. The wild-type allele *StCDF1.1* is commonly present in genotypes with a prolonged growing season and dependency on short-day conditions for tuberization. *StCDF1.2* and *StCDF1.3* alleles are found in early maturing genotypes adapted to long-day conditions. Breeders are exploiting the dosage dependent effect of *StCDF1* alleles on maturity to breed cultivars adapted to different growing condition and market segments. For example, the very late maturing starch cultivars Altus and Avenger are quadruplex for *StCDF1.1* (Hoopes et al., 2022) resulting in a prolonged growing season which maximize accumulation of starch. While long growing seasons are usually associated with higher yields in cultivars, our pot-grown seedlings show that early genotypes, bearing *StCDF1.2* or *StCDF1.3* haplotypes yielded more than late genotypes homozygous for *StCDF1.1*. This counter-intuitive outcome can be explained by the delayed tuberization of genotypes homozygous for *StCDF1.1*. Our 1.1/1.1 cohort with delayed tuberization is genetically different in comparison to commercial varieties, which are selected for early tuber set, independent of maturity. Interestingly, in our population the positive effect of *StCDF1.3* on tuber yield was larger than the one of *StCDF1.2*. Similarly, a stronger effect of *StCDF1.3* over *StCDF1.2* on early maturity was also estimated in the offspring the tetraploid cultivars Altus and Columba (Hoopes et al., 2022). Those results are consistent with the finding that *StFLORE*, the long non-coding RNA and antisense transcript of *StCDF1*, is disrupted in *StCDF1.3* but not in *StCDF1.2* (Ramírez Gonzales et al., 2021). Disruption of *StFLORE* makes *StCDF1.3* heterozygotes theoretically less susceptible to translation inhibition than *StCDF1.2* heterozygotes. While the effect of maturity on yield is usually controlled with a co-factor, our large population allowed us to perform independent QTL discovery within three maturity cohorts and unravel minor effect QTLs specific to a given maturity class. With this approach we identified five minor effect QTLs, all of them specific to a single cohort, suggesting GxG interaction between *StCDF1* allele combinations and those minor-effect QTLs. Given the epistatic effect of *StCDF1* allele combinations over the minor effect QTLs identified, we argue that deciphering the genetic component(s) of yield should be performed in fixed *StCDF1* background. For example, non-*StCDF1* factors contributing to early tuberization might substantially contribute to yield in late maturing genotypes, and must have been selected for in cultivars such as Altus and Avenger (Hoopes et al., 2022), but their effects will be masked in early maturing genotypes bearing *StCDF1.2* or *StCDF1.3* alleles. Because of the pleiotropic effect of maturity on

starch content which later influence frying colour and cooking type, this argument may also hold for other traits.

Candidate genes for pollen shed

While poor male fertility is a historical issue in potato breeding (Krantz, 1924), the ongoing re-invention of potato into a diploid F_1 hybrid crop (Jansky et al., 2016; Lindhout et al., 2011) is turning this trait into a major breeding target. Fertility starts with the ability to produce flowers for which we could not find QTL in our population. Male fertility is subsequently determined by the ability to produce and release pollen for which we identified one major QTL *PPE2* on chromosome *2* with a top bin-marker located at 37.65 Mb and a LOD-1.5 confidence interval ranging from 33.6 Mb to 38.1 Mb. Studying TRD at the fruiting stage followed by a bulked segregant analysis for floral bud abortion, Zhang et al., (2021) identified *StDYT1 (Soltu.DM.02G019340) as a regulator of stamen development and thus affecting fertility. StDYT1 is located approximately at 33.6 Mb on the chromosome 2 of DMv6.1 assembly, right at the boundary of the LOD-1.5 confidence interval of QppE2* and could be considered as a candidate gene. We identify two other candidate genes involved in anther development more closely located to the top bin-marker of *QppE2*: *Soltu.DM.02G024440* and *Soltu.DM.02G023940*. *Soltu.DM.02G024440* is homologous to the *S. lycopersicum transcription factor Tomato MADS box gene 6* (*TM6*). *TM6* RNAi lines and *TM6* deletions mutant *ms-15* display homeotic defects primarily in stamen and are male sterile (Cao et al., 2019; de Martino et al., 2006). Since carpelloid stamen were only observed on a few clones out of the entire population, it is unlikely that a loss-of-function allele of *Soltu.DM.02G024440* is segregating our population. However, altered expression level of *Soltu.DM.02G024440* resulting in aberrant anthers development and subsequent poor pollen shed cannot be excluded. The second candidate gene, *Soltu.DM.02G023940*, is homologous to the *A. thaliana* homeobox transcription factor *WUSCHEL* (*WUS*). WUS is regulating stem cell activity in the organizing centre of floral meristem and is implicated in floral patterning (Ikeda et al., 2009). In addition, (Deyhle et al., 2007) reported the involvement of WUS in anther development more particularly in the differentiation of stomium cells. During normal development, stomium cells undergo cell wall thickening and degenerate which leads to rupture of the anther and pollen release. In *wus* mutants, this process is repressed resulting in indehiscent anthers and no pollen release which resemble our poor pollen shed phenotype.

Data availability

The sequencing data are available from the ENA under the BioProject ID PRJEB56778. The software OutcrossSeqDiploidR is available at https://github.com/ccrclot/OutcrossSeqDiploidR. The phenotyping data, genotyping

data and, all the R code to reproduce the results and figures of this article are available at https://doi.org/10.6084/m9.figshare.21533004.v1.

Author contribution statement

CRC: conceived the research, collected and analysed phenotypical data, constructed libraries, implemented OutcrossSeqDiploidR, performed bioinformatic and genetic analysis, wrote the manuscript. XW: collected and analysed phenotypical data, constructed libraries. JK: collected and analysed phenotypical data. ATN: contributed to Smooth Descent results. JB: constructed libraries. RGFV: supervised research, reviewed manuscript. RF: supervised research, reviewed manuscript. HJvE: obtained funding, conceived and supervised research, edited the manuscript.

Acknowledgements

We thank the members of the public-private partnership "A new method for potato breeding: the 'Fixation-Restitution' approach" and SusCrop ERANET funded project "DIFFUGAT: Diploid Inbreds For Fixation, and Unreduced Gametes for Tetraploidy" (Averis Seeds B.V., Bejo Zaden B.V., Danespo A/S, Germicopa, Den Hartigh B.V., SaKa Pflanzenzucht GmbH & Co. KG, C. Meijer B.V., and Teagasc) for providing their support. We are grateful to Christian W.B. Bachem, Sara Bergonzi and, Li Shi for their input on *StCDF1* discussion. Our colleagues of Unifarm are acknowledged for plant care in the greenhouse.

Supporting information

Supplementary files are available at: https://doi.org/10.1007/s11540-023-09627-7

Suppl. Fig. 1: Distribution of mean haploid genome coverage in population CE-XW. The red dot indicates the population mean while the vertical dash line indicates the median.

Suppl. Fig. 2: Five Mb moving average of the number of heterozygous parent-private variants across bins of 100 kb. Clone C data are displayed in orange and clone E in blue.

Suppl. Fig. 3: Relationship between mean haploid genome coverage and the nuber of genotyping errors corrected by Smooth Descent. The level of genotypic noise indicated by the total number of predicted recombination events is displayed with a pseudo-color ranging from low noise in blue to high noise in red. The minimal haploid genome coverage necessary for effective error correction by Smooth Descent is indicated by the vertical dash line.

Suppl. Fig. 4: χ^2 value for zygotic transmission ratio distortion plotted over physical distances (Mb) for each chromosome. Threshold of significance (black dashed line) were determined with χ^2 test with $\alpha = 0.0004$.

Suppl. Fig. 5: Location of significant QTL for total tuber fresh weight in the different StCDF1 cohorts: a) 1.1/1.1 cohort, b) 1.2/1.1 cohort, c) -/1.3 cohort. The X axis represents physical position (Mb), the Y axis represents -log10(p-value) and threshold of significance is indicated by the black dashed line. Clone C data are displayed in orange and clone E in blue.

Suppl. Text 1: Details on the crumpled mutation.

Chapter 4

Desynapsis in potato is caused by *StMSH4* mutant alleles and leads to either highly uniform unreduced pollen or sterility

Authors

Corentin R. Clot[1], Dennis Klein[1], Joey Koopman[1], Cees Schuit[2], Christel J.M. Engelen[1], Ronald C.B. Hutten[1], Matthijs Brouwer[1], Richard G.F. Visser[1], Martina Juranić[1], Herman J. van Eck[1]

Affiliations

[1] Plant Breeding, Wageningen University and Research, Po Box 386, 6700 AJ Wageningen, Netherlands
[2] Bejo Zaden B.V., Trambaan 1, 1749 CZ, Warmenhuizen, The Netherlands

Preprint in BioRxiv (2023)

https://doi.org/10.1101/2023.02.23.529759

Abstract

The balanced segregation of homologous chromosomes during meiosis is essential for fertility and is mediated by crossovers. A strong reduction of crossovers leads to desynapsis, a process in which pairing of homologous chromosomes is abolished before metaphase I. This results in a random segregation of univalent and the production of unbalanced and sterile gametes. However, if desynapsis is combined with another meiotic alteration that restitutes the first meiotic division, then uniform and balanced unreduced gametes, essentially composed of non-recombinant homologs, are produced. This mitosis-like division is of interest to breeders because it transmits most of the parental heterozygosity to the gametes. In potato, desynapsis is a recessive trait that was tentatively mapped to chromosome *8*. In this article, we have fine-mapped the position of the desynapsis locus and identified *StMSH4*, an essential component of the class I crossover pathway, as the most likely candidate gene. A seven base-pair insertion in the second exon of *StMSH4* was found to be associated with desynapsis in our mapping population. We also identified a second allele with a 3820 base-pair insertion and confirmed that both alleles cannot complement each other. Such non-functional alleles appeared to be common in potato cultivars. More than half of the varieties we tested are carriers of mutational load at the *StMSH4* locus. With this new information, breeders can choose to remove desynaptic alleles from their germplasm to improve fertility or to use them to produce highly uniform unreduced gametes in alternative breeding schemes.

Introduction

Meiosis is a specialized type of cellular division, essential for sexually reproducing organisms, which generates four haploid spores out of a single diploid mother cell. This is achieved via two successive divisions, meiosis I and II, following a single S-phase. While meiosis II resembles a haploid mitosis where sister chromatids are separated, meiosis I is characterized by pairing and subsequent separation of homologous chromosomes. During the prophase of meiosis I, homologous chromosomes pair with each other in a process known as synapsis. The connexion between homologs is maintained until anaphase I by crossovers which ensure their proper positioning and segregation by providing a counterforce to the pole directed spindle forces (Page and Hawley, 2003). Indeed, if chiasmata are not formed and synapsis cannot be maintained, such as in asynaptic *Arabidopsis thaliana asy1* mutants, univalents will segregate randomly, ultimately producing aneuploid nonviable gametes (Armstrong et al., 2002). Desynapsis indicates a condition with a similar gametic outcome and is characterized by the failure to maintain synapsis between most homologs after a seemingly initially normal pairing. Desynapsis can result from a drastic reduction in crossover numbers, as observed in organisms with a mutation in the class I crossover pathway (Lynn et al. 2007; Pyatnitskaya et al. 2019). This pathway, also known as ZMM pathway (an acronym for Zip1–4, Msh4–5 and Mer3), is responsible for 75-85% of crossovers in *A. thaliana* (Serrentino and Borde, 2012). In addition to their role in chromosome segregation, crossovers are necessary for recombination events which reshuffle parental chromosomes into unique genetic combinations. Although recombination and segregation are essential for genetic diversity, they can pose difficulties for breeders of heterozygous outcrossing crops, as it makes it challenging to preserve previously chosen combinations of beneficial alleles. The highly heterozygous autotetraploid potato (*Solanum tuberosum*) with its slow increase of genetic gains is one embodiment of such a crop. To circumvent this challenge, the potato breeding community is currently putting a lot of efforts in the conversion of the allogamous tetraploid germplasm into a diploid self-compatible germplasm. This new germplasm, compatible with inbreeding, should enable breeding strategies that preserve cumulative genetic gains and delivers true potato seeds (TPS) F_1 varieties (de Vries et al., 2023; Jansky et al., 2016; Lindhout et al., 2011; Zhang et al., 2021). Alternatively, if recombination is supressed, as in the case of asynaptic *StDMC1* RNAi mutants (Kumar et al., 2023), the genetic makeup of the parents can be fixed and potentially transmitted to their offspring. This makes it possible to produce TPS without the need for extensive inbreeding. This concept was first explored about three decades ago using desynaptic potato clones that produce unreduced gametes (Jongedijk, 1991). Monogenic recessive desynaptic mutants were identified in dihaploids from variety Chippewa and in *S. tuberosum* Group Tuberosum and Phureja hybrids. Those mutations were proven to be allelic and unified under the locus name *Ds-1* (Jongedijk

and Ramanna, 1988). Locus *Ds-1* was tentatively mapped to chromosome *8* in the bi-parental diploid population CE (Jacobs et al., 1995). Desynaptic mutants are characterised by a 90% reduction in crossover number affecting equally male and female meiosis resulting in the formation of ~1 bivalent per meiocytes (Jongedijk and Ramanna, 1989). Assuming that during meiosis I each univalent has an equal chance of moving to either pole, the probability to obtain a gamete with a normal chromosomal set up is $\frac{1}{2^{n-b}}$ with n being the haploid number of chromosomes and b being the number of remaining bivalents. For a desynaptic diploid potato clone, we can estimate that a single reduced gamete out of over 2000 will be balanced, which leads to sterility. However, fertility can be rescued by a second meiotic alteration producing highly uniform unreduced gametes by First Division Restitution (FDR) (Ramanna, 1983). Those unreduced gametes, also known as 2n gametes (2nG), are formed when the outcome of meiosis I is restituted by the mis-orientation of meiosis II spindle such as in *ps1* and *jason A. thaliana* mutants (De Storme and Geelen, 2011; D'Erfurth et al., 2008). Whether the first division was balanced or chaotic has no impact and meiosis II can progress with the equational division of the entire chromosomal complement. This phenomenon was recently exploited to rescue male fertility in haploid *A. thaliana* (Aboobucker et al. 2023). FDR 2n pollen production is not rare in potato and has been reported in the parents of the population used to map *Ds-1* (Mok and Peloquin, 1975b; Ramanna, 1979).

In the current study, we exploit the joint segregation of desynapsis and FDR 2n pollen to fine map the *Ds-1* locus on the short arm of chromosome *8*. We identified *StMSH4* as candidate gene and discovered a 7 bp insertion in the second exon of the allele associated with desynapsis. Mining the growing number of potato assemblies, we discovered another allele with a 3820 bp insertion at the same position and confirmed that both alleles cannot complement each other. We subsequently found that non-functional *StMSH4* insertion alleles are prevalent in European cultivars. Finally, we discuss the opportunities and limitations offered by desynapsis in the context of potato breeding.

Materials and Methods

Plant materials

The diploid mapping population CE-XW comprising 1536 full-sibs descends from a cross between two heterozygous potato clones named C (USW5337.3) and E (77.2102.37) with mixed ancestry of *Solanum tuberosum* Group Tuberosum and Phureja, and *S. vernei*. The seedlings were grown under standard greenhouse conditions at Unifarm (Wageningen University and Research) in 2020 (Clot et al., 2023b) and a subset of 500 individuals were grown from tubers the following year.

Tubers were planted the 12th of April 2021 in five-litre pots and grown outdoors in a screen cage equipped with sprinkler irrigation. The diploid populations CRH and ERH, used for a complementation test, descend from crosses between the *S. tuberosum* diploid clone RH89-039-16, used as male parent and clones C and E used as female parents. Populations CRH and ERH were sown the 29th of March 2022. For CRH and ERH populations, 100 seedlings were transplanted in 11x11 cm pots and grown in a greenhouse at ambient temperatures and under 16 hours of light.

Phenotyping pollen stainability and desynapsis

During the flowering stage of the CE-XW population, which occurred between the seventh and the tenth week post-sowing in 2020 and between the fourth and nineth week post-planting in 2021, one pollen sample per flowering individual was collected. Pollen samples were extracted from a freshly opened flower at anthesis using a vibrator pin (modified electric toothbrush). Pollen was spread on a glass slide and stained with a simplified version of Alexander staining (Peterson et al., 2010) and observed under bright field using a Axiophot Zeiss microscope equipped with a Neofluar 10x/0.30 lens. For each sample, four random field views containing approximately 50 pollen grains each were used to visually estimate the proportion of stained pollen relative to the unstained or shrivelled pollen grains. Within the fraction of stained pollen, the proportion of 2n pollen, having a ~15% larger diameter, was assessed. Combining those two observations, we identified desynaptic individuals. In a wild-type background the effect of desynapsis on pollen stainability is undistinguishable from other forms of sterility. In the presence of another meiotic alteration, resulting in FDR, the unbalanced chromosome segregation of first meiotic division is restored (Fig. 1a). This results in a variable proportion of stained balanced 2n pollen grains next to the shrivelled, unstained, and unbalanced reduced pollen grains. The dichotomous key shown in Fig. 1b allows a binary classification of pollen samples into synaptic and desynaptic classes. Samples with a combination of large stained and small unstained pollen grains were considered desynaptic. Samples with at least 5% of stained pollen grains and including at least one small and stained pollen grain were considered synaptic, irrespective of the level of 2n pollen production.

The same phenotyping protocol, with an increased stringency for synaptic classification set at 5% of stainability of reduced pollen grains, was applied to the CRH and ERH populations which flowered between the nineth and the eleventh week post-sowing. When in doubt about pollen size, pictures were taken with the Zeiss Axiomcam ICc 5 colour camera and pollen diameter was measured with Zen 2.3 lite software. Pollen grains with a diameter above 23 µm were consider as 2n pollen.

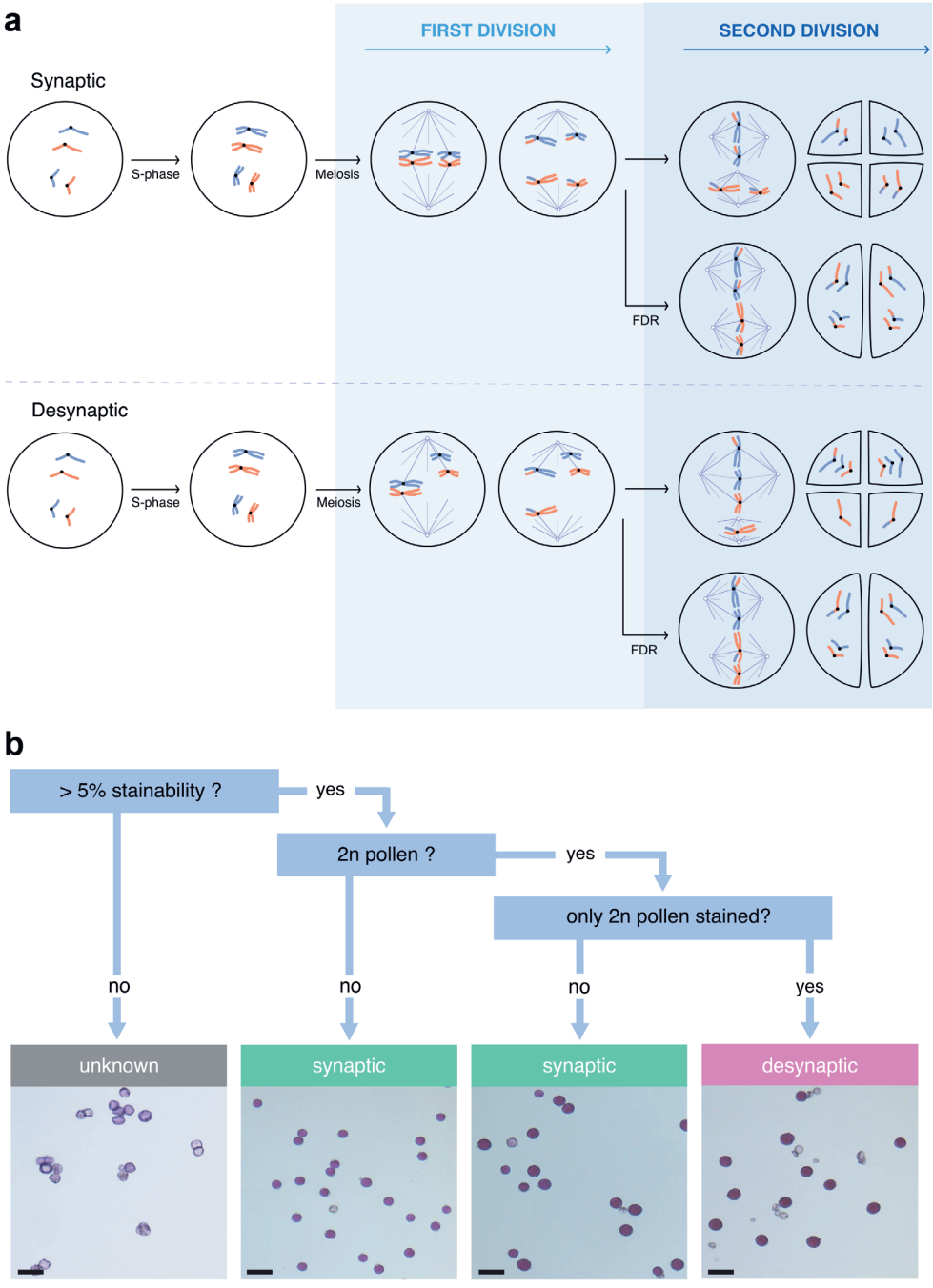

Figure 1: **a)** Schematic representation of synaptic and desynaptic meiosis producing reduce and FDR 2n pollen. **b)** Dichotomous key used to phenotype desynapsis in population CE-XW via pollen microscopy. Scale bar is 50 µm.

Genetic analysis

Marker data and map construction of population CE-XW is detailed in Clot et al. (2023). A total of 4894 female and 4740 male markers segregating across 1461 individuals were used. QTL mapping was performed using the package polyqtlR version 0.0.6 (Bourke et al., 2018). The function singleMarkerRegression was used to fit an additive model at each marker position returning the $-\log_{10}$ p-value of model fit per marker. The significance thresholds for QTL detection were determined via permutation tests on the phenotypic values with N = 1000 cycles and $\alpha = 0.05$. After an initial QTL discovery using separate maternal and paternal markers, QTL discovery was performed at full classification obtained by merging parental markers of identical physical map position. This full classification was used for the recombinant analysis presented in Suppl. Table 1.

Candidate gene exploration

Potato orthologs of *Arabidopsis thaliana* ZMM genes (including *AtMLH1* and *AtMLH3*), necessary for class I crossovers, were identified by blasting the *A. thaliana* protein sequences against DM v6.1 Protein high confidence gene model on SpudDB (http://spuddb.uga.edu/blast.shtml) with default parameters. Potato orthologs were named according to the *A. thaliana* gene name prefixed by St (for *Solanum tuberosum)* instead of At.

Retrieving haplotypes of the *StMSH4* locus

Haplotypes of *StMSH4* were mined from a collection of *de novo* assembled potato genomes. A local blast database was build using Nucleotide-Nucleotide BLAST 2.8.1+ (Camacho et al., 2009) with a *S. lycopersicum* cv. Heinz 170 build SL5.0 genome (Zhou et al., 2022) and 18 de-novo assembled potato genomes: M6 v.4.1 (Leisner et al., 2018); DM v6.1 (Pham et al. 2020); Solyntus v.1.1 (van Lieshout et al., 2020a); RH89-039-16 v.3 (Zhou et al. 2020); A6-26, E4-63, E86-69, PG5068, PG6359 and PG0019 (Tang et al., 2022a) ; Atlantic v.2.0, Castle Russet v.2.0, Avenger, Altus, Columba and, Spunta (Hoopes et al., 2022); Otava v.1 (Sun et al., 2022); C88 v.1 (Bao et al., 2022). The genomic sequence of Soltu.DM.08G000640 (*StMSH4* in DM v6.1) was used as query and the results were manually curated to stitch together partial hits within 5-kb of each other, assumed to be truncated by structural variations. Incomplete hits matching with scaffold ends were removed from the analysis. Genic and 500 bp upstream and downstream regions from those genome assemblies, homologous Soltu.DM.08G000640, were collected for haplotype analysis. The *StMSH4.f1* haplotype, associated with the desynapsis phenotype in our mapping population, was constructed in Geneious Prime 2022.2.2 (https://www.geneious.com) using the merged bam files (ENA BioProject ID PRJEB56778) of the 102 clones phenotyped as desynaptic and genotyped with allelic combination h1h3 at 1.15 Mb on chromosome *8.*

Multiple sequences comparison

Multiple sequence comparison of *StMSH4* alleles was performed in Geneious Prime 2022.2.2 using MUSCLE 3.8.425 (Edgar, 2004) with default parameters. The dendrogram was generated using the neighbour-joining algorithm with default settings.

KASP marker analysis

Leaf samples of complementation test populations CRH and ERH were sent to Bejo (Warmenhuizen, The Netherlands) for Kompetitive allele-specific PCR (KASP) marker analysis (LGC Genomics GmbH, Berlin, Germany) following to manufacturer protocols. KASP markers distinguishing between putatively functional and non-functional *StMSH4* alleles were developed. Primers are listed in Suppl. Table 2. KASP assay results were visualized using SNPviewer (lgcgroup.com/products/genotyping-software/snpviewer) to confirm correct segregation and genotype calling.

K-mer based exploration of *StMSH4* insertion alleles in commercial germplasm

A *k*-mer based method was used to investigate the presence of *StMSH4* insertion alleles in commercial tetraploid germplasm. Whole genome sequencing reads (150bp pair-ends) of 134 tetraploid potato varieties (unpublished data) were *k*-merized into 31-mers using KMC v.3.1.0 (Kokot et al. 2017) with default parameters and option -k31. Five additional *k*-mer sets (MSH4_777:T, MSH4_777:C, StMSH4.t1/t2_start, StMSH4.t1/t2_end and, StMSH4.f1 shown in Suppl. Table 3) were generated by *k*-merizing sequences of 39 nucleotides overlapping with the T/C polymorphism of KASP marker MSH4_777, sequences of 60 nucleotides centred around both insertion boundaries of the transposon insertion of *StMSH4.t1* as well as a sequence of 67 nucleotides centred around the 7 bp footprint of *StMSH4.f1*. Those five *k*-mer sets were intersected with the *k*-mer sets of 134 varieties. *K*-mers lacking specificity, present in virtually all varieties, were remove from the analysis which resulted in final sets of 9 *k*-mers specific to *MSH4_777:C*, 9 *k*-mers specific to *MSH4_777:T*, 31 *k*-mers specific to the footprint, 22 *k*-mers specific to the transposon start and, 23 *k*-mers specific to the transposon end. The absence/presence of *k*-mers in varieties, along with specific k-mers frequencies, were used to infer the distribution of specific *StMSH4* haplotypes in commercial germplasm. The dosage of *MSH4_777:C* was estimated by dividing the *k*-mer frequencies of MSH4_777:C with the sum of MSH4_777:C and MSH4_777:T *k*-mer frequencies.

Results

Desynapsis is controlled by a single locus on chromosome *8*

Stained pollen samples from 1345 seedlings of the CE-XW population were observed under a microscope and classified according to the dichotomous key presented in Fig. 1b. These microscope observations allowed classification of 134 individuals as desynaptic and 912 as synaptic while 299 offspring remained unclassified due to either low pollen stainability, low proportion of 2n pollen or uncertainty about pollen ploidy. To validate this classification, a subset of 500 individuals were regrown from tubers the following year, from which 470 pollen samples could be collected. In 107 desynaptic and 333 synaptic clones we observed the same phenotype in both years. Five desynaptic clones were classified as normal in the 2nd year and 25 clones with normal synapsis were classified as desynaptic in the 2nd year. These 30 conflicting classifications (6.3%) were specifically related to individuals with low pollen shed and were discarded from further analysis. Ultimately, we mapped the *Ds-1* locus using 106 desynaptic and 857 synaptic individuals. While the classification of the desynaptic individuals was based on two years data, this was not the case of all synaptic individuals. Nonetheless, we decided to use all genotyped individuals classified as synaptic based on a single year observation, considering that the increase in power offered by a larger cohort will compensate for putative misclassification.

Using this binary phenotypic classification in a single marker regression with female and male markers, we mapped the *Ds-1* locus to the north arm of chromosome *8* at 1.15 Mb both on the female and the male physical map. Merging parental markers of identical physical map positions, resulting in full classification, also localised *Ds-1* at 1.15 Mb. (Fig. 2a-b). A recessive inheritance of desynapsis implies that the recessive phenotype will match only with one of the allele combinations h1h3, h1h4, h2h3 and h2h4. Indeed, three marker classes accurately predict synaptic plants where h1h4, h2h3 and h2h4 are associated with the *ds-1/Ds-1*, *Ds-1/ds-1* and *Ds-1/Ds-1* synaptic genotypes respectively, with only one or two phenotypic misclassifications, as shown in the mosaic plot (Fig. 2c). The allele combination h1h3 however, indicative of the *ds-1/ds-1* desynaptic genotypes showed 102 true positive and 79 false negative classifications. These false negatives suggest a second 1:1 segregating locus compensating the *ds-1/ds-1* mutants. However, a new QTL analysis for desynapsis within the cohort of individuals with the allelic combination h1h3 at 1.15 Mb on chromosome *8* did not identify such a compensatory locus (Suppl. Fig. 1). Upon microscopic re-examination of the slides of 20 random false positive clones, five are better classified as ambiguous due to a low pollen shed, one was confirmed synaptic and 14 were reclassified as desynaptic. Those 14 clones show a stainability above 50% with more than 95% of stained pollen grain being 2n. The occasional observation of a few small and stained pollen grains in those clones were in retrospect also visible in 10

randomly re-examined true positive desynaptic clones. This suggests that desynaptic plants can produce up to 3% of stained small pollen grains, assumed to be haploid or aneuploid cells with a stainable cytoplasm. Overall, false positives were essentially explained by erroneous phenotypic classifications due to a too mild threshold to classify a pollen sample as synaptic based on a single stained small pollen grain, or the potential misidentification of 2n pollen as n pollen in desynaptic clones with elevated levels of 2n pollen production. Finally, after a recombinant analysis in the subset of individuals curated for false positives and with unambiguous recombination breakpoints, we identified the first 1.9 Mb of chromosome *8* as candidate region for desynapsis (Suppl. Table 1).

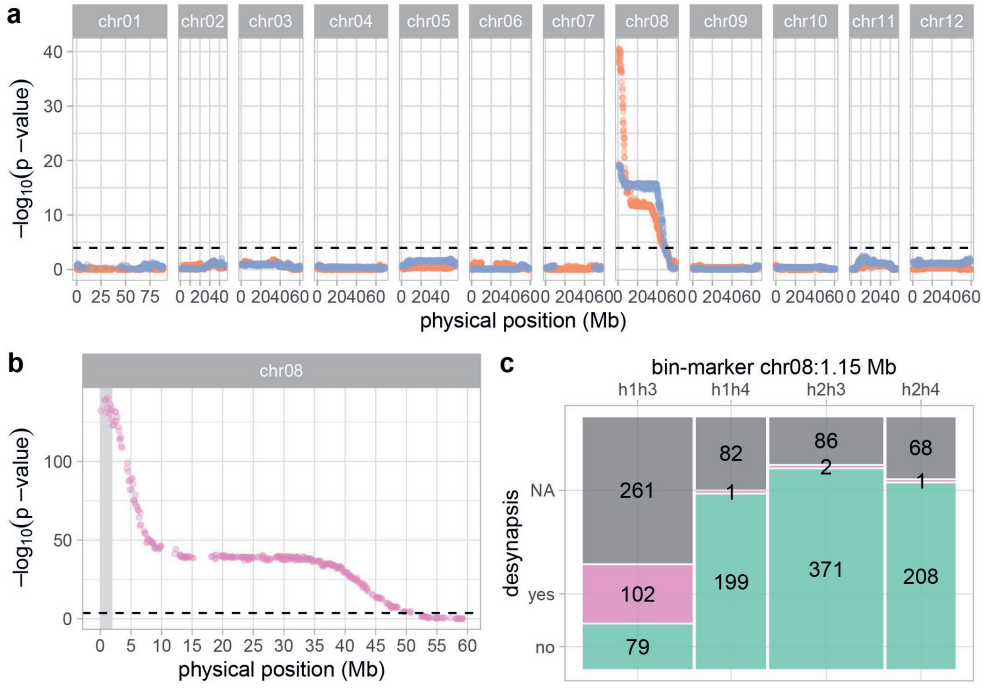

Figure 2: **a-b)** Significance of the association between markers and the phenotypic classification of desynapsis (n=963). The X axis represents the physical position (Mb), the Y axis represents -$\log_{10}$(*p*-value) and, the threshold of significance is indicated by the black dashed line. In panel **a)**, clone C marker data are displayed in orange and clone E in blue while in panel **b)** the integrated marker allele combination h1h3 is displayed in pink and the candidate region for *Ds-1* is highlighted in grey **c)** Mosaic plots illustrating the goodness of fit between phenotypic and genotypic observations for the markers at 1.15Mb on chromosome *8*. Individuals phenotyped as synaptic or desynaptic are displayed in green and pink respectively, and phenotypically unclassified individuals in grey.

Candidate genes in the ZMM pathway

Low amounts of stained n pollen in desynaptic CE material were explained by Jongedijk and Ramanna (1989) with cytological observations. They showed a 90% reduction of

crossovers in meiocytes of desynaptic CE plants resulting in unbalanced chromosomal segregation. Class I crossovers, resulting from the ZMM pathway, were shown to represent 75-85% of the crossovers in *A. thaliana* (Serrentino and Borde, 2012). Therefore, a gene involved in the ZMM pathway may explain desynapsis in our material. We collected the physical positions of the potato orthologs of *A. thaliana* ZMM genes, including AtMLH1 and AtMLH3, and report their names and positions as found on the potato reference genome DM v6.1 in Table 1. All *A. thaliana* genes were associated with a single ortholog apart from the duplicated genes *AtZYP1a* and *AtZYP1b* which matched a single gene in DM v6.1 (*Soltu.DM.04G029680*) and *AtMSH5* and *AtMLH3* whose orthologs in DM were annotated as four consecutive genes (*Soltu.DM.10G007760/70/80/90* and *Soltu.DM.04G016510/20/30/40*). Strikingly, *Soltu.DM.08G000640* the ortholog of *AtMSH4*, from now on referred to as *StMSH4*, is located around 1.14 Mb on chromosome *8*, in the middle of the candidate region for the *Ds-1* locus. In *A. thaliana*, *AtMSH4* mutants exhibit a severe reduction in fertility due to an 85% reduction in chiasmata frequency at metaphase I leading to univalents and unbalanced chromosomal segregation (Higgins et al., 2004). This phenotypic description is identical to the phenotype observed in our material.

Table 1: Identification of S. tuberosum orthologs of the A. thaliana ZMM genes including AtMLH1/3 and their positions on the potato reference genome (DM v6.1).

A. thaliana		***S. tuberosum* (DM v6.1)**			
gene name	**gene ID**	**ortholog name**	**ortholog ID**	**position**	**references**
AtZYP1a/b	*AT1G22260*	*StZYP1*	*Soltu.DM.04G029680*	chr04:61,021,784 – 61,029,101	(Higgins et al., 2005)
AtHEI10	*AT1G53490*	*StHEI10*	*Soltu.DM.08G006360*	chr08:9,024,306 – 9,028,660	(Chelysheva et al., 2012)
AtMSH4	*AT4G17380*	*StMSH4*	*Soltu.DM.08G000640*	chr08:1,132,003 – 1,145,869	(Higgins et al., 2004)
AtMSH5	*AT3G20475*	*StMSH5*	*Soltu.DM.10G007760 /70/80/90*	chr10:10,573,090 – 10,606,226	(Higgins et al., 2008)
AtZIP4	*AT5G48390*	*StZIP4*	*Soltu.DM.09G015170*	chr09:39,928,465 – 39,932,216	(Chelysheva et al., 2007)
AtSHOC1	*AT5G52290*	*StSHOC1*	*Soltu.DM.03G017560*	chr03:41,931,056 – 41,945,565	(Macaisne et al., 2008)
AtPTD	*AT1G12790*	*StPTD*	*Soltu.DM.06G021190*	chr06:47,889,378 – 47,892,355	(Wijeratne et al., 2006)
AtMER3	*AT3G27730*	*StMER3*	*Soltu.DM.02G033190*	chr02:44,748,723 – 44,756,970	(Chen et al. 2005; Mercier et al. 2005)
AtMLH1	*AT4G09140*	*StMLH1*	*Soltu.DM.04G016510 /20/30/40*	chr04:33,317,255 – 33,354,788	(Dion et al., 2007)
AtMLH3	*AT4G35520*	*StMLH3*	*Soltu.DM.02G022580*	chr02:36,320,407 – 36,351,765	(Jackson et al., 2006)

Natural diversity of *StMSH4* haplotypes in potato germplasm

Sequence data of *StMSH4,* being a plausible candidate gene for the *Ds-1* locus, were retrieved with BLAST from 18 de novo assembled genomes. This resulted in the identification of at least 19 unique haplotypes. We aligned these sequences with the haplotype of CE-XW desynaptic individuals and calculated a neighbor-joining tree of the haplotypes and rooted the tree using the haplotype from *S. lycopersicum* as outgroup (Fig. 3a). The most common haplotype, *StMSH4.1*, was identified in six different clones: DM, RH, Atlantic, PG6359, Otava and, C88. The mutant haplotype associated with desynapsis in our mapping population was identical to the second most common haplotype, named *StMSH4.f1* where the f indicates a footprint. This footprint haplotype was found in Colomba, Otava and C88 and is characterized by a seven bp insertion within the second exon of *StMSH4* as annotated in DM v6.1. Strikingly, much longer insertions of 3820 and 3819 bp are observed at the exact same position in haplotypes named StMSH4.t1 and StMSH4.t2, where the t indicates a transposon insertion (Fig. 3b). These transposon insertion mutants were found in clones RH89-039-16 and Spunta, respectively, and are 99.9% identical with only 10 SNPs and one indel observed within the insertion. Albeit the different insertions observed in haplotypes *StMSH4.t1*, *StMSH4.t2* and *StMSH4.f1*, the remainder of the haplotypes are identical to each other (Suppl. File 1). We submitted the *StMSH4.t1* insert sequence to BLASTn against the RepetDB database (Amselem et al., 2019) “Solanum_tuberosum consensus from ReptDB v2” with defaults parameters. The best hit, with 93% sequence identity, returned the consensus sequence Stub_TedenovoGr-B-G5183-Map16. This sequence is classified as a Class II TIR transposon and is present in 294 copies and 342 fragments in DM v4.03 and amounts for a cumulative genome coverage of 335,463 bp. In conclusion, we identified *StMSH4* as a candidate gene, and we assume that haplotypes with a premature stop codon due to a transposon insertion or footprint are associated with desynapsis.

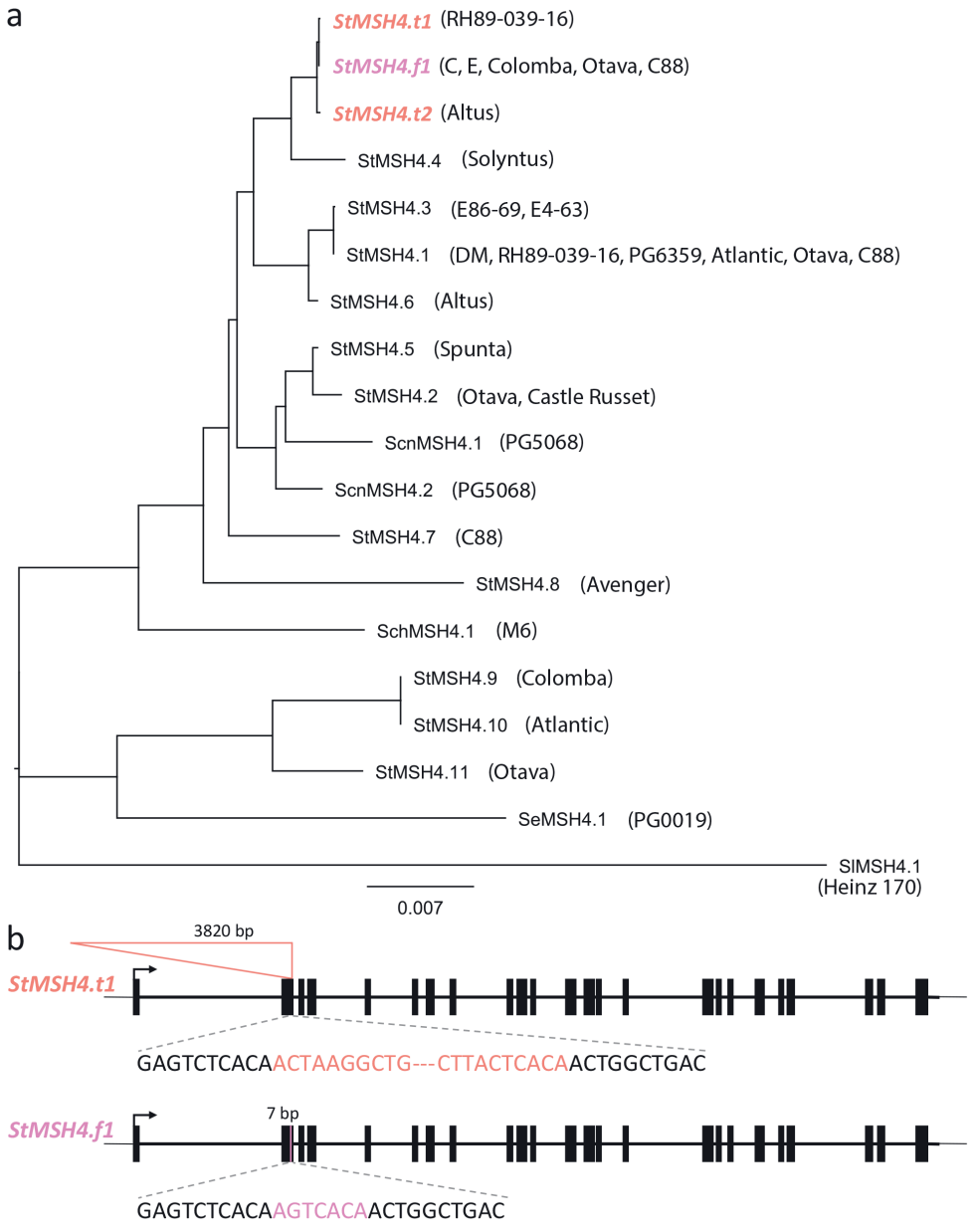

Figure 3: a) Phylogenetic dendrogram of StMSH4 haplotypes retrieved from de novo assemblies. Clone from which the various haplotypes were identified are indicated in between brackets. Haplotypes with insertions in the second exon of StMSH4 are indicates in pink and red. **b)** Haplotypes StMSH4.t1 and StMSH4.f1 show insertions of 3820 bp and 7bp in the second exon of candidate gene StMSH4.

Complementation test

The evidence that the *Ds-1* locus is equal to *StMSH4* is only based on a positional co-localisation and phenotypical similarity with *Atmsh4* mutants. To validate our positional evidence with molecular evidence we designed a genetic complementation assay using different insertion mutant versions of *StMSH4* as found in different potato clones. We crossed clones C and E, both heterozygous for *StMSH4.f1*, with clone RH89-039-16, heterozygous for *StMSH4.t1*, generating populations CRH and ERH. KASP marker MSH4_777 and MSH4_168, both tagging *StMSH4.t1* and *StMSH4.f1*, were used to follow the segregation of either haplotype in the progenies of CRH and ERH. As expected, in both populations the KASP markers segregated in a 1:2:1 fashion and about one quarter of offspring, homozygous for the KASP alleles associated with the insertion alleles are assumed to carry the haplotype combination *StMSH4.t1*/ *StMSH4.f1*. We phenotyped population CRH and ERH for desynapsis without prior knowledge on the seedling genotypes. Out of CRH 99 seedlings, 47 plants did not flower or shed sufficient pollen for microscopic observation and 3 plants displayed insufficient pollen stainability. Among the remaining 49 plants, 8 were classified as desynaptic and 41 as synaptic. Similarly, for the 100 seedlings of ERH, 49 plants did not flower or shed sufficient pollen for microscopic observation and 3 plants displayed insufficient pollen stainability. Among the remaining 48 plants, 7 plants were classified as desynaptic and 41 plants as synaptic. We observed a perfect correlation between the desynapsis classification and the two KASP markers distinguishing *StMSH4.t1* and *StMSH4.f1* from the functional haplotypes (Fig. 4). All phenotyped plants bearing the haplotype combination *StMSH4.t1*/ *StMSH4.f1* were classified as desynaptic indicating that *StMSH4.t1* and *StMSH4.f1* alleles cannot complement each other and are non-functional. From now on, those non-functional alleles can be referred to under the unified notation *Stmsh4.*

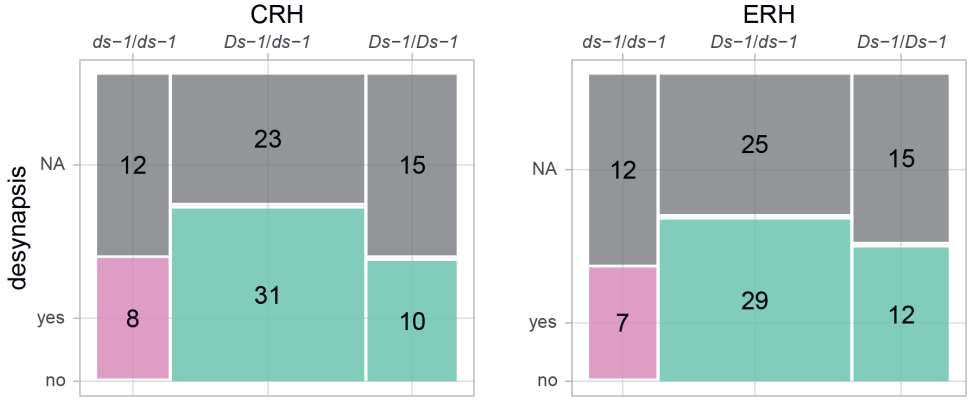

Figure 4: Mosaic plots illustrating the agreement between desynapsis phenotype and genotypes obtained with KASP markers tagging StMSH4 haplotypes in two complementation studies. CRH and ERH indicate the

offspring from a cross between clone C or E with clone RH89-039-16. Ds-1 indicates StMSH4 alleles with no insertion and ds-1 stand for either StMSH4.f1 or StMSH4.t1. Individuals phenotyped as synaptic or desynaptic are displayed in green and pink respectively, and phenotypically unclassified individuals in grey.

Occurrence of *StMSH4* insertion mutants in commercial tetraploid varieties

To assess the occurrence of *Stmsh4* alleles in commercial germplasm, we intersected the *k*-mer sets of 134 re-sequenced tetraploid varieties with the *k*-mer sets uniquely tagging the KASP marker allele *MSH4_777:C*, the footprint region of *StMSH4.f1* and the junctions at both ends of *StMSH4.t1* transposon insertion site (Suppl. Table 3). Surprisingly, we observed at least one of the mutant haplotypes *StMSH4.t1* or *StMSH4.t2* in 70 out of the 134 varieties, because in these varieties all *k*-mers unique to the junctions of both insertion ends were present. Furthermore, the footprint-specific k-mers were identified in 21 varieties including 14 varieties already positive for the transposon insertion. Two more varieties, Royal and Summer Delight, were found positive for only two *k*-mers specific for the *StMSH4.f1* footprint haplotype. After retrieving reads from these clones containing those two *k*-mers and aligning them to DM v6.1, we identified in Royal had a 4 bp footprint instead of the 7 bp insertion found in *StMSH4.f1*, indicating a new haplotype called *StMSH4.f2* (Suppl. Fig. 2a). More strikingly Summer Delight retrieved reads mapped partially to the footprint location and partially to another region 28,460 bp downstream. This suggests that the transposon excision event caused a structural rearrangement producing another more complex footprint haplotype (*StMSH4.f3*) (Suppl. Fig. 2b). Whether the *StMSH4.f2* and *StMSH4.f3* alleles are also non-function remains to be demonstrated. Finally, we observed a perfect correlation between the 79 varieties positive for *k*-mers tagging *MSH4_777:C* and varieties positive for either *StMSH4* insertion mutants. This observation confirms that our KASP marker MSH4_777 accurately predicts the insertion alleles and can be used to estimate their combined dosage. We estimated that a total of 55 varieties were nulliplex for *MSH4_777:C*, while 45 were simplex, 26 were duplex and 8 were triplex. Hence, with a minor allele frequency (MAF) estimated at 22.6%, deleterious *StMSH4* insertion mutants are abundantly present in commercial germplasm.

Discussion

Phenotypic classification of desynapsis via pollen microscopy

Rather than relying on laborious cytological observations of male meiocytes to phenotype desynapsis, we exploited the combination of desynapsis and FDR 2n pollen in our population to phenotype this trait through pollen microscopy. Either method has its advantages, where meiocyte observations would allow to capture the reduced chromosome pairing, while pollen observations allow a quick and dirty classification of

a large population. A disadvantage of this latter approach is that desynapsis could only be positively classified, and distinguished from other causes of male sterility, when combined with FDR 2n pollen production. This requirement for 2n pollen resulted in 261 unclassified ds-*1*/*ds-1* individuals (59.0%), compared to 236 unclassified *Ds-1*/— individuals (23.2%). Classification of synaptic offspring was highly accurate with only four misclassifications (0.5%), allowing a high mapping accuracy despite 79 misclassifications (43.6%) of desynaptic offspring. We learned that those misclassifications were predominantly caused by two misleading phenotypic observations: 1) almost all stainable pollen grains were unreduced and we lacked small sized pollen grains to see that, and 2) we observed a few stainable small pollen grains while this possibility was not expected in desynaptic clones (initial threshold 0%). Stainability of small pollen grains in a desynaptic individual suggests that balanced segregation may occur by chance alone or that despite an unbalanced genome the cytoplasm is still stainable. This led to a new threshold (<5%), which was successfully applied during the phenotypic classification of the complementation test populations.

From candidate region to candidate gene

In our mapping experiment, desynapsis was clearly associated with the allelic combination h1h3 on the north arm of chromosome *8*. However, the combined effect of phenotypic misclassification and noisy skim-sequencing data limited our fine mapping resolution. Nonetheless, we performed a recombinant analysis on a curated set of individuals with consistent phenotypic and genotypic data and identified the first 1.9 Mb of chromosome *8* as candidate region for *Ds-1*. Based on DM v6.1 annotation, this region contains 129 genes among which 7, including *StMSH4*, are putatively involved in reproduction. Articulating the genetic and cytological descriptions of desynapsis in potato (Jongedijk and Ramanna 1989; Jongedijk et al. 1991) with the more recently uncovered function of the ZMM proteins in *A. thaliana* (Pyatnitskaya et al. 2019), we could focus on *StMSH4* as candidate gene for desynapsis. This exemplifies how knowledge obtained from model organism can facilitate discoveries in crops, at least for conserved process such as meiosis.

Different *StMSH4* insertions alleles

In this study we identified a variety of *StMSH4* alleles, including non-functional alleles due to a transposon insertion in the second exon of this gene or to a 7bp footprint at a position coinciding with the transposon insertion site. Among them, the transposons insertion alleles *StMSH4.t1* and *StMSH4.t2*, indistinguishable based on our *k*-mer based analysis, were the most widespread (Suppl. Table 3). The low sequence divergence between *StMSH4.t1* and *StMSH4.t2*, limited to 10 SNPs and a 1 bp indel within the insertion sequence, can be explained in three ways. Either those two alleles

are *bona fide* and result from 1) independent transposition events at the same location in the same ancestral allele (very unlikely), or 2) from a single transposon insertion event and subsequent divergency restrained to the transposon sequence, or 3) these differences are artifacts caused by the challenge of assembling repetitive sequences present in high copy number in the genome. *StMSH4.f1*, the 7 bp footprint allele causing desynapsis in our mapping population, was also relatively common among tetraploid varieties. However, we could also identify two rarer footprint alleles (Suppl. Fig. 2) and cannot exclude the future discovery of more footprints resulting from independent excision events of the widespread transposon alleles.

MSH4 and meiotic adaptation to polyploidy

In allotetraploid *Brassica napus*, a reduction in the number of functional copies of MSH4 to a single copy prevents crossovers between homeologous chromosomes without affecting the total number of crossovers (Gonzalo et al., 2019). In this context, reducing *MSH4* dosage seems to help stabilising allopolyploid meiosis by favouring homologous chromosomes as recombination partner. Contrary to allopolyploids, the stabilization of meiosis in autopolyploids, such as *S. tuberosum*, does not rely on recombination partner choice but on avoiding multivalent formation, in particular the combination of trivalent and univalent (Bomblies, 2022). This can be mediated by increasing crossover interference strength and ultimately decreasing crossovers number to a minimum of one per pair (Morgan et al. 2021). Interestingly, a direct partner of MSH4, the ZMM protein and synaptonemal complex central element ZIP1 harbour signature of selection for meiotic stabilization in autotetraploid *Arabidopsis arenosa* (Yant et al., 2013). Recent cytological investigation of tetraploid potato cultivars revealed that clone Sante display significantly less multivalent than clones Maris Peers and Cara (Choudhary et al. 2020). Remarkably, Maris Peer and Cara are quadruplex for functional *StMSH4* alleles (Suppl. Table 3) and Sante appears to bear one copy of *StMSH4.t1* (S. Oome (HZPC), personal communication). Whether reducing the number of functional *MSH4* alleles is an adaptation to auto/allo-polyploidy in potato (i.e to minimize multivalents in autotetraploid *S. tuberosum* or homeologous crossovers in allotetraploid wild relatives such as *S. stoloniferum*) remains speculative but could be simply investigated with functional experiments.

Desynapsis: a double edge-sword for breeders

At a first glance, the production of sterile n gametes due to desynapsis appears to be a detrimental phenotype that should be removed from potato breeding programs. Suboptimal fertility has always been an issue for potato breeders (Krantz, 1924) and maintaining fertility upon inbreeding is central to the reinvention of potato as a diploid inbred line-based crop. With a MAF estimated at 22.6% in tetraploid varieties, *Stmsh4* alleles are contributing to the notorious fertility problems of dihaploids induced from cultivars. Ignoring double reduction, the probability to induce a dihaploid with a given

number of *Stmsh4* copies can be calculated using equation (1) with p being the ploidy of the parent P, $\mathrm{P}(P_n)$ being the probability of parent P to bear n of copies of *Stmsh4* and, $\mathrm{P}(G_k)$ being the probability of a gamete G (future dihaploid) to bear k copies of *Stmsh4*.

(1) $$\mathrm{P}(G_k) = \sum_{n=0}^{p} \mathrm{P}(P_n) \times \frac{\mathrm{C}_k^n \times \mathrm{C}_{(p/2)-k}^{p-n}}{\mathrm{C}_{p/2}^{p}}$$

Extrapolating from the estimated dosages of *Stmsh4* in commercial varieties (Suppl. Table 3), we can estimate that 6.4% of all induced dihaploids will be homozygous for *Stmsh4* and 33.5% of them will be heterozygous. While these dihaploids are often used to introduce quality and resistance traits to the diploid gene pool, they will also introduce *Stmsh4* alleles, thus hampering future inbreeding efforts. Indeed, one quarter of the S1 population obtained by selfing a clone heterozygous for *Stmsh4* will be desynaptic. Desynapsis will only be observed and selected against when attempting to produce an S2, hereby wasting labour and greenhouse space. We therefore recommend to breeders that aim at developing fertile diploid inbred lines to remove *StMSH4.f1* and *StMSH4.t1*/*t2* haplotypes from their germplasm.

Looking at desynapsis from another angle, one could envisage to combine it with another meiotic mutation, FDR unreduced gametes (FDR 2nG), and exploit these highly heterozygous and highly uniform unreduced gametes to create tetraploid varieties via sexual polyploidisation. Interploidy breeding schemes exploiting the simplicity of diploid breeding and the heterozygosity offered by tetraploids have long been proposed in potato (Chase, 1963; Hutten, 1994). In those schemes, tetraploid varieties are produced by unilateral (4x × 2x) or bilateral (2x × 2x) sexual polyploidisation with diploid clones producing 2nG. However, not all 2nG are made equal with FDR 2nG retaining ~80% of parental heterozygosity compared with ~40% for 2nG formed by a second division restitution (SDR) (Douches and Quiros, 1988a, 1988b; Jongedijk et al., 1991a; Peloquin et al., 2008). This higher heterozygosity of FDR 2nG has been linked with a significant yield increase in progenies when compared with SDR 2nG (Hutten et al., 1994a; Kidane-Mariam and Peloquin, 1975; Mendiburu and Peloquin, 1977). In desynaptic clones, the significant reduction in crossover number boosts the heterozygosity of FDR 2nG to 94.1% (Jongedijk et al., 1991a) and thus, could contribute to an even higher heterosis in the tetraploid progeny. Moreover, the increased uniformity of these 2nG will also be instrumental to the development of uniform tetraploid varieties grown from true seeds (TPS varieties), without the necessity to develop inbred parents. Practically, the breeders can generate diploid progenitor material which are heterozygous for *Stmsh4*. These progenitor clones can be used as recurrent parent at the diploid level, but after one generation of selfing the

desynaptic descendants producing FDR 2nG can be used for commercial true seed production. Likewise, desynaptic descendants can also be obtained in hybrid progeny descending from a cross between diploid parents that both carry an *Stmsh4* allele.

Despite being less common than FDR 2n pollen production, the formation of FDR 2n megaspores has been reported in desynaptic potato clones by Jongedijk et al (1991). Those authors also discussed the potential use of such mutants to produce heterozygous and uniform tetraploid TPS varieties via bilateral sexual polyploidisation. While being facilitated by the potential of marker assisted selection for desynapsis, further research on the genetic regulation of FDR 2nG production both on the male and the female side remains essential to efficiently exploit desynapsis. Importantly, the genetic regulation of FDR 2nG production is expected to be more complex on the female than on the male side because of the successive type of cytokinesis of female meiosis. FDR 2n pollen could be achieved with a single mutation as in the cases of *Atps1* and *Atjason* mutants (de Storme and Geelen, 2011; d'Erfurth et al., 2008). On the other hand, a combination of mutations leading to the loss of cohesion between sister-chromatids during meiosis I and the omission of the second meiotic division should be associated with desynapsis to obtain both male and female near non-recombinant unreduced gametes.

Author contribution statement

CRC: conceived research, collected and analysed phenotypical data, preformed QTL analysis, complementation test, allele mining, wrote manuscript. DK and JK: collected and analysed phenotypical data, performed candidate gene exploration. CS: developed and run KASP assays. CJME and RCBH advised on phenotypical observation, edited breeding discussion. MB provided support for allele mining. RGFV: supervised research, edited manuscript. MJ: supervised research, edited manuscript. HJvE: obtained funding, conceived and supervised research, edited manuscript.

Acknowledgements

We thank the members of the public-private partnership "A new method for potato breeding: the 'Fixation-Restitution' approach" and SusCrop ERANET funded project "DIFFUGAT: Diploid Inbreds For Fixation, and Unreduced Gametes for Tetraploidy" (Averis Seeds B.V., Bejo Zaden B.V., Danespo A/S, Germicopa, Den Hartigh B.V., SaKa Pflanzenzucht GmbH & Co. KG, C. Meijer B.V., and Teagasc) for providing their support.

Supporting Information

Supplementary files are available at: https://github.com/ccrclot/PhD_thesis_Supp_info

Suppl. Fig. 1: Significance of the association between parental markers and the phenotypic classification of desynapsis within the cohort of individuals with the allelic combination h1h3 at 1.15 Mb on chromosome *8* (n=181). The X axis represents the physical position (Mb), the Y axis represents -log10(*p*-value) and, the threshold of significance is indicated by the black dashed line.

Suppl. Fig. 2: Reads containing footprint specific *k*-mers extracted from a) Roya and b) Summer Delight aligned to DM v6.1.

Suppl. Table 1: Recombinant analysis in the first 5 Mb of chromosome *8* for a curated set of genotypes with markers as column and genotypes as row. The second column represent the desynapsis classification and the first row indicates the -log10(*p*-value) of marker association with the desynapsis classification in the entire cohort. Haplotype combination h1h3 is represented by a 1 while any other haplotype combination is represented by a 0.

Suppl. Table 2: KASP primers used to follow the segregation of *StMSH4.t1* and *StMSH4.f1* in population CRH and ERH.

Suppl. Table 3: Correlation of *k*-mers specific to the C allele of KASP MSH4_777 and *k*-mers specific to either the transposon insertion in *StMSH4.t1* and *StMSH4.t2* or the 7 bp insertion of *StMSH4.f1*. *K*-mers are shown in rows and varieties in columns. Numbers indicate *k*-mers frequencies in a given variety.

Suppl. File 1: Multiple sequence alignment of *MSH4* genomic sequences and *Soltu.DM.08G000640.1* cDNA.

Chapter 5

Identification of two mutant *JASON-RELATED* genes associated with unreduced pollen production in potato

Authors

Corentin R. Clot[1], Lea Vexler[1,3], Maria de La O Leyva-Perez[3], Peter Bourke[1], Christel J.M. Engelen[1], Ronald C.B. Hutten[1], José van de Belt[2], Erik Wijnker[2], Dan Milbourne[3], Richard G.F. Visser[1], Martina Juranić[1], Herman J. van Eck[1]

Affiliations

[1] Plant Breeding, Wageningen University and Research, Po Box 386, 6700 AJ Wageningen, Netherlands

[2] Laboratory of Genetics, Wageningen University and Research, Po Box 16, 6700 AA Wageningen, Netherlands

[3] Teagasc, Crops Research, Oak Park, Carlow, R93 XE12, Ireland.

Abstract

In diploid potato the production of unreduced gametes with a diploid (2n) rather than a haploid (n) number of chromosomes has been widely reported. Besides of their evolutionary important role in sexual polyploidisation, unreduced gametes also have practical value for potato breeding as bridge between diploid and tetraploid germplasm. Although early articles argued for a monogenic recessive inheritance, the genetic basis of unreduced pollen production in potato has remained elusive. Here, three diploid full-sib populations were genotyped with an amplicon sequencing approach and phenotyped for unreduced pollen production across two growing seasons. We identified two minor-effect and three major-effect QTLs regulating this trait. The two QTLs with the largest effect displayed a recessive inheritance and an additive interaction. Both QTLs colocalised with genes encoding for putative *AtJAS* homologs, a key regulator of meiosis II spindle orientation in *Arabidopsis thaliana*. The function of these candidate genes is consistent with the cytological phenotype of mis-oriented metaphase II plates observed in the parental clones. The alleles associated with elevated level of unreduced pollen showed deleterious mutation events: an exonic transposon insert causing a premature stop, and an amino acid change within a highly conserved domain. Taken together, our findings shed light on the natural variation underlying unreduced pollen production in potato and will facilitate interploidy breeding by enabling marker assisted selection for this trait.

Introduction

Unreduced gametes, also known as 2n gametes, are gametes with a diploid (2n) rather than a haploid (n) number of chromosomes. Sexual polyploidisation through unreduced gametes is proposed as a major route to polyploid formation (Ramsey and Schemske, 1998; Tayalé and Parisod, 2013). Polyploids are common among flowering plants and overrepresented in crops (Otto and Whitton 2000; Salman-Minkov et al. 2016). It is suggested that polyploids were selected pre- or post-domestication because of their greater phenotypical plasticity, larger organ size and reproductive isolation from wild-relatives (Bretagnolle and Thompson, 1995; Dempewolf et al., 2012; Sattler et al., 2016). Unreduced gametes are also an important tool for plant breeders who used them to link germplasm of different ploidy levels and for the induction of new polyploids (Ramanna and Jacobsen 2003; Younis et al. 2014). Unreduced gametes generally result from an altered meiotic process affecting either chromosomal segregation or cell division. The most reported mechanism of male meiotic restitution is the misorientation of metaphase II (MII) plate and spindles also known as parallel spindles (Bretagnolle and Thompson, 1995). In some species, simultaneous cytokinesis, the misorientation of MII plate can lead to the production of unreduced gametes that contain homologous chromosomes and are genetically equivalent to a First Division Restitution (FDR). Alternatively, unreduced gametes obtained from an omission of the second division are composed of sister chromatids and are genetically equivalent to a Second Division Restitution (SDR). Those restitution mechanisms can be caused by a combination of environmental and genetic factors (de Storme and Geelen, 2014, 2013).

Here we are interested in the genetic basis of unreduced pollen production in potato (*Solanum tuberosum*). The production of unreduced pollen via misorientation of MII plates and spindles has been commonly recorded among tuber-bearing *Solanum* species (Watanabe and Peloquin, 1989). The inheritance of this meiotic restitution mechanism and two independent types of premature cytokinesis were described as independent monogenic recessively inherited traits (Mok and Peloquin, 1975b, 1975a). The subsequent re-investigation of those cytological mechanisms by Ramanna (1979) drew a more complex picture of the situation where the co-occurrence of various restitution mechanisms in a single clone could not be excluded. Likewise, analysis of the genetic control of unreduced pollen formation during cycles of recurrent selection and in diallel crosses suggested a more complex polygenic inheritance (Dongyu et al., 1995; Ortiz and Peloquin, 1992). While unreduced gametes have been successfully used to introgress resistance genes from diploid wild relatives of potato to tetraploid cultivars (Ortiz et al. 1994; Capo et al. 2002), genetic loci involved in their production have never been identified. Meanwhile, the study of *Arabidopsis thaliana* mutants uncovered several meiotic pathways regulating unreduced gamete formation (de

Storme and Geelen, 2013). Key players in those pathways include *AtPS1* (*A. thaliana Parallel Spindle 1*) and *AtJAS* (*A. thaliana JASON*) which are known to regulate the orientation of the MII plate and spindle. Knock-out of either of those genes results in the formation of FDR unreduced pollen (de Storme and Geelen, 2011; d'Erfurth et al., 2008). On the other hand, loss of function of *TAM* (*TARDY ASYNCHRONOUS MEIOSIS*) or *OSD1* (*OMISSION OF SECOND DIVISION*) leads to the failure of entering meiosis II and to the production of SDR dyads (d'Erfurth et al., 2010, 2009).

Through a comprehensive study spanning multiple years and several populations, we discovered a total of five QTLs regulating unreduced pollen production diploid potato. Two major-effect QTLs co-localised with genes sharing homology with *AtJAS*. We also identified deleterious sequence variants in those genes, namely a transposon insertion within an exon and an amino acid substitution within a domain that is highly conserved across vascular plants. Besides facilitating interploidy breeding for potato, our findings provide the first examples of natural genetic variation associated with unreduced pollen production.

Materials and Methods

Plant material

In this study, we used three diploid *S. tuberosum* mapping populations: FRW19-112, IVP16-560, and CE-XW, which resulted from the crosses IVP92-057-3 × IVP92-030-14, RH89-039-16 × IVP10-281-1, and C (USW 5337.3) × E (77.2102.37), respectively. Detailed information on these clones and their related pedigrees is provided in Fig. 1. In previous growing seasons the parental clones RH89-039-18, IVP92-057-3, C, and E produced unreduced pollen, which was confirmed to be functional through successful 4x × 2x crosses (Hutten et al., 1994a; Park et al., 2007).The three mapping populations have a shared ancestry, as C is a grand-parent of RH89-039-16 and both C and E are grand-parents of IVP92-057-3. In 2020, 94 descendants from population FRW19-112 were grow from true potato seeds in five litre pots and under 16 hours of light in an open ground greenhouse compartment with drip irrigation and an evaporative cooling system. In 2019, 124 descendants from population IVP16-560 were grown from seed tubers in the same greenhouse conditions. In 2020, this population was extended to 191 clones grown from seed tubers in a single row field (51°57'05.7"N; 5°38'02.5"E) planted the 17th of April. In 2020, 225 CE-XW seedlings grown under greenhouse conditions as described in (Clot et al., 2023b), were phenotyped for unreduced pollen production. In 2022, this population was extended to 268 clones grown from seed tubers in 19 cm pots at ambient temperatures and under 16 hours of light.

Analysis of male meiosis

To observe the male meiosis progression and chromosomal behaviour of the parental clones of population FRW19-112 and IVP16-560, we fixed floral buds of 3 to 4 mm parental clones in Carnoy solution (3:1 EtOH (99.8%): glacial acetic acid). We refreshed the Carnoy solution during the harvest day and kept overnight at 4 ºC. Next, we washed the buds twice with 70% EtOH and stored them at 4º C and followed the chromosome spread technique described by Jones and Heslop-Harrison (1996). After rinsing the fixated buds in water, we dissected the anthers and washed them a second time in water. The anthers were then incubated at 37 ºC for 1.2h in a 1:1 mixture of pectolytic enzymes and 10 mM citrate buffer (pH 4.39). After enzymatic digestion, we macerated a single anther on a slide in a small drop of 60% acetic acid and stirred it gently on a hotplate at 55 ºC for 1 min. We then flooded the slide with freshly made Carnoy solution and dried it on a hotplate for 2 min. Finally, we stained the slides with 12 mL of DAPI diluted in Vectashield (300ng/µL) and mounted them with a coverslip. We kept the slides 4 ºC until we observed them using a Axio Imager Z2 microscope equipped with an external light source (X-cite series 120 EXFO), a 120W high pressure metal halide lamp, a DAPI reflector and an Axiocam 506 camera. We conducted a comparative analysis of progression of meiosis by examining five anthers per clone spanning stages from pachytene to tetrad. For each anther, we recorded the meiotic stage of at least 200 and up to 250 meiocytes. Similarly, we compared the proportion of different MII plate orientations based on at least three and up to five anthers per clone. For each anther, 20 to 70 meiocytes at metaphase II were observed and classified as having normal, parallel or fused MII plates following Ramanna (1979). For each orientation, we performed a statistical comparison of clone specific medians using the nonparametric Kruskal-Wallis test and the Conover-Inman post hoc test (α = 0.05) corrected for multiple comparison with Holm procedure (Conover and Iman, 1979; Holm, 1979).

Phenotyping unreduced pollen production

For each growing season, we collected up to four pollen samples per clone on different days throughout the flowering period. The samples were composed of bulked pollen extracted using a vibrator pin (modified electric toothbrush) from two flowers at anthesis with freshly opened anther pores. We stained the pollen samples using a simplified version of Alexander staining (Peterson et al., 2010), and examined them under bright field using a Axiophot Zeiss microscope equipped with a Neofluar 10x/0.30 lens. For each pollen sample, we photographed four field views using a Zeiss Axiomcam ICc 5 colour camera and Zen 2.3 lite software. We considered these four images as a single observation and analysed them using Fiji (Schindelin et al., 2012) and a modified version of the macro developed by Tello et al. (2018) (Suppl. file 1) to determine pollen diameter. For each full-sib population, we delimited a symmetrical

interval of +/- 3 µm centred around the lowest mode of the pollen diameter across the entire offspring. Pollen falling within this interval was classified as reduced, while larger pollen was classified as unreduced. Smaller pollen was considered malformed and removed from the dataset. We then used this pollen classification to calculate the proportion of unreduced pollen for each observation. To ensure the reliability of this proportion, we excluded samples containing less than 50 pollen grains. We also excluded the observations of two triploid individuals identified in population FRW19-112 and one identified in population IVP16-560. Furthermore, we removed observations from 48 previously characterized desynaptic individuals from population CE-XW (Clot et al., 2023a). After these exclusions, we obtained a total of 81 phenotyped FRW12-112 individuals, 187 phenotyped IVP16-560 individuals (119 in 2019 and 175 in 2020) and 247 phenotyped CE-XW individuals (165 in 2020 and 218 in 2022).

BLUPs and heritability estimation

We used Generalized Linear Mixed Models (GLMM), allowing for a binomial distribution and random effects, to analyse the level of unreduced pollen production in all three populations. We fitted the model described by equation (1) to population IVP16-560 and CE-XW and the simplified model described by equation (2) to population FRW19-112. In both models, logit is the link function between the linear predictors η_{ijk} (or η_{ik}) and the observations π_{ijk} (or π_{ik}) of the *k*-th record of the *i*-th genotype in the *j*-th year, μ is the population mean, (G_i) is the random effect of genotype *i*, Y_j is the fixed effect of year *j* and $\left(GY_{ij}\right)$ is the genotype *i* by year *j* interaction. Models were fitted with the R package glmmTMB version 1.1.2.3 (Brooks et al., 2017). We used likelihood ratio tests to assess the significance of random effects via the chi-square statistic. We tested the significance of the fixed effect of year with the Wald's test and the t-as-z-approach implemented in glmmTMB. Finally, we extracted predictions for the random genotype effects (BLUPs) for single year and multiyear models and used them as phenotypes in the subsequent QTL analyses.

$$\text{logit}\left(\pi_{ijk}\right) = \eta_{ijk} = \mu + (G_i) + Y_j + \left(GY_{ij}\right) \quad (1)$$

$$\text{logit}(\pi_{ik}) = \eta_{ik} = \mu + (G_i) \quad (2)$$

We estimated broad-sense heritability was using equation (3) (Piepho and Möhring 2007) where σ_g^2 is the genotypic variance estimated using the GLMM equations described above and $\overline{v}_{\Delta..}^{BLUE}$ is the mean variance of a difference of two genotypic BLUEs estimated using the same equations but considering G_i as the fixed effect of genotype *i*.

(3) $$H^2_{\text{Piepho}} = \frac{\sigma^2_g}{\sigma^2_g + \bar{v}^{\text{BLUE}}_{\Delta..}/2}$$

Whole genome sequencing of parental clones

In a previous study (Clot et al., 2023b), we sequenced clone C and E with paired-end reads, quality trimmed those reads with fastp version 0.19.5 (Chen et al., 2018), aligned them to DM v6.1 with BWA-MEM v.0.7.17 (Li and Durbin, 2009) and called variants with bcftools v.1.13 (Danecek et al., 2021). Here, we sequenced the parental clones of population IVP16-560 with 150bp paired-end reads using BGISEQ-500 and retrieved sequencing reads of RH89-039-16 and IVP10-281-1 from BioProject ID PRJEB2504 and PRJEB36551 respectively. Following the alignment pipeline of (Clot et al., 2023b), we obtained a mean haploid coverage of 10.5x, 16.9x, 24.7x and 23.9x for clone RH89-039-16, IVP10-281-1, IVP92-057-3 and IVP92-030-14 respectively. We then visualised those alignments in candidate gene regions using IGV (Robinson et al., 2011). Finally, we extract variants identified in the vicinity of *Soltu.DM.07G014730* (+/- 500bp) and annotated them using snpEff v4.3 (Cingolani et al., 2012).

Genotyping by amplicon sequencing and linkage map construction

We generated marker data for all three populations using the amplicon sequencing strategy potatoMASH (Leyva-Pérez et al., 2022). DNA extraction, library preparation and bioinformatic processing of sequencing reads to obtain marker dosages were performed according to Leyva-Pérez et al. (2022), with the exception that we aligned short-reads to DM v6.1 (Pham et al., 2020) instead of DM v4.03. Unique amplicon alleles, referred to as haplotags, were used as genetic markers. We obtained a total of 890, 822 and 873 segregating markers for population FRW19-112, IVP16-560 and CE-XW respectively. Next, we used polymapR version 1.1.2 (Bourke et al., 2018) and followed the package vignette to automate the linkage map construction in all three populations. First, we removed uninformative markers with high missing rates (>20%) or strongly distorted transmission ratios and manually imputed missing parental dosages based on offspring segregation, when possible. We also removed poorly genotyped individuals missing more than 20% of markers. Next, we calculated recombination frequencies for all marker-pairs, clustered the markers into 12 chromosomal groups and ordered them via multidimensional scaling (Preedy and Hackett, 2016). Finally, we phased the markers into parental homologues using PolyoriginR version 0.0.3 (Zheng et al., 2021) with a recombination rate set at 1.25 cM/Mb. The final phased linkage maps were composed of 755, 655 and 737 markers segregating across 91, 186 and 233 individuals for populations FRW19-112, IVP16-560 and CE-XW respectively (Table 1). Using Marey maps (Chakravarti, 1991) (Suppl. Fig. 2-4), we confirmed that the genetic order of our linkage maps follow the expected

physical order of potatoMASH amplicons on DM v6.1 except for a previously described paracentric inversion on the long arm of chromosome *3* (Tang et al., 2022a).

QTL mapping

We combined BLUPs and phased linkage maps in polyqtlR version 0.0.9 (Bourke et al., 2021) to perform a QTL analysis following the package vignette. We estimated offspring Identity-by-descent (IBD) probabilities using a Hidden Markov Model with a prior error ranging from 0.01 to 0.05 per chromosome. We then performed an IBD-based QTL analysis with the QTLscan function by regressing genotypic BLUPs on the parental homologue probabilities. LOD significance thresholds were determined via permutation tests on the BLUPs values with N = 1000 cycles and $\alpha = 0.05$. Next, we identified markers significantly associated with unreduced pollen production and examined the effect of marker dosage on unreduced pollen production using the nonparametric Kruskal–Wallis test and the Conover-Inman post hoc test ($\alpha = 0.05$) corrected for multiple comparison with Holm procedure.

Candidate genes exploration

We delimitated a LOD minus 1.5 confidence interval around each QTL peak and extracted the physical position of makers flanking this interval. When those flanking markers corresponded to the first or last marker of the linkage group, the physical confidence interval was extended to the most distal coordinates of the chromosome. We then narrowed the candidate region for each major-effect QTL by intersecting confidence intervals across years and populations. Next, we combined the high confidence gene model annotation of DM v6.1 with their GOSlim annotations and descriptions based on their best hit with the *A. thaliana* proteome TAIR10. We then retrieved all DM v6.1 annotated genes within our candidate regions and filtered them by GOSlim terms for reproduction (GO:0000003) or cell cycle (GO:0007049). Finally, we manually reviewed the TAIR10 curator summary and computational description of those short-listed genes to identify candidate genes for unreduced pollen production.

Sequence variation in genes encoding for JASON-RELATED proteins

To retrieve *StJR1* alleles from DM v6.1 and Otava (Sun et al., 2022), we constructed a local blast database with these two assemblies using Nucleotide-Nucleotide BLAST 2.8.1+ (Camacho et al., 2009). We queried against this database the genomic sequence of *Solyc07g045010.3.1*, the *JR1* allele of *S. lycopersicum* cv. Heinz 1706 SL4.0 genome (Su et al., 2021), with default parameters. We manually curated the results by stitching together partial hits within 5-kb of each other and extracted genic and 500 bp upstream and downstream regions homologous to *Solyc07g045010.3.1*. We performed a multiple sequence alignment of *StJR1* alleles and

Solyc07g045010.3.1 coding sequence in Geneious Prime 2022.2.2 (https://www.geneious.com) using MUSCLE 3.8.425 (Edgar, 2004). Based on this alignment we translated *JR1* exons using the build-in translate function of Geneious Prime 2022.2.2. To investigate allelic variation in *StJR2*, we blasted on EnsemblPlants (Yates et al., 2022) with default parameters the amplicon allele associated with high level of unreduced pollen on chromosome *12* against RH89-039-16 v3 (Zhou et al., 2020). We then retrieved the protein sequence the *StJR2* alleles in coupling and repulsion phase with this amplicon. We aligned those protein sequence, together with a set of JASON-RELATED proteins from the panel of plants used by Erilova *et al.* (2009) using MUSCLE 3.8.425.

Results

Misorientation of metaphase II plate observed in parental clones

Mok and Peloquin (1975a) and Ramanna (1979) studied the meiosis of the historical clones C (USW5337.3) and E (77.2102.37), the parents of our CE-XW population. Both studies report the presence of parallel and fused metaphase II (MII) spindles and, more generally, the misorientation of MII plate causing the formation of unreduced pollen. It is also known that the unreduced pollen produced by RH89-039-16 are genetically equivalent to FDR (Park et al., 2007), a typical outcome of MII plate misorientation. Although we suspected that MII plate misorientation was also the cause of unreduced pollen production in clones RH89-039-16 and IVP92-057-3 due to their shared ancestry with clones C and E, we opted to validate this assumption through cytological observations. We used the chromosome spread technique to examine the meiotic progression and chromosomal behaviour of clones RH89-039-16 and IVP92-057-3 known to produce unreduced pollen, and of clones IVP10-281-1 and IVP92-030-14 that are not able to do so. As expected, clones RH89-039-16 and IVP92-057-3 produced a mixture of tetrads (50% and 30%), dyads (45% and 62%), and triads (3% and 7%), and clones IVP10-281-1 and IVP92-030-14 essentially produced tetrads (98% and 99%) (Suppl. Fig. 2a). For clones RH89-039-16 and IVP92-057-3, we also occasionally observed the presence of nuclear fusion during telophase II resulting in 2% and 1% of unbalanced products (Suppl. Fig. 2d). The meiotic progression of clone IVP92-057-3 was asynchronous, with early meiotic stages such as pachytene and diakinesis occasionally observed next to dyads, which could suggest occasional omission of the second meiotic division. This was not the case for the three other clones where only consecutive meiotic stages could be observed simultaneously (Suppl. Fig. 2).

After this general outlook on the progression of meiosis, we focused our attention on MII. During this stage, the chromosomal behaviour of the two clones producing unreduced pollen differ from the other two clones. Roughly 80% of the meiocytes of

clones IVP10-281-1 and IVP92-030-14 showed the expected perpendicular orientation of MII plates, while the remaining ~20% displayed a parallel orientation. In contrast, for the unreduced pollen producing clones RH89-039-16 and IVP92-057-3, only ~50% of meiocytes had normally oriented MII plates, with the other half of the meiocytes displayed either parallel or fused MII plates. Approximately 35% of their meiocytes exhibited parallel MII plates, which correspond to 1.75 fold increase when compared with the ~20% reported for IVP10-281-1 and IVP92-030-14. We observed an even more striking increase in proportion for the meiocytes with fused MII plates, which represented ~15% of RH89-039-16 and IVP92-057-3 meiocytes, while this phenotype was virtually absent in clones IVP10-281-1 and IVP92-030-14 (Fig. 1). Overall, misorientation of MII plates appears to be the main meiotic restitution mechanism of RH89-039-16 and IVP92-057-3. Nonetheless, the presence of additional restitution mechanisms such as partial omission of meiosis II and fusion of nuclei cannot be excluded.

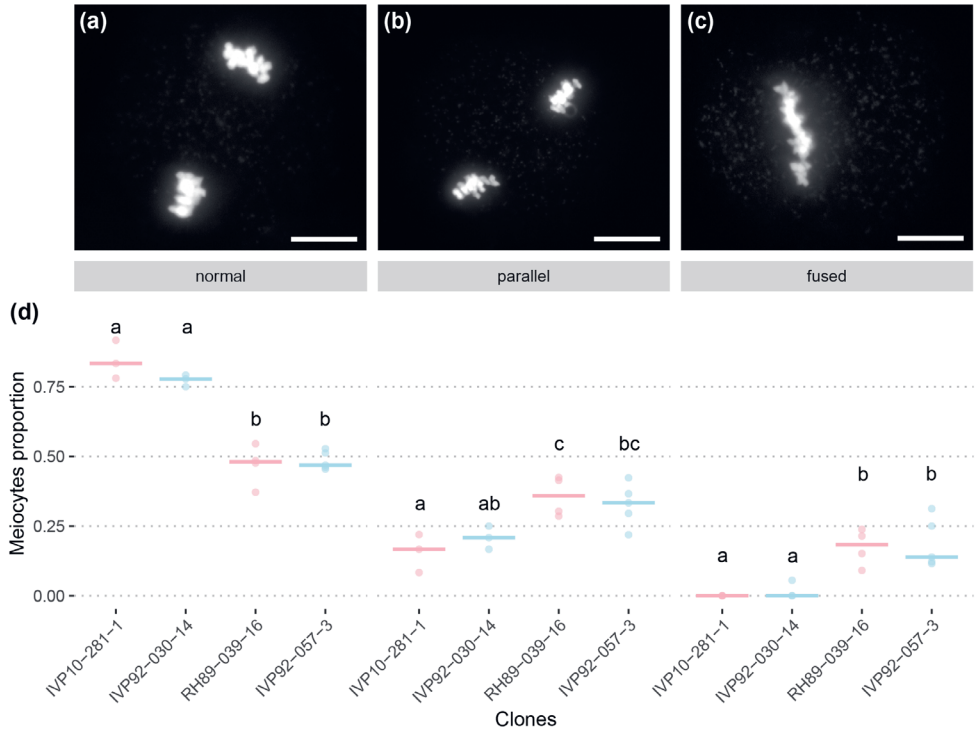

Figure 1: Distinctive metaphase II orientation phenotypes in parental clones. (a-c) DAPI-stained chromosome spreads of male meiocytes at metaphase II. (a) normal and (b) parallel orientation of metaphase II plates, (c) fused metaphase II plates. (d) Proportion of meiocytes with normally oriented (left), parallel (middle) or fused metaphase II plates (right) for the parental clones of population FRW19-112 in pink and IVP16-560 in blue. Median values per orientation and per clone are indicated by vertical bars. Significantly different median values between clones for a given orientation are indicated by different letters. Scale bars = 10 µm.

Proportion of unreduced pollen production is a heritable quantitative trait

Such detailed cytological observations are not scalable to entire populations. Therefore, we exploited the correlation between pollen size and ploidy level to phenotype unreduced pollen production using microscopic image analysis of pollen grains (Fig.1a-b). In the CE-XW population, the distribution of pollen diameter was bimodal (Fig. 1g), with the first mode corresponding to reduced pollen and the second mode to unreduced pollen. However, in populations FW19-112 and IVP16-560, the low level of unreduced pollen did not result in a visible second mode (Fig. 1c,e). The mode of reduced pollen diameter showed mild variation across populations, with values approximately centred around 19 µm, 20 µm, and 20.5 µm for CE-XW, IVP16-560, and FRW19-112, respectively. Taking this into consideration, we defined a symmetrical interval of +/- 3 µm around the mode of reduced pollen diameter in each population. Pollen falling within this interval was classified as reduced, while larger pollen was classified as unreduced. Next, we calculated the proportion of unreduced pollen for each observation and plotted the distribution of mean unreduced pollen production per clone for the three populations (Fig. 1d,f,h). Interestingly, all three distributions were skewed toward the left with 54%, 63% and 33% of the offspring in populations FRW19-112, IVP16-560 and CE-XW producing less than 5% of unreduced pollen. The highest levels of unreduced pollen production were observed in population CE-XW, with 26% of clones producing over 50% of unreduced pollen. Such levels of unreduced pollen were virtually absent in the other two populations. Overall, the three populations displayed a continuous distribution for the proportion of unreduced pollen produced. As it was challenging to define a threshold to classify descendant with or without unreduced pollen, we chose to treat the proportion of unreduced pollen as a quantitative trait.

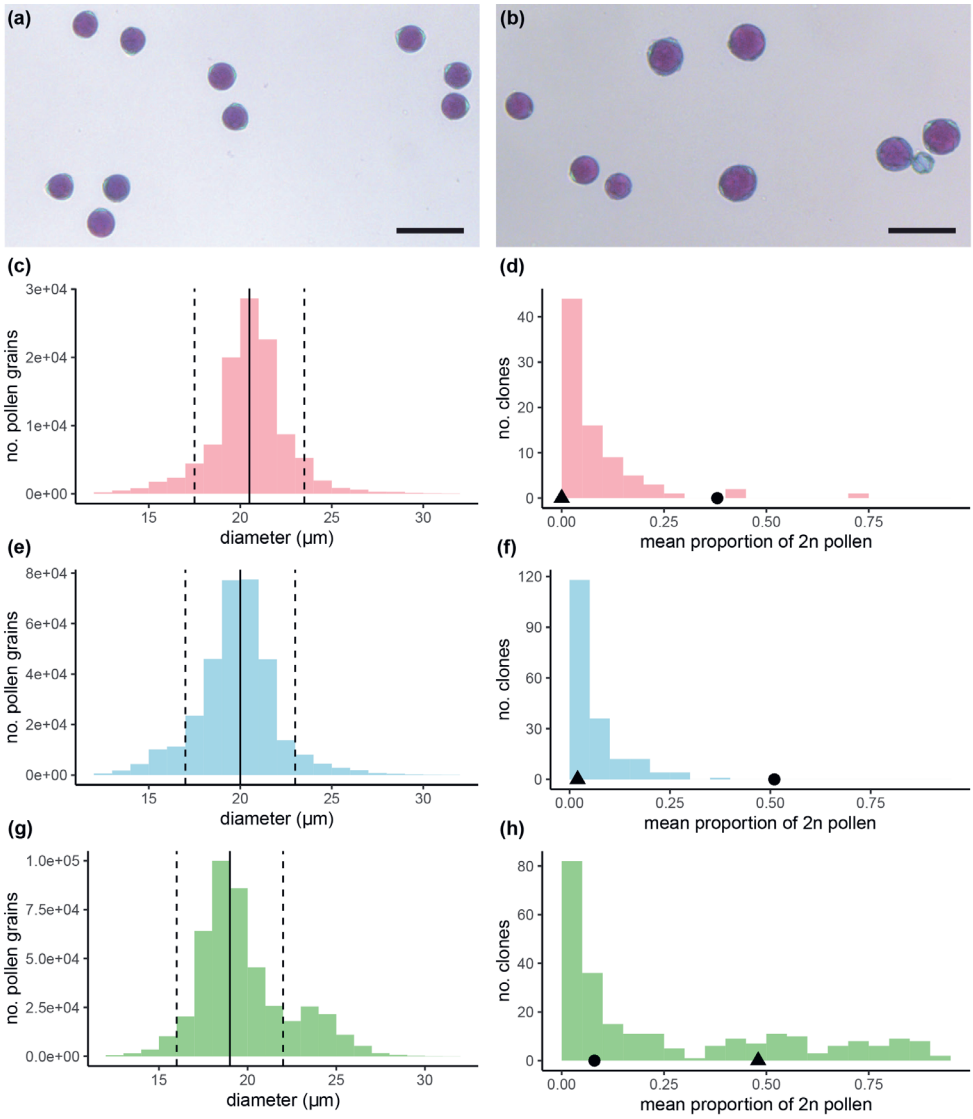

Figure 2: Microscope pictures of (**a**) normal reduced pollen, and (**b**) a mixture of reduce and unreduced pollen. The scale-bar represents 50 µm. Distribution of pollen diameter (**c**, **e**, **g**) and proportion of unreduced pollen (**d**, **f, h**) are shown in pink for population FRW19-112, in blue for IVP16-560, and in green for CE-XW. The vertical dashed lined represents the 3µm symmetrical interval centred around the mode of reduce pollen diameter indicated by the full vertical line. For each population, maternal and paternal mean unreduced pollen proportion are indicated by a circle and a triangle respectively. Scale bars = 50 µm.

As expected with discrete proportions, our data violated the linear model assumptions of normality and homoscedasticity of residuals. Indeed, the variance between repeated observations in time and across growing seasons was higher for clones producing over 10% of unreduced pollen than for those producing less (Fig. S6-S8). We therefore used

the GLMM Eq (1) (which is simplified to Eq (2) for FRW19-112) with a logit link and a binomial distribution to fit our data. In all three populations, the variable genotype explained most of the variance. Inclusion of the genotype by year interaction improved model accuracy for populations IVP16-560 and CE-XW (Table 1). The fixed effect of year was also significant for those two populations, particularly for population IVP16-560. Consistent with the visual examination of the data (Fig. 7), population IVP16-560 showed a higher intercept for the field trial of 2020 compared with the greenhouse one of 2019. We then extracted genotypic BLUPs for subsequent QTL analysis and estimated broad-sense heritability using the genetic variance (σ_g^2) and the mean variance of difference between BLUEs ($\overline{v}_{\Delta..}^{BLUE}$) as part of Eq (3). We treated genotypes as random effect while extracting genotypic BLUPs and estimating σ_g^2. On the other hand, we treated them as fixed effect while estimating $\overline{v}_{\Delta..}^{BLUE}$, which measures the precision of pairwise comparisons. We obtained a high broad-sense heritability in all three populations with estimates of 0.971 for FRW19-112, 0.896 for IVP16-560, and 0.923 for CE-XW, indicating a strong influence of the genotype on the level of unreduced pollen production.

Table 1: Partitioning of variance components and broad-sense heritability estimates for unreduced pollen production in the three mapping populations

	FRW19-112		IVP16-560		CE-XW	
Random Effect	**Variance Component**	**Std Dev**	**Variance Component**	**Std Dev**	**Variance Component**	**Std Dev**
genotypes	2.672	1.635	1.236**	1.112	4.239**	2.059
genotypes* year	NA	NA	0.429**	0.655	1.293**	1.137
Fixed Effect	**Estimate**	**Std Error**	**Estimate**	**Std Error**	**Estimate**	**Std Error**
Intercept	-3.341	0.184	-4.023	0.111	-1.816	0.166
2nd year	NA	NA	0.818**	0.094	-0.229*	0.136
$\overline{v}_{\Delta..}^{BLUE}$	0.158		0.286		0.714	
H^2_{Piepho}	0.971		0.896		0.923	

** Significant with p-value < 0.001; * Significant with p-value < 0.05

Several QTLs regulate unreduced pollen production, two co-localise with *AtJAS* homologs

We used an amplicon sequencing approach to genotype our three populations and generate phased linkage maps (Suppl. Table 1). We then identified QTLs for unreduced pollen production by regressing previously extracted genotypic BLUPs against parental homolog probabilities. Seven QTLs were identified: one in population FRW19-112, four in population IVP16-560, and two in population CE-XW (Table 2, Suppl. Fig. 9). Notably, we detected a large-effect QTL on the long arm of chromosome *12* in all three

populations (Fig. 2a), which explained 25.1%, 26.2%, and 29.0% of BLUPs variance in FRW19-112, IVP16-560, and CE-XW, respectively. Additionally, we identified two population-specific large-effect QTLs: q2nP_IVP_5 (for QTL 2n pollen IVP16-560 chromosome *5*) and q2nP_CE_7 (for QTL 2n pollen CE-XW chromosome 7), which explained 19.0% and 38.9% of BLUPs variance, respectively. All those QTLs were consistently detected across years in populations IVP16-560 and CE-XW. In population IVP16-560, we also found two lesser-effect QTLs on chromosome *9* and *11*, each explaining less than 10% of BLUPs variance.

Our amplicon sequencing method results in marker loci with multiple alleles (or haplotags). In all three populations, the haplotag representing the sequence of the DM reference genome, located at ~53.8 Mb on DM v6.1 chromosome *12* was associated with a high proportion of unreduced gametes. Individuals homozygous for this DM allele had the highest level of unreduced pollen with median value of 13.0% for FRW19-112, 42.2% for CE-XW and 7.5% for IVP16-560 (Fig. 2b-d). In the later population, individuals heterozygous or without the DM alleles how a similarly low proportion of unreduced pollen with median values of 2.7% and 2.4%, respectively, suggesting a recessive mode of inheritance. However, this could not be confirmed in population FRW19-112, where the inheritance appears to fit a dominant model despite the median being 2.5-fold higher for DM homozygotes than for heterozygotes. This discrepancy could be due to the smaller population size of FRW19-112 (n=77) and the transmission ratio distortion observed against the DM allele (Fig. 2b). We observed that FRW19-112 and IVP16-560 parents are heterozygous for the DM allele and therefore that *q2nP_FRW_12* and *q2nP_IVP_12* segregate from both parents. In contrast, in population CE-XW, *q2nP_CE_12* segregates from clone E only while clone C is homozygous for the DM allele suggesting that *q2nP_CE_12* is contributing to the formation of unreduced pollen of clone C. For *q2nP_CE_7,* the haplotag associated with elevated unreduced pollen production was also identical to DM and located at ~44.3 Mb on chromosome *7.* We observed a 10-fold increase in unreduced pollen production between individuals homozygous and heterozygous for the DM haplotag, indicating compatibility with a recessive model of inheritance (Fig. 2f). This QTL segregates from clone C only while clone E is homozygous for the DM allele suggesting that *q2nP_CE_7* is contributing to unreduced pollen production in clone E. In contrast, *q2nP_IVP_5* segregates from both parents and appears to follow an additive model (Fig. 2e). Finally, we analysed dosage variation at *q2nP_CE_7* and *q2nP_CE_12* simultaneously (Fig. 2g). Individuals heterozygous at for the DM haplotag at both loci produce the lowest proportion of unreduced pollen (median of 2.4%), while individuals homozygous for the DM haplotag at either locus produce a significantly higher proportion of unreduced pollen (median of 23.7% for *q2nP_CE_7* and 15.2% for *q2nP_CE_12*). Hence, taken independently, the effect of *q2nP_CE_12* was of the same order of magnitude as *q2nP_FRW_*12 and *q2nP_IVP_*12. Remarkably, individuals

homozygous for both QTLs produced a median level of unreduced pollen of 69.6%, a value 29-fold higher than that of double heterozygotes, suggesting at least additivity and potentially synergistic interaction between *q2nP_CE_7* and *q2nP_CE_12*.

To identify candidate genes associated with unreduced pollen production, we used a LOD minus 1.5 confidence interval around the major effect QTLs (Table 2) and narrowed three candidate gene regions by intersecting overlapping intervals. This resulted in a 4.5 Mb region on chromosome *5* (from ~51.1 Mb to ~55.6 Mb), a 2.8 Mb region on chromosome *7* (from ~43.4 Mb to ~46.2 Mb), and a 1.9 Mb region on chromosome *12* (from ~52.8 Mb to ~54.7 Mb). We then extracted DM v6.1 annotated genes in these regions and selected the ones predicted to be involved in cell cycle or reproduction based on homology with *A. thaliana* genes (Suppl. Table 2). We identified 454 genes within the candidate region on chromosome *5*, including 36 involved in cell cycle or reproduction. Among them, we found no genes with a direct link to unreduced gametes. However, we identified three genes worth noting: *Soltu.DM.05G022860* with homology to *AtPOR* involved in tubulin complex assembly and cytokinesis (Mayer et al., 1999), *Soltu.DM.05G023410* with homology to *AtRGL2* involved in gibberellic acid signalling and, *Soltu.DM.05G023970* with homology to *AtRAD50* involved in meiotic double strand break repair (Gallego et al., 2001). The candidate region on chromosome *7* contained 166 genes including 13 involved in cell cycle or reproduction. Among them we identified *Soltu.DM.07G014730* as the most likely candidate gene due to its homology with the meiotic regulator *AtJAS* whose mutants are known to produced unreduced pollen by restitution of the first division (de Storme and Geelen, 2011). Finally, in the candidate region on chromosome *12* we found 167 genes including 16 involved in cell cycle or reproduction. Strikingly, another gene, *Soltu.DM.12G023840,* with homology *AtJAS* was found. Following the naming convention of Cabout *et al.* (2017), we will refer to *Soltu.DM.07G014730* and *Soltu.DM.12G023840* as alleles of *StJR1* and *StJR2* (for *S. tuberosum JASON-RELATED 1* and *2*), respectively.

Table 2: Summary of QTLs for unreduced pollen production identified in the three mapping populations

QTL ID	population	size[1]	year	LOD	chr	pos[2] (cM)	PVE[3]	CI[4] (Mb)
q2nP_FRW_12	FRW19-112	78	2020	4.9	12	61.4	25.1%	51.4 - 59.5
q2nP_IVP_19_5	IVP16-560	113	2019	6.9	5	58.2	24.6%	51.1 - 55.5
q2nP_IVP_19_11	IVP16-560	113	2019	3.5	11	12.3	13.4%	0 - 41.3
q2nP_IVP_19_12	IVP16-560	113	2019	6.2	12	44.0	22.2%	10.2 - 55.5
q2nP_IVP_20_5	IVP16-560	170	2020	7.5	5	58.2	18.3%	50.2 - 55.5
q2nP_IVP_20_9	IVP16-560	170	2020	4.7	9	6.8	11.9%	0 - 3.7
q2nP_IVP_20_11	IVP16-560	170	2020	4.8	11	40.1	12.3%	10.8 - 43.1
q2nP_IVP_20_12	IVP16-560	170	2020	5.5	12	42.7	13.9%	50.2 - 54.7
q2nP_IVP_5	IVP16-560	180	combined	7.5	5	58.2	19.0%	51.1 - 55.5
q2nP_IVP_9	IVP16-560	180	combined	4.7	9	6.8	9.0%	0 - 64.6
q2nP_IVP_11	IVP16-560	180	combined	4.8	11	40.1	9.5%	10.0 - 46.8
q2nP_IVP_12	IVP16-560	180	combined	5.5	12	42.7	16.2%	10.2 - 55.5
q2nP_CE_20_7	CE-XW	117	2020	10.7	7	42.6	34.2%	43.4 - 46.2
q2nP_CE_20_12	CE-XW	117	2020	5.6	12	61.1	19.9%	52.8 - 58.5
q2nP_CE_22_7	CE-XW	171	2022	17.2	7	43.6	37.1%	43.4 - 46.2
q2nP_CE_22_12	CE-XW	171	2022	13.0	12	61.1	29.6%	52.8 - 56.2
q2nP_CE_7	CE-XW	187	combined	19.6	7	42.6	38.4%	43.4 - 46.2
q2nP_CE_12	CE-XW	187	combined	13.9	12	61.1	29.0%	52.8 - 55.5

[1] Size refers to the effective population size tested across years; [2] Pos refers to the genetic position of QTLs; [3] PVE refers to the proportion of variance explained by the QTLs; [4] CI refers to the LOD minus 1.5 physical confidence interval around QTLs

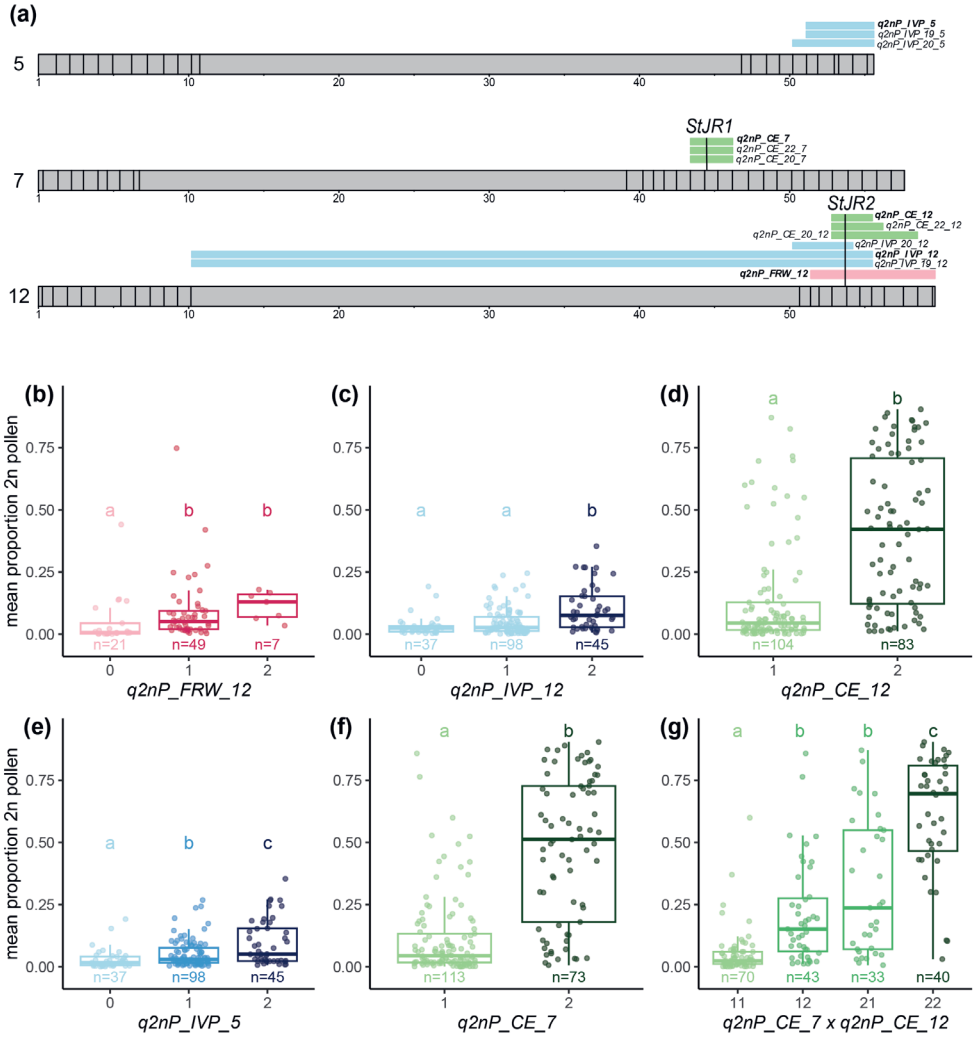

Figure 3: Position and effect of three major effect QTL involved in unreduced pollen production (**a**) Horizontal bars present the physical length of a potato chromosome (in Mb) including the marker positions in euchromatic chromosome arms as vertical lines. The LOD minus 1.5 confidence intervals around major-effect QTLs in population FRW19-112, IVP16-560 and CE-XW are shown in pink, blue and green respectively. (**b-g**) Dosage effect of haplotags associated with unreduced pollen production for each major-effect QTL. Each points represent the average level of unreduced pollen production of one clone. Significantly different medians trait values between haplotag dosage groups (α = 0.05) are indicated with different letters and colour shades.

Sequence variation across *StJR1* and *StJR2* alleles suggest functional defects

Sequence variation in the candidate genes *StJR1* and *StJR2* was examined to identify variants with putative effect on the functionality of the JASON-RELATED proteins. Upon

closer inspection of the annotation of the reference genome DM v6.1 in the *StJR1* region, we noticed that two neighbouring genes separated by 2,109 bp, *Soltu.DM.07G014730* and *Soltu.DM.07G014740*, were annotated as *AtJAS* homologs. The structure of those two genes, each with five exons, differed from the *JR1* gene structure having seven exons in *A. thaliana*, *S. lycopersicum* and in other potato clones such as the Otava. Once *Soltu.DM.07G014730* and *Soltu.DM.07G014740* were aligned with the corresponding genomic regions in Otava and with the coding sequence of the tomato *JR1* gene *Solyc07g045010.3.1.*, we realized that the annotation in DM v6.1 was affected by a 16 bp deletion and a 2,103 bp insertion in the 5th exon of *StJR1* resulting in two partial *JR1* genes. We queried this insertion against the RepetDB database (Amselem et al., 2019) which classified it as a Class I LINE transposon. Hence, we will from now on refer to the DM allele of *StJR1 as StJR1.t1* where t stands for transposon. *In silico* translation of the disrupted 5th exon of *StJR1.t1* led to a premature stop codon (Fig. 4a). The visual examination of clones C and E short reads aligned to DM v6.1 confirmed that clone E is homozygous for the *StJR1.t1* while clone C is heterozygous (Suppl. Fig. 10). Next, we retrieved from the genome assembly of RH89-039-16, the maternal clone of population FRW19-112, *StJR2* alleles in coupling and repulsion phase with the chromosome *12* haplotag associated with high level of unreduced pollen. Assuming that *StJR2* is the causal gene behind *q2nP_FRW19-112_12*, the allele in repulsion phase with this haplotag (*RHC12H2G0793.2*, from now on *StJR2.1*) must be wild-type while the allele in coupling phase (*RHC12H1G0690.2*, from now on *StJR2.2*) must be mutant. We noted 18 amino acid changes between the StJR2.1 and StJR2.2 proteins one of which, the conversion of glutamine to an alanine (Glu>Ala) at position 432, was located within the highly conserved C-terminal domain of JASON-RELATED proteins (Fig. 4b) (Erilova et al., 2009). This Glu>Ala conversion is due to a T>G missense mutation on the second last nucleotide of *StJR2* 5th exon, located at 53,693,790 bp on DMv6.1, and annotated by snpEff as having a moderate effect on StJR2. Changing a polar Glu by a nonpolar Ala in such a conserved region is likely to affect the function of this protein. Consistently with the identification of QTLs for unreduced pollen production in this region of chromosome *12* in our three mapping populations, we found back the T>G SNP in a heterozygous condition in the six parental clones our three mapping populations (Suppl. Fig. 11). While this is explaining the bi-parental segregation observed for *q2nP_FRW_12* and *q2nP_IVP_12*, it is insufficient to account for the uniparental segregation observed for *q2nP_CE_12* and would therefore suggest that clone C is heterozygous for *StJR2.2* and yet another putative mutant allele.

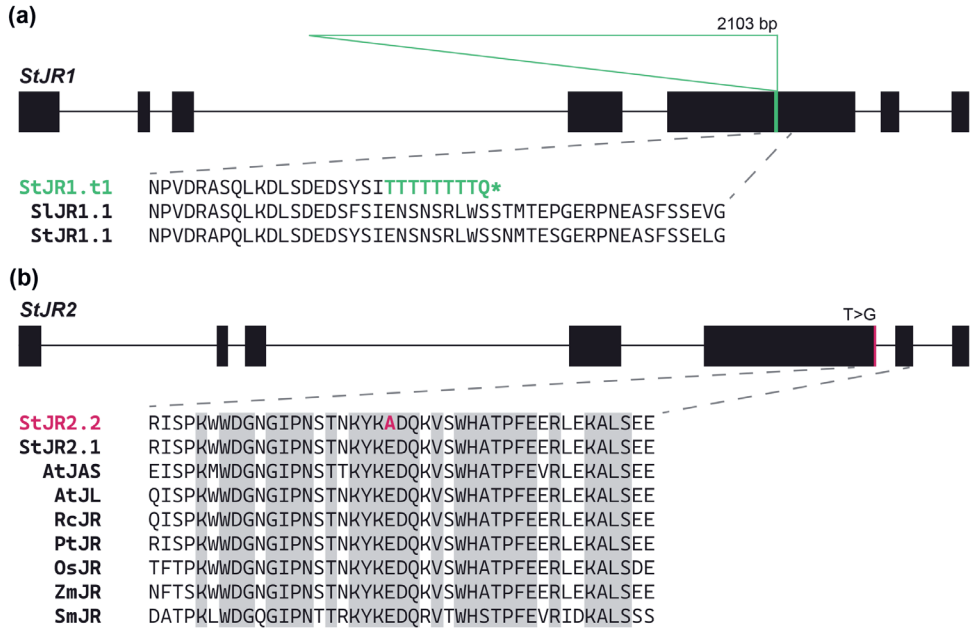

Figure 4: Functional and mutant *StJR1* and *StJR1alleles* (**a**) Diagram of the exon intron structure of *StJR1*. The position of the deletion identified in *StJR1.t1* is highlighted by a green stripe and the position of the insertion by a green triangle. The sequence of JR proteins overlapping with the indel site is displayed for the DM allele (*StJR1.t1*), the tomato *Solyc07g045010.3.1* allele (*SlJR1.1*) and the Otava *St2-St07G406640* allele (*StJR1.1*). (**b**) Diagram of the exon intron structure of *StJR2*. The position of the missense mutation associated with high level of unreduced gamete production in RH89-039-16 is highlighted by a pink stripe. The C-terminal sequence of a set of JASON-RELATED proteins overlapping with the missense mutation is displayed. Amino acid residues with 100% sequence conservation are highlighted in grey. At: *Arabidopsis thaliana*, Rc: *Ricinus communis*, Pt: *Populus trichocarpa*, Os: Oryza sativa, Zm: *Zea mays*, Sm: *Selaginella moellendorffii*.

Discussion

The intricate layers of complexity of unreduced pollen production in potato

Production of unreduced gamete is undeniably a complex trait, influenced both by environmental and genetic factors (de Storme and Geelen, 2014, 2013). In this study, the main cause of variation was genetic as indicated by the high broad sense heritability observed in our three-mapping population. However, even when environmental influence is accounted for, the genetic basis of unreduced pollen in potato remains complex and layered. The first layer of complexity is a mechanistic one. Unreduced pollen can result from various meiotic restitution mechanisms. While in mutagenized *A. thaliana* such mutants can be studied independently or in controlled combinations, in potato clones several meiotic alterations can co-occur (Ramanna, 1979). We aimed to reduce this variability by working with parental clones sharing ancestry with the

historical clones C and E, known to produced unreduced pollen via misorientation of MII plates and spindles (Mok and Peloquin, 1975a; Ramanna, 1979). We confirmed with cytological observations that clone RH89-039-16 and IVP92-057-3 indeed displayed this restitution mechanism. However, we also observed the co-occurrence of additional meiotic aberrations such as partial omission of the second division and nuclear fusion. The second layer of complexity is linked with the quantitative phenotype of unreduced pollen production. In all three populations variation in unreduced pollen production was continuous. Therefore, simple classification into producers and non-producers of unreduced pollen depends on an arbitrarily threshold. This has happened in earlier work, notably with inconsistent threshold values across studies (Iwanaga and Peloquin, 1982; Mok and Peloquin, 1975b, 1975a). This threshold based classification endorsed the conclusion that unreduced pollen production in potato was under the control of independent monogenic and recessive mutations, one affecting spindle orientation and two affecting cytokinesis (Mok and Peloquin, 1975b). In contrast, we adopted a quantitative approach, as previously advocated by Dongyu *et al.* (1995), and identified two minor-effect and three major-effect QTLs involve in unreduced pollen production. We also expect additional QTLs underlying MII plates and spindle orientation, but not segregating in our populations. Indeed, clones RH89-039-16 and IVP92-057-3 were both heterozygous at all QTLs positions while producing about 38% and 51% unreduced pollen, respectively. Overall, the involvement of at least 5 loci in unreduced pollen production follows the multiple mechanism of meiotic restitution and the involvement of several genes for each mechanism. The third layer of complexity is allelic heterogeneity. For any given QTL and the underlying candidate gene, more than one haplotype can be mutant, based on different sequence variants. This seems to be the case of *q2n_CE_12* for which both alleles of clone C contribute to unreduced pollen production, while this clone is heterozygous for the T>G missense mutation in *StIR2.2*, suggesting the presence of a second non-functional allele. Taken together, those layers of complexity explain why, prior to this study, the genetic regulation of unreduced pollen in potato has remained elusive for more than 50 years after the descriptions of an allegedly monogenic mendelian inheritance.

AtJAS homologs as candidate genes

The misorientation of MII plates and spindles is the most frequently reported mechanism of male meiotic restitution in plants (Bretagnolle and Thompson, 1995; de Storme and Geelen, 2013). Nonetheless, only three genes, *AtPS1* (d'Erfurth et al., 2008), *AtJAS* (de Storme and Geelen, 2011) and *AFH14* (Li et al. 2010), are known to be involved MII spindle and plate orientation. While *AFH14* regulates microtubule and microfilament arrays, *AtJAS* and *AtPS1* form a mini regulatory network controlling MII spindle orientation. Although similar in gametic outcome, the loss-of-function

phenotype of *Atjas1* differs from *Atps1* in the unique loss of the MII organelle band required to prevent spindles interaction (Brownfield et al., 2015). In *A. thaliana* two forms of AtJAS are produced: a long version with an N-terminal Golgi localization signal and a short version localizing to the plasma membrane (Cabout et al., 2017). The short version of AtJAS is upregulated during meiosis and therefore assumed to be the one responsible for the organelle band maintenance. Another protein, AtJL (for *A. thaliana* JASON-like), present a high degree of sequence similarity with AtJAS, notably within the conserved C-terminal domain, but lack the N-terminal Golgi extension (Cabout et al., 2017; Erilova et al., 2009). Although *AtJL* function is unknown, its predominant expression during reproductive development and its localisation to the plasma membrane suggest a meiotic function similar to *AtJAS*.

In this study, we identified the misorientation of MII plates as the principal meiotic restitution mechanism in the parents of our mapping populations. This phenotype fits well to our candidate genes in potato with homology to *AtJAS*: *StJR1* located within the candidate region of *q2nP_CE_7* and *StJR2* located within the shared candidate region of *q2nP_FRW_*12, *q2nP_IVP_*12 and, *q2nP_CE_*12. Candidate gene *StJR1* present the typical N-terminal extension of *AtJAS* and is therefore its putative ortholog. Mutagenized *AtJAS* alleles with a premature stop codon prior to the conserved C-terminal domain, in the 3rd or 5th exon, lead to the formation of unreduced pollen in *A. thaliana* (Erilova et al., 2009). Similarly, the transposon insertion in *StJR1.t1*, leading to a premature stop, is located upstream of this conserved C-terminal domain and is likely the causal polymorphism behind *q2nP_CE_7*. On the other hand, candidate gene *StJR2* is lacking the N-terminal Golgi localization signal and is therefore a putative *AtJL* ortholog. *AtJAS*, *AtJL* and their homologs in other plant species share a highly conserved C-terminal domain (Erilova et al., 2009). Although the function of this C-terminal domain is yet unknown, its conservation throughout vascular plants suggests its importance. In our three mapping populations, a Glu>Ala substitution within this conserved domain was associated with unreduced pollen production, but molecular validation is needed to demonstrate that this amino acid replacement is the causal mutation behind *q2nP_FRW_*12, *q2nP_IVP_*12 and, *q2nP_CE_12*. More generally, this would indicate that JR1 and JR2 proteins are equally important for proper MII plates and spindles orientation. Indeed, in single mutants one functional JR protein seems to partially compensate for loss of the other JR protein while in double mutants the severity of the phenotype drastically increases.

Interploidy breeding with unreduced gametes

Unreduced gametes have been successfully used in potato breeding program to introduce resistance genes to various pathogens from diploid wild relatives of potato to tetraploid landraces and cultivars (Ortiz *et al.* 1994; Capo *et al.* 2002). An interploidy breeding scheme combining the ease of diploid breeding and the heterozygosity of

tetraploids was proposed 60 years ago (Chase, 1963; Hutten et al., 1994a). However, such a breeding scheme has not been adopted widely, partly due to the lack understanding of the genetic control of unreduced gamete production (Dongyu et al., 1995; Ortiz and Peloquin, 1992). The identification of major-effect QTLs and candidate genes underlying unreduced pollen production enables, for the first time, marker assisted selection (MAS) for this trait. In addition, a dominant gene conferring self-compatibility to diploid potato was identified (Clot et al., 2020; Eggers et al., 2021; Hosaka and Hanneman, 1998a; Ma et al., 2021), prompting the potato breeding community towards a transition from tetraploid clonal selection to a diploid F_1 hybrid breeding system (Bethke et al., 2022; Jansky et al., 2016; Lindhout et al., 2016). F_1 hybrid breeding is still in its early stages, and large-scale breeding efforts are required to overcome inbreeding depression due to a high genetic load (Zhang et al., 2019, 2021). In this context, MAS for unreduced pollen production provides the opportunity to cross improved diploids with tetraploid elite material, allowing breeders to select clonally propagated tetraploid varieties, while the diploid F_1 hybrid system matures. MAS for unreduced pollen can also be combined with MAS for reduced meiotic recombinations (previously referred to as desynapsis) (Clot et al., 2023a; Jongedijk and Ramanna, 1989) to facilitate to the creation of diploid clones producing near non-recombinant FDR male gametes. These relatively uniform and heterozygous gametes could be used to produce relatively uniform tetraploid varieties grown from true seeds, by crossing a heterozygous diploid clone with a partially inbred tetraploid clone. Finally, the additivity of *q2nP_CE_7* and *q2nP_CE_*12 offers breeders the ability to fine-tune the level of unreduced pollen production in their clones. Although, the optimal level of unreduced pollen needed to simultaneously allow high seed set at the diploid and tetraploid level is currently unknown. Unpublished data from the crossing booklets of the Wageningen University potato breeding program indicate that for interploidy crosses, as little as 5% of stainable unreduced pollen is sufficient to induce a seed set comparable to a normal tetraploid cross. Similar observations have been made in *Rosa hybrida* (Gao et al., 2019) and could be explained by the competitiveness of unreduced pollen during certation associated with the masking of gametophytic mutational load (Husband, 2016). In this scenario, breeders interested in interploidy breeding may benefit from selecting clones bearing a single QTL for unreduced pollen production allowing for interploidy crosses while maintaining fertility at the diploid level.

Evolutionary relevance of unreduced gametes

Before being a useful breeding tool, unreduced gametes are precursor to significant evolutionary innovations such as polyploidy and apomixis (Bicknell and Koltunow, 2004; Tayalé and Parisod, 2013). Although the role of genetic mutations on unreduced pollen formation were demonstrated in *A. thaliana* (de Storme and Geelen, 2013), little is known about the natural genetic variation underlying this trait (Kreiner et al., 2017a).

To our knowledge, the candidate alleles *StJR1.t1* and *StJR2.2* are the first natural alleles known to be associated with unreduced pollen production. Kreiner et al. (2017b) extensively studied the rate of unreduced pollen formation across natural populations of *Brassicaceae* species and noticed the presence of rare individuals with elevated unreduced pollen production. Their rare occurrence and elevated level of unreduced pollen could be explained by the additivity of recessive mutations such as *q2nP_CE_7* and *q2nP_CE_12* identified in our mapping population. Mutations leading to unreduced pollen formation can be considered disadvantageous as they are expected to reduce male fitness in natural diploid population. As such, they are likely maintained in a dynamic equilibrium between the rate of mutation and the efficacy of purifying selection (Kreiner et al. 2017b). In vegetatively propagated species, purifying selection on sexual process is relaxed and can allow the maintenance of unreduced pollen mutations at a higher rate (Kreiner et al. 2017a). The ubiquity of unreduced pollen in tuber-bearing *Solanum*, as reported by Watanabe and Peloquin (1989), may be explained by this phenomenon. Those unreduced gametes have played a major role in shaping the tetraploid genome of cultivated potato. Firstly, because tetraploid cultivar are likely derived from the sexual polyploidization of early diploid landraces (Hardigan et al., 2017). Secondly, because they facilitate interploidy and interspecific gene flow and can therefore account for the prevalent wild *Solanum* introgression in tetraploid cultivars (Hardigan et al., 2017; Hoopes et al., 2022).

Author contribution statement

CRC: conceived research, collected and analysed phenotypical data, preformed linkage mapping, QTL analysis and allele mining, wrote manuscript. MdlOLP and LV: genotyped mapping populations. PB: advised on statistical modelling, linkage map construction, and QTL discovery. CJME and RCBH: created populations and advised on phenotypical observations. JvdB, EW and MJ: supervised cytological experiments. DM: supervised genotyping. RGFV: supervised research, edited manuscript. MJ: supervised research, edited manuscript. HJvE: obtained funding, conceived and supervised research, edited manuscript.

Acknowledgements

We thank the members of the public-private partnership “A new method for potato breeding: the ‘Fixation-Restitution’ approach” and “DIFFUGAT: Diploid Inbreds For Fixation, and Unreduced Gametes for Tetraploidy” (Averis Seeds B.V., Bejo Zaden B.V., Danespo A/S, Germicopa, Den Hartigh B.V., C. Meijer B.V., and Teagasc) for providing their support.

Supporting Information

Supplementary files are available at: https://github.com/ccrclot/PhD_thesis_Supp_info

Suppl. Fig. 1: Pedigrees of three bi-parental diploid potato populations. Parental clones of populations FRW19-112, IVP16-560 and CE-XW are highlighted in pink, blue and green respectively. Bold arrows represent dihaploidisation.

Suppl. Fig. 2: Marey maps in which genetic distance (cM) is plotted over physical distance (Mb). RH89-039-16 markers are displayed in blue, IVP10-281-1 markers in yellow and shared markers in black. Triangles indicate the position of CENH3-binding domains on DM v6.1.

Suppl. Fig. 3: Marey maps in which genetic distance (cM) is plotted over physical distance (Mb). IVP92-057-3 markers are displayed in pink, IVP92-030-14 markers in green and shared markers in black. Triangles indicate the position of CENH3-binding domains on DM v6.1.

Suppl. Fig. 4: Marey maps in which genetic distance (cM) is plotted over physical distance (Mb). C markers are displayed in orange, E markers in purple and shared markers in black. Triangles indicate the position of CENH3-binding domains on DM v6.1.

Suppl. Fig. 5: (a) Number of meiocytes at giving meiotic stage observed in the parental clones of population FRW19-119 (in shades of pink) and IVP16-560 (in shades of blue). Each row represents the meiocytes observed in a single anther. (b-e) DAPI-stained chromosome spreads of male meiocytes: tetrad (b), triad (c), dyad (d), nuclei fusion (e). Scale bars = 10 µm.

Suppl. Fig. 6: Distribution of the proportion of 2n pollen production across the clones of population FRW19-112.

Suppl. Fig. 7: Distribution of the proportion of 2n pollen production across the clones of population IVP16-560. Observations from 2019 and 2020 are shown in red and blue respectively.

Suppl. Fig. 8: Distribution of the proportion of 2n pollen production across the clones of population CE-XW. Observations from 2020 and 2022 are shown in red and blue respectively.

Suppl. Fig. 9: LOD profiles of QTL scan for genotypic BLUPs in FRW19-112 (a), IVP16-560 (b) and CE-XW (c). Significance thresholds, as determined by permutation tests (N = 1000, α = 0.05), are shown as dashed red lines (data were re-scaled so that these

overlap). For IVP16-560 (b) and CE-XW (c), data collected the first and second year are shown in red and blue respectively while the combined year data are shown in purple.

Suppl. Fig. 10: IGV view of the short reads of clone C and E aligned against the reference genome DM v6.1 in the region of StJR1.t1. StJR1.t1 is annotated as two genes due to a 2,103 bp insertion. While clone E is homozygous for StJR1.t1, clone C is heterozygous and display pair-end reads coloured in red with an insert size of more than 2,000 bp.

Suppl. Fig. 11: IGV view of the short reads of clone RH89-039-16, IVP10-281-1, IVP92-057-3, IVP92-030-14, C and E aligned against the reference genome DM v6.1. This view focus on the 5th exon of StJR2 where a T>G missense mutation leading to a Glu>Ala substitution in the highly conserved C-terminal domain of StJR2 is found in all 6 clones.

Suppl. Table 1: Overview of genetic lengths and marker densities per linkage group in the three mapping populations.

Suppl. Table 2: List of candidate genes predicted to be involved in cell cycle or reproduction within three QTLs candidate regions for unreduced pollen production.

Suppl. File 1: ImageJ macro pollen diameter extraction derived from Tello et al. (2018).

Chapter 6

General discussion

Despite the economic and nutritional importance of the potato crop, its genetic improvement lagged behind that of other major food crops. This limitation is primarily attributed to the difficulty of breeding an allogamous, heterozygous, and tetraploid crop. In this thesis, the objective was to explore the genetic basis of naturally occurring alterations of potato reproductive biology that can be combined to enhance the efficiency of potato breeding. I focused on three major targets: self-compatibility, crossover reduction and unreduced pollen production, and investigated their inheritance using a forward genetics strategy. In the first section of this final chapter, I will provide a reflection on the methods and resources used throughout this thesis. I will begin by discussing the advantages and limitations offered by forward genetics in comparison to reverse genetics within the context of this. Recognizing the increasing amounts of FAIR-ly available genomic data, I will argue that "translational genomics" could speed up the discovery of valuable natural mutants in potato. In the second section of this chapter, I will begin by emphasizing how the knowledge gain on the Self-incompatibility locus *Sli* has facilitated the development of self-compatible diploid germplasm. I will then proceed to discuss the challenges associated with inbreeding depression that must be overcome to generate inbred lines out of heterozygous *Sli*-bearing clones. In the third section, acknowledging the difficulty of this task, I will take a step back and question whether the breeding advantages associated with inbred lines are dependent on genome-wide homozygosity. I will then argue that, in the context of Fixation-Restitution breeding, both incremental improvement of progenitor clones and exploitation of heterosis can be achieved without relying on highly homozygous inbred lines. In the fourth section of this chapter, I will emphasize fundamental differences between the different potato breeding strategies proposed thus far. I will highlight how Fixation-Restitution breeding encompasses multiple advantages and show how its characteristics can position it as a "bridging approach" between different breeding strategies. I will conclude this section by providing practical recommendations regarding the reproductive mutations identified in this thesis. Additionally, I will highlight some of the remaining questions that need to be addressed to ensure the viability of innovative potato breeding strategies. Finally, in the last section of this chapter I will provide an outlook on reproductive biology research in potato, with a particular focus on the identification of deleterious sequence variants that contribute to male sterility.

Reflections on the methods and resources employed in this thesis

Was forward genetics the most appropriate strategy?

In this thesis, we employed a forward genetics strategy, specifically a *k*-mer based bulk-segregant analysis in **Chapter 2** and QTL mapping approaches in **Chapter 3-5**, to

identify loci that regulate reproductive traits relevant to potato breeding. Our choice of forward genetics, as well as these methods and target traits, reflects the history and expertise of Plant Breeding at Wageningen University & Research where this thesis was produced. Firstly, the reproductive biology of potato has been a topic of interest for Plant Breeding for over 50 years (Hermsen and Verdenius, 1973). In this context, diploid potato clones exhibiting self-compatibility (Olsder and Hermsen, 1976), crossovers shortage (Jongedijk and Ramanna, 1989) and unreduced gamete production (Ramanna, 1979) have been well characterized. Although some of those traits have already been exploited in breeding programs (Hutten, 1994; Lindhout et al., 2011) their inheritance remained poorly understood. Secondly, Plant Breeding has an expertise in forward genetics, as demonstrated by the development of genetic association and mapping tools (Bourke et al., 2018; Prodhomme et al., 2019; Stam, 1993; van Eck et al., 2017; van Os et al., 2005) and by the identification of natural allelic diversity causing variation in major breeding targets in potato such as foliage maturity and tuber shape and flesh colour. High resolution mapping of those traits facilitated the cloning of *StCDF1* (Kloosterman et al., 2013) *Zep* (Wolters et al., 2010) and *StOFP20* respectively (van Eck et al., 2022; Wu et al., 2018).

In this thesis research, as in those earlier studies, forward genetics proved to be an effective approach. We successfully developed KASP markers predicting self-compatibility (**Chapter 2**) and crossovers shortage (**Chapter 4**) and identified major-effect QTLs regulating pollen shed (**Chapter 3**) and unreduced pollen production (**Chapter 5**). However, forward genetics is a phenotyping-intensive and therefore time-consuming approach. This is exemplified by **Chapter 5** where we phenotyped three bi-parental populations across multiple growing seasons. The combination of lengthy forward genetics with the multitude of phenotypes investigated during this thesis project was at the expense of fine mapping and functional validation of the candidate genes we identified. Instead, we focused on discovering sequence variants with predictive power that enables marker assisted selection (MAS) of these phenotypic traits in breeding programs.

Could we have used reverse genetics instead? Knowing that the genetic basis of reproductive traits is extensively studied in model organism, we could have used reverse genetics to obtain comparable results, potentially within a shorter timeframe. For instance, the CRISPR–Cas9 system was used to generate self-compatible diploid potato clones by knocking out the *S-RNase* gene responsible for self- incompatibility (Enciso-Rodriguez et al., 2019; Ye et al., 2018). Leveraging the knowledge of meiosis from *A. thaliana*, we could have attempted to engineer diploid potato clones producing non-recombinant First Division Restitution (FDR) unreduced pollen via genome editing. For example, knocking-out the potato homolog of *AtSPO11-1* (Grelon et al., 2001) would have likely produced clones lacking meiotic recombinations. Likewise, knocking-

out the potato homologs of *AtJAS* (de Storme and Geelen, 2011) or *AtPS1* (Erilova et al., 2009) would have likely generated clones producing FDR unreduced pollen. Alternatively, transiently repressing the expression of those meiotic genes using virus-induced gene silencing could also have also been a viable option (Calvo-Baltanás et al., 2022; Calvo-Baltanás et al., 2020). Such reverse genetics approaches have the advantage of demonstrating (with stronger evidence than our genetic mapping approaches) the involvement of a specific gene in a phenotype. However, these methods would not have allowed us to investigate the natural allelic diversity underlying our target phenotypes. Concretely, we would have missed the prevalence of *Sli* (**Chapter 2**) and *StMSH4* mutant alleles (**Chapter 4**) in potato tetraploid germplasm and overlooked the polygenic basis of unreduced pollen production (**Chapter 5**). Moreover, the current EU regulation considers genome-edited organism as GMOs in the sense of Directive 2001/18/EC and as such their release into the environment is bound to many regulatory permits and considerations (van der Meer et al., 2023).

The potato genomics revolution facilitates allele mining

Due to the challenges of assembling heterozygous and polyploid genomes, the potato genetics community relied for about decade on the genome assembly of the double monohaploid clone DM derived from a *S. tuberosum* group Andigenum (formerly *S. phureja*) accession (PSGC, 2011). Although the DM reference genome was improved over time (Pham et al., 2020; Sharma et al., 2013; Yang et al., 2023), it remained a single-haplotype assembly of an unimproved clone which is insufficient to capture the genomic diversity of cultivated potato. In recent years, the development of long-read sequencing and long-range information technologies, coupled with algorithmic advancement, has accelerated the process of genome assembly (Pucker et al., 2022). In addition, the phasing of sequence reads into heterozygous assemblies was facilitated by several strategies exploiting information derived from the parental genomes or derived from gametes and offspring genomes (Koren et al., 2018; Shi et al., 2019; Zhou et al., 2020). These methodological advancements resulted in the release of two new diploid potato assemblies in 2020, one heterozygous and one homozygous (van Lieshout et al., 2020b; Zhou et al., 2020), and of eight phased tetraploid genome assemblies in 2022 (Bao et al., 2022; Hoopes et al., 2022; Sun et al., 2022). Taken together, those assemblies represent a set of 36 haplotypes per chromosome which offer a better glimpse into the allelic diversity of *S. tuberosum*.

Throughout this thesis, we have increasingly used those assemblies to explore sequence variation in the various candidate genes we identified. In **Chapter 4**, this strategy allowed us to identify *StMSH4.f1* and *StMSH4.t1*, two mutant alleles of *StMSH4* harbouring insertions of different lengths both located at the same position in second exon of this gene. Originally, we only identified *StMSH4.f1* in our mapping population. However, the *in silico* discovery of *StMSH4.t1* allowed us to confirm the

association of both mutant alleles with crossover shortage via a complementation test. Likewise in **Chapter 5**, we exploited the phased genome assembly of clone RH89-039-16, the parent of our mapping population FRW19-112, to extract the mutant allele of the candidate gene *StJR2* associated with unreduced pollen production. We could also easily compare different alleles of the other candidate gene *StJR1* and highlight the presence of an exonic transposon insertion both in the DM assembly and in our mapping population CE-XW. More generally, the availability of several refence genomes limits the effect of reference bias which can be as drastic as presence-absence variation (PAV) of the causal gene. This was the case for the *StOFP20* gene regulating tuber shape which can be found in the assemblies of clones M6 and RH89-039-16 but is absent in the assembly of DM (van Eck et al., 2022; Wu et al., 2018; Zhou et al., 2020).

Returning to the previous section where I discussed reverse and forward genetics, I want to highlight that the growing number of high-quality genome assemblies enables a different kind of reverse genetics strategy based on gene annotation and allele mining. This strategy broadly defined as "translational genomics" (Kang et al., 2016) relies on the conservation of gene order and sequence homology across closely related species to transfer gene annotation from one species to another. Based on this annotation, researchers can identify genes susceptible to regulate a trait of interest. Sequence variants predicted to disrupt the function of such genes can then be identified with prediction tools such as snpEff (Cingolani et al., 2012) and subjected to phenotypical inquiry. For instance, entire genome annotation can be transferred between assemblies of closely related species such as tomato and potato using Liftoff (Shumate and Salzberg, 2021). Homologous genes identified between assemblies can then be compared to detect PAV and sequence variants using LiftoffTools (Shumate and Salzberg, 2022). This strategy can also be applied when causal genes for a given phenotype are unknown, but QTLs and candidate genes are reported in the literature (Pancaldi et al., 2023, 2022). In retrospect, this approach would have identified the mutant alleles *StMSH4.f1, StMSH4.1* (**Chapter 4**) and *StJR1.t1* (**Chapter 5**), based on their homology with the well described *A. thaliana* genes *AtMSH4* (Higgins et al., 2004) and *AtJAS* (Erilova et al., 2009) and their insertion polymorphisms resulting in premature stop codons. In that way, the lengthy forward genetics approach could have been circumvented and it would have been possible to directly set up complementation tests and functional validation experiments. Last but not least, this strategy relies on natural genetic variation and is therefore not affected by the regulation set on genome-edited crops. Ironically, in this context, the high genetic load of potato is advantageous as it increases the likelihood of finding non-functional mutants of interest. However, as we will see in the subsequent sections, this genetic load poses significant limitations to inbreeding efforts in potato.

Converting an outbreeder into a selfer

The genetic basis of self-compatibility in potato

While self-compatible diploid *S. tuberosum* has been reported occasionally in the 70s (de Jong and Rowe, 1971; Olsder and Hermsen, 1976), it is the identification of the *Sli* locus in a *S. chacoense* accession in the late 90s (Hosaka and Hanneman, 1998a, 1998b) that truly marked the beginning of a long journey toward the development of diploid potato inbred lines. Although *Sli* could be introgressed from *S. chacoense* to *S. tuberosum* (Lindhout et al., 2011; Phumichai et al., 2006), its usage in breeding programs was limited by the high levels of tuber glycoalkaloids and the long stolons associated with *S. chacoense* linkage drag (Enciso-Rodriguez et al., 2019; Leisner et al., 2018). Meanwhile, the lack of knowledge regarding the genetic basis of self-compatibility of *S. tuberosum* origin limited its practical utilisation. In **Chapter 2**, we discovered that the *Sli* haplotype was prevalent in the tetraploid cultivated germplasm. In retrospect, this finding could explain the earlier reports of self-compatible dihaploids potato clones. Recognizing that breeders would rather introgress *Sli* from *S. tuberosum* clones than from *S. chacoense* accessions, we developed and published KASP markers specific to the *Sli* haplotype. Those markers were quickly adopted by the potato genetics community (Kaiser et al., 2021; Song and Endelman, 2022). It was subsequently shown, both in *S. chacoense* and *S. tuberosum* clones, that the *Sli* gene encodes for a pollen-expressed F-box protein functioning as a general S-RNase inhibitor (Eggers et al., 2021; Ma et al., 2021). The knowledge gained on *Sli* identity and on its prevalence in *S. tuberosum* germplasm facilitated the development of self-compatible diploid potato clones with agronomical value (Alsahlany et al., 2021; Kaiser et al., 2021; Song and Endelman, 2022).

Selfing of diploids results in the loss of vigour and fertility

Once self-incompatibility was overcome, potato breeders and geneticists realised that poor fertility and inbreeding depression were the next obstacles on the road to the development of diploids suitable for inbreeding and repetitive backcrossing. Inbreeding depression is characterized by a decrease of vigour and fertility due to an increase in homozygosity. This phenomenon can be largely explained by the presence of recessive deleterious mutations whose detrimental effects on phenotype are exposed through inbreeding (Charlesworth and Willis, 2009). Expanding this model to multiple loci, the two homologous chromosomes (homologs) of a heterozygous individual will contain different sets of deleterious and beneficial alleles. Most of the deleterious alleles of one homolog will be masked by the beneficial alleles of the other. Upon inbreeding, some deleterious alleles will reach fixations and their unmasking in the inbreds will contribute to a loss of vigour and fertility. In potato, severe inbreeding depression affecting vigour, yield and fertility has been long reported in tetraploid and diploid inbred

populations (de Jong and Rowe, 1971; Golmirzaie et al., 1998; Krantz, 1924; Phumichai and Hosaka, 2006) as well as in dihaploids (Hutten et al., 1995; Manrique-Carpintero et al., 2018). This is commonly explained as the result of the accumulation of deleterious mutations which remain masked more easily in tetraploids. Furthermore, purifying selection of sexual offspring is limited due to vegetative propagation and long generation time (Mirzaghaderi and Hörandl, 2016; Otto, 2007). Prediction of deleterious mutation based on resequencing data have highlighted the severity of the mutational load in potato and the correlation between the number of deleterious alleles and the overall level of heterozygosity (Achakkagari et al., 2022; Zhang et al., 2019, 2021).

Each of those mutations can have different effect size and inbreeding depression can result from the exposure of a few highly deleterious alleles as well as from the combined effect of many slightly deleterious mutations (Charlesworth and Willis, 2009). In breeding programs, the identification and removal of large-effect deleterious mutation can be facilitated by studying Transmission Ratio Distortion (TRD) (Zhang et al., 2019, 2021). This strategy was applied to three diploid potato selfing populations resulting in identification of 15 TRD loci associated with deleterious phenotypes including the abnormal development of crumpled offspring and chlorophyl deficiency in yellow margin offspring (Zhang et al., 2019). In **Chapter 3**, we used a similar approach in a backcross population and identified nine gametic and one zygotic TRD confirming the prevalence of large-effect deleterious mutations in potato. When those large-effect deleterious mutations are dispersed across the genome, they can be removed in populations of moderate size through phenotypical and marker assisted selection. However, when two large-effect deleterious mutations are tightly linked in repulsion phase, large segregating populations are needed to identify the rare recombinants containing the two favourable alleles on the same haplotype. This so-called Hill-Robertson effect (Hill and Robertson, 1966) is exemplified in potato by the detrimental recessive alleles *ws1* and *pa1* which remained linked in repulsion phase in a diploid F_2 population of 880 individuals and could only be recombined in an additional population of 1,200 descendants (Zhou et al., 2020).

Assuming that such efforts become routine and that large-effect deleterious mutations are ultimately removed from the potato inbreeding germplasm, other complications may arise. For example, in **Chapter 3**, we identified four pairs of loci displaying non-independent assortment of alleles. This type of genetic signal can be explained by various phenomena including Bateson–Dobzhansky–Muller incompatibilities (Bateson, 1909; Dobzhansky, 1936; Muller, 1942) and epistatic interactions between deleterious mutations. From a breeding point of view, the non-independent assortment of alleles will limit the diversity of inbred lines that can be generated. Moreover, at this moment it seems that only a few lineages can reach a high degree of homozygosity. If indeed the

ability to sustain several rounds of selfing can only be established in a few clones with a specific genetic background, this would cause considerable genetic erosion. The genetic bottleneck, experienced by the first generations of potato inbred lines, could explain the limited heterosis recently reported in a large panel of about 800 experimental F_1 hybrid potatoes (Adams et al., 2022). Another point of concern is that the removal of large-effect deleterious mutations does not eliminate inbreeding depression as a whole. Low vigour can also result from the cumulative effect of many mildly deleterious alleles (Charlesworth and Willis, 2009; Wu et al., 2023; Zhang et al., 2021). The challenge of identifying and removing such minor-effect deleterious mutations is comparable to the challenge of identifying and combining the multitude of minor-effect QTLs underlying complex quantitative traits. According to genetic models that describe purging through inbreeding, mildly harmful alleles can take hundreds or even thousands of generations to be eliminated by natural selection (Wang et al., 1999).

Fixation-Restitution breeding alleviates the need for genome-wide homozygosity

From the previous section, it is evident that inbreeding with the goal to reach genome wide homozygosity is a challenging task. Although potato breeders can rely on genomic-assisted inbreeding, success is by no means guarantied. Therefore, it is worth reconsidering whether the creation of inbred lines is a necessary step for accelerating the fixation of genetic gains in potato breeding. The advantages of using inbred lines in breeding are twofold. Firstly, inbred lines can be incrementally improved throughout introgression and subsequent fixation of beneficial alleles. Secondly, inbred lines can be crossed to generate F_1 hybrids which facilitate the systematic exploitation of hybrid vigour. Importantly, neither of these aspects is inherently dependent on genome-wide homozygosity which is only a prerequisite to produce uniform F_1 hybrid varieties. The Fixation-Restitution breeding strategy proposes an alternative approach to achieve these goals without the need for full homozygosity. Incremental improvement can be achieved through repeated backcrossing using self-compatible diploid clones as recurrent parents. Hybrid vigour can be achieved to some extend by crossing diploids, but interploidy crosses offer more potential for specific combining abilities between up to four different alleles at the tetraploid level. These two aspects of incremental improvement and hybrid vigour as embodied in the steps of Fixation and Restitution will be further discussed in this section.

Accumulation of beneficial alleles in self-compatible recurrent parents

Backcross breeding is an effective method used to introgress one locus or a few loci controlling trait(s) of interest from an often wild or unimproved individual called the

donor parent, to an elite individual with agronomical value called the recurrent (or recipient) parent. After an initial cross between the donor and the recurrent parent, descendants are repetitively backcrossed with the recurrent parent. With each backcross generation, the proportion of the donor genome is halved. Once most of the donor genome is eliminated, the introgressed segment(s) can be fixed via one round of selfing. When an inbred line is used as recurrent parent, backcross breeding results in the formation of a Near Isogenic Line (NIL). This process allows breeders to introduce valuable traits into an elite germplasm without altering its overall genetic composition.

In this context, the value of self-compatible diploids, that can tolerate a certain level of inbreeding and can be utilized as recurrent parents, is beyond dispute. However, recurrent parents do not necessarily have to be inbred lines. Self-compatible clones that are heterozygous or partially inbred can also be used for backcrossing. In this modified backcrossing scheme, the proportion of the donor genome will also be halved every generation. However, only half of the heterozygous recurrent parent genome will remain heterozygous, while the other half will be homozygous for either haplotype of the recurrent parent. Therefore, after a few backcrosses most of the donor parent alleles will be replaced by the elite alleles of the recurrent parent and the level of homozygosity will theoretically reach 50%. Finally, one final round of selfing (BC_nS_1) or full-sib mating (BC_nF_1) between advanced backcross individuals will allow the fixation of the introgressed donor locus. Depending on the strategy used, the theoretical homozygosity level is expected to reach 75% for BC_nS_1 and 60% for BC_nF_1. The resulting improved descendants can serve once again as recurrent parent for a new cycle of introgression. With every introgression cycle, the recurrent parent will gradually increase in homozygosity at each introgressed locus and at locus that can be fixed without affecting vigour and fertility.

The potential increase in mutational load due to increased homozygosity in backcross breeding is a valid concern. It is challenging to provide a definitive answer to whether this will lead to a dead end (fertility loss). However, insights can be gained from the expanding body of literature on inbred diploid potato. For instance, it is clear that phenotypic selection for vigour and fertility between generations of inbreeding maintain a higher level of heterozygosity than theoretical estimates (Leisner et al., 2018; Peterson et al., 2016; Phumichai et al., 2006; Song and Endelman, 2022; van Lieshout et al., 2020b). Likewise, selection of the most vigorous and fertile descendant during backcross breeding should maintain a level of heterozygosity relatively close to the one of the recurrent parent and thus limit the extent of inbreeding depression. Genomic regions with negligible impact on vigour and fertility could reach fixation while other regions with deleterious mutations in repulsion phase will remain heterozygous. Moreover, if a genome-wide marker system is used to facilitate the removal of the

donor genome during backcrossing, it can also be used to maintain high level of heterozygosity in the recovered recipient genome.

The main drawback of using non-inbred clones as recurrent parents is that backcross breeding cannot be done blindly. Breeders must ensure that each BC_nF_1 has maintained the desired agronomic characteristics of the original recurrent parent. This can be achieved by implementing a step of clonal selection at the conclusion of backcross breeding. Selection can be based on phenotype but can also be assisted by markers if the genetic basis of the desired traits is known. However, the need for this time-consuming clonal selection step can be minimized by employing partially inbred recurrent parents that have been previously fixed for beneficial alleles.

Recurrent parents are preferably the result of a diploid breeding program that is composed of elite diploids with good agronomical and quality traits. Interestingly, many domestication traits are associated with recessively inherited loss-of-function mutations (Dwivedi et al., 2023). If recessive inheritance also applies to quality traits (such as glycoalkaloid content, oxidative browning, and stolon dispersal), it implies the absence of dominant alleles for poor quality in a recurrent parent. Although recurrent parents may be fixed for recessive alleles that contribute to good quality, these alleles may not be identical-by-descent, and therefore, may not result in an elevated level of homozygosity. In summary, self-compatibility and moderate inbreeding tolerance, rather than genome-wide homozygosity, can be used to transform potato breeding from a historically slow and non-accumulative method into a rapid and incremental process.

Restitution of heterozygosity at the tetraploid level

Recurrent parents can be intercrossed to further accumulate desirable traits. For instance, resistance (*R*) genes that are linked in repulsion phase and cannot be fixed within a single genotype can be combined in hybrids derived from crosses between recurrent parents possessing complementary sets of *R* genes. The resulting hybrids are solely intended to be used as progenitor clones in interploidy crosses, because alleles at heterozygous loci will segregate in their haploid gametes, but their diploid FDR gametes will maintain most of parental heterozygosity. This brings forward the second element of Fixation-Restitution breeding: the use of FDR unreduced pollen for interploidy crosses. Except for a few distal loci affected by recombination, most beneficial alleles of the diploid progenitor will be transmitted to tetraploid offspring, among which superior descendants can be identified via clonal selection and vegetatively propagated as uniform clonal varieties.

In F_1 hybrid breeding heterosis must be maximized at the diploid level and relies on the identification of heterotic pools. This aspect is still in an early stage of development (Adams et al., 2022) and requires laborious testing of many inbred lines for general and specific combining ability (GCA and SCA). On the other hand, the level of heterosis in

Fixation-Restitution diploid hybrids is only marginally relevant since these clones are not varieties but rather progenitors of future tetraploid varieties. The true contribution to heterosis come from the interploidy cross. This 4x × 2x cross is an implicit SCA experiment where the male FDR gametes of the diploid parent are combined with the many unique allele combinations found in the female gametes of the tetraploid parent. Earlier studies have already reported that heterotic responses were striking in 4x × 2x progenies (Hutten et al., 1994a; Mendiburu and Peloquin, 1977, 1971). Therefore, it seems evident that heterosis can be effectively harnessed without the need to develop potato inbred lines.

How to design a more efficient potato breeding strategy?

A variety of proposed potato breeding strategies

In **Chapter 1** I introduced conventional potato breeding along with a variety of potential alternative breeding strategies. In this section, I would like to highlight the defining characteristics of the different breeding schemes and evaluate their merits. The **conventional breeding** strategy, which is still producing most modern cultivars, is based on clonal selection of progenies obtained by intercrossing highly heterozygous tetraploids. The first alternative is the **analytical breeding** strategy (Chase, 1963) which makes use of diploids to facilitate genetic studies and selection, before returning to the tetraploid level through interploidy crosses. As in conventional breeding, clonal selection is applied. Another alternative aims to breed for true potato seed (TPS) varieties. It was adopted by CIP (International Potato Centre) to solve logistic and phytosanitary issues of vegetative propagation in countries without a formal seed tuber system. I refer to this as the **CIP-TPS breeding** strategy. CIP-TPS varieties can be obtained from tetraploids, as well as from interploidy crosses as in analytical breeding. The more recently proposed **diploid F_1 hybrid breeding** aim to convert an obligate outbreeding species into a selfer (Jansky et al., 2016; Lindhout et al., 2011). In outbreeders the parental alleles are dispersed through meiosis across a segregating offspring, without the ability to reach fixation of genetic gains. The use of self-compatible diploid inbreds allows to fixe and accumulate beneficial alleles while removing detrimental ones. Subsequently, two inbred lines can be crosses to produce a uniform diploid F_1 hybrid variety that is released to the market as TPS. Finally, the **Fixation-Restitution breeding** strategy proposes to exploit self-compatible diploid clones as recurrent parents in repetitive backcrossing scheme to facilitate the accumulation of beneficial alleles. Making use of unreduced pollen, the resulting improved diploids can be crossed with elite tetraploid cultivars to generate heterotic tetraploid progenies. Depending on their uniformity, these progenies can become TPS varieties (**Fixation-Restitution TPS**) as in CIP-TPS breeding or be subject to clonal

selection (**Fixation-Restitution clonal**) to identify superior individuals which can become vegetatively propagated varieties.

Key differences between breeding strategies

Except classical breeding, all the above mentioned strategies are enabled by at least one and up to three alterations of potato reproductive biology (self-compatibility, crossover reduction, unreduced pollen) whose genetic bases were uncovered in **Chapters 2, 4 and 5**. Key differences between those breeding strategies are summarized in Table 1 and include genetic aspects and aspects related to the nature of the propagule (TPS or seed tubers). In this discussion, I do not wish to dive into the detailed agronomical, physiological, phytosanitary and logistic considerations associated with formal seed systems or TPS technology (Lindhout and Struik, 2023; van Dijk et al., 2022, 2021). Instead, I will focus on the genetic advantages and limitations offered by the different breeding approaches.

Table 1: Summary of key differences between potato breeding strategies

	reproductive alterations[1]	parents ploidy	variety ploidy	recurrent parents	inbreeding depression	propagule	uniformity
classical breeding		4x	4x	no	none	tubers	high
analytical breeding	2nP	4x, 2x	4x	no	none	tubers	high
CIP-TPS	(2nP, dsds)	4x, 2x	4x	no	limited	TPS	limited
diploid F_1 hybrid	SC	2x	2x	yes	high	TPS	high
Fix-Res clonal	SC, 2nP, (dsds)	4x, 2x	4x	yes	limited	tubers	high
Fix-Res TPS	SC, 2nP, (dsds)	4x, 2x	4x	yes	limited	TPS	limited

[1] Reproductive alterations exploited in each breeding strategy. SC indicates self-compatibility, 2nP indicates 2n (unreduced) pollen, dsds indicates desynapsis (crossover reduction). Optional alterations are denotated in between parenthesis. Fix-Res stands for Fixation-Restitution.

The fundamental distinction between these breeding strategies lies in their ability to enable fixation of beneficial alleles and incremental improvement of progenitor clones. Although using inbred lines can facilitate this process, as discussed is the previous section, genome-wide homozygosity is not an absolute requirement. Incremental improvement can also be achieved by using self-compatible heterozygous clones as recurrent parents in modified backcrossing schemes. Therefore, both F_1 hybrid breeding and Fixation-Restitution breeding strategies have the potential to significantly enhance potato breeding efficiency by facilitating the fixation of genetic gains.

Another important distinction lies in the homozygosity requirement imposed on progenitor clones. While the uniformity of clonal variety is ensured by vegetative propagation, the uniformity of TPS varieties relies on a minimal allelic segregation in the parental gametes. In diploid F_1 hybrid breeding, the uniformity of TPS varieties is

achieved via highly homozygous inbred parents. However, as mentioned earlier, high homozygosity in potato is accompanied by severe inbreeding depression. On the other hand, the uniformity expected from tetraploid TPS varieties has historically been set at a lower level and may not necessitate a similarly high level of homozygosity. Relatively uniform tetraploid progenies have been obtained in CIP-TPS and can be obtained via Fixation-Restitution TPS breeding by 1) fixating DUS (Distinctness Uniformity Stability) traits in the tetraploid female parent and 2) ensuring that the diploid male progenitor is homozygous for mutant *Stmsh4* and *Stjr1* or *Stjr2* alleles and thus produces near non-recombinant unreduced male gametes. Although homozygosity is not required for uniformity in Fixation-Restitution breeding, the diploid recurrent parents used in this breeding strategy must possess a moderate level of inbreeding tolerance to be incrementally improved via repeated backcrossing. These partially inbred clones can be fixated in genomic regions associated with agronomical value while other parts of the genome can remain heterozygous to minimize inbreeding depression.

Finally, the ploidy level by itself has important implications for gaining genetic knowledge and for breeding efforts. On the one hand, breeding at the diploid level is advantageous, because the distribution of trait values has more extreme values and is less balanced as compared with observations in tetraploids. This is due to a greater buffering of recessive alleles and a greater variation in allelic dosages at the tetraploid level. Besides of this phenotypical aspect, the fixation of beneficial alleles and the exposure and removal of recessive detrimental ones is more easily achieved in diploids (Haldane 1930). Moreover, and despite the constant improvement of software dedicated to polyploids (Rosyara et al. 2016; Bourke et al. 2018, 2021; Amadeu et al. 2021; Zheng et al. 2021), genetic studies in diploids are more straightforward than their tetraploid equivalent because of a simpler inheritance and a lower level of allelic interactions. For instance, in diploid populations complex traits such as yield can be precisely dissected with two-dimensional QTL scans including epistatic interactions between all pairs of loci (Marand et al. 2019). Conversely, the computational complexity of the equivalent analysis at the tetraploid level is prohibitive. Due to the genetic simplicity of diploids, several genes regulating important traits for potato breeding, such as tuber shape (van Eck et al. 1994), flesh colour (Wolters et al. 2010) and foliage maturity (Kloosterman et al. 2013), have been successfully fine mapped and ultimately cloned in diploid populations. Likewise, all the genetic analysis presented in this thesis were performed at the diploid level and tetraploid panels were only used for allele mining.

On the other hand, breeding for tetraploid varieties also has beneficial aspects. Firstly, the presence of four copies of each chromosome offers more space to combine beneficial alleles linked in repulsion phase. This is of particular interest for *R* genes which many reside in only a limited number of *R* gene clusters. For example, the wart

disease resistance genes *Sen1* and *Sen3* (Prodhomme et al., 2019), the late blight resistance gene *Rpi-cap1*, the PVY resistance $Ny_{(o,n)sto}$ (van Eck et al., 2017) and the root-knot nematode resistance $R_{Mc1\text{-}fen}$ (Draaistra, 2006) are located in *R* gene cluster(s) on the north arm of chromosome *11.* The challenge of combining all those genes in a single cultivar can be alleviated by bringing them on the different homologs of a tetraploid variety. Secondly, the four homologs of a tetraploid provide a better buffering against deleterious recessive mutations. Recent resequencing studies have revealed that tetraploid cultivars exhibit a higher number of deleterious mutations compared to diploid clones (Lian et al., 2019). However, they also exhibit a lower level of homozygous deleterious mutations (or expressed burden) due to inter-homolog complementation (Wu et al., 2023). Those findings agree with the often-reported correlation between heterozygosity and good field performances in potato (Mendoza and Haynes, 1974). An elegant demonstration of this phenomenon is provided by studies comparing the performance of tetraploid progenies derived from FDR and SDR unreduced gametes. FDR unreduced gametes are composed of homologous chromosomes and thus are estimated to transmit ~80% of the parental heterozygosity while this value drops to ~40% for SDR unreduced gametes, which are composed of siter-chromatids (Douches and Quiros, 1988a, 1988b; Jongedijk et al., 1991a; Peloquin et al., 2008). The increased transmission of heterozygosity permitted by FDR unreduced gametes has been demonstrated to enhance the yield and tuber size of progenies (Hutten et al., 1994a; Mendiburu and Peloquin, 1977, 1971). Although the correlation between heterozygosity and yield in potato is undeniable, there could be a threshold beyond which heterozygosity increase does not translate to higher yield (Sanford and Hanneman, 1982). It is yet to be determined whether heterozygous diploid clones can reliably offer sufficient buffering against deleterious mutations to maximize yield, or if a greater number of homologs are required for this purpose. Thirdly, tetraploid cultivars may benefit from the phenomenon of progressive heterosis whereby heterosis in double-cross tetraploid hybrids goes beyond the level seen in their single-cross parents, even though the ratio of positive and negative alleles has not been altered (Washburn et al., 2019; Washburn and Birchler, 2014). This phenomenon, unique to polyploid hybrids, remains poorly understood, and should be further investigated by potato geneticists considering its potential application in breeding. Finally, tetraploidy is associated with a larger cell and organ size which has been shown to increase biomass in *A. thaliana* (Corneillie et al., 2019) and in potato (Uijtewaal et al., 1987b).

The flexibility of Fixation-Restitution breeding

Taking all those differences into account, I believe that Fixation-Restitution breeding is a highly promising strategy that combines advantages of diploid and tetraploid breeding. Importantly, I do not see Fixation-Restitution breeding as opposed to other breeding

strategies but rather as a flexible system which, depending on the breeding goals and the evolution of the germplasm, can morph into analytical, CIP-TPS or diploid F_1 hybrid breeding (Fig. 1).

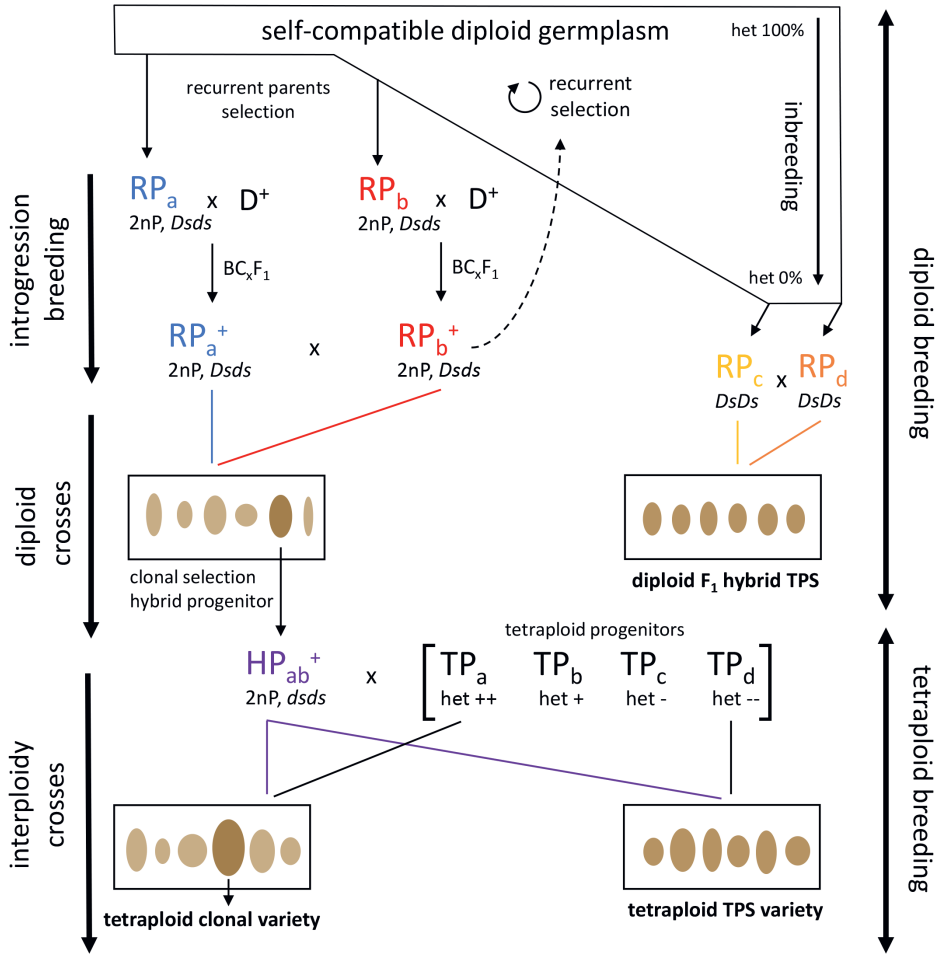

Figure 1: **Fixation-Restitution breeding scheme may morph into F_1 hybrid breeding.** Fixation-Restitution breeding begins with self-compatible diploid elite material selected for excellent fertility and agronomical value. In this germplasm mutations are selected or introduced at the *Sli, StJR1, StJR2* and *StMSH4* locus. This results in self-compatible recurrent parents with normal synapsis, a moderate proportion of FDR pollen. The diagonal box in the top indicates variable levels of homozygosity of diploids, where in the extreme left Fixation-Restitution may use heterozygous recurrent parents, and in the extreme right the uniformity of F_1 hybrids demands high homozygosity. Those recurrent parents (RP) can be incrementally improved by introgression of valuable alleles from a donor clone (D^+) and selection against the unwanted donor background during backcrosses. Improved recurrent parents (RP^+) can be reintegrated in the recurrent selection pool but can also be hybridised with other recurrent parents to create hybrid progenitor clones (HP^+). Depending on the level of heterozygosity (het) of the recurrent parents and the heterotic response observed, uniform progenies with agronomical potential can be commercialised as diploid F_1 hybrid TPS varieties. Alternatively, RP^+ and HP^+ clones can be used as male parents in interploidy crosses

with female tetraploid parents (TP). Again, depending on the heterozygosity of the tetraploid parent (het +/-), promising uniform progenies could be commercialised as tetraploid TPS varieties. In every case superior descendants can be selected and commercialised as clonally propagated tetraploid variety. To enable sexual polyploidisation, recurrent parents must be fixed for least one major-effect QTL for unreduced pollen production (2nP). Additionally, the uniformity of unreduced pollen (and of the resulting tetraploid progenies) can be improved by a reducing the number of crossovers. This can be achieved by selecting diploid fertile recurrent parent heterozygous for a recessive mutation in *StMSH4* (*Dsds*). By intercrossing two recurrent parents, one quarter of the progeny will be homozygous (*dsds*) and thus sterile at the diploid level but able to produce near non-recombinant unreduced pollen.

Within the context of Fixation-Restitution breeding, potato breeders can initiate the development of self-compatible and moderately inbreeding tolerant clones, which can be incrementally improved when used as recurrent parent. Once inbreeding depression is overcome and heterotic pools are defined, two highly homozygous and complementary diploid clones can be crossed to produce TPS propagated diploid F_1 hybrid varieties. Otherwise, as long as inbreeding depression remains a challenge and heterosis is lacking at the diploid level, superior diploid hybrid clones can be clonally selected from progenies lacking uniformity. By ensuring that the recurrent parents are fixed for *Stjr1* and heterozygous for *Stjr2* mutant alleles or vice versa (**Chapter 5**), these superior diploid hybrid clones can be used as male progenitors in interploidy crosses. Subsequent clonal selection within the resulting tetraploid progenies will identify descendants with variety potential that combined tetraploid heterozygosity with the beneficial alleles stacked in their diploid ancestors. Moreover, if TPS is preferred, tetraploid progenies with increase uniformity can be produced by fixating DUS traits in the tetraploid female parent and ensuring that the diploid male progenitor is homozygous for mutant *Stmsh4* alleles (**Chapter 4**).

Overall, Fixation-Restitution breeding lowers the homozygosity requirement necessary for using diploid self-compatible clones in breeding schemes of commercial potential. By connecting the historical tetraploid germplasm with the self-compatible elite diploid germplasm under development, Fixation-Restitution breeding has the potential to play a crucial bridging role in the ongoing potato breeding revolution. Once diploid F_1 hybrid breeding fulfils its promises, it may eventually render Fixation-Restitution breeding obsolete. However, if the advantages of tetraploidy and challenges associated with inbreeding depression prove to be too formidable, the recurrent parents and interploidy crosses of Fixation-Restitution breeding could deliver most of tomorrow's potato cultivars.

Remaining questions and practical recommendations

Many practical questions remain to be answered to assess the viability and commercial relevance of innovative potato breeding schemes. Some of these questions are shared between the Fixation-Restitution breeding and the diploid F_1 hybrid breeding strategies (de Vries et al., 2023) and include: "How to broaden and subdivide the diploid inbred

germplasm into heterotic and market class specific groups?", "What kind of marker system must be used to follow inbreeding and introgressions?" and "At which stage, with which intensity and in which environment selection for simple and complex traits must be applied?". Some questions are specific to the targeted propagule such as "How to develop a male sterility system facilitating TPS production?". Other questions are more specific to Fixation-Restitution breeding, for example: "How will vigour and fertility be transmitted along with quality traits during repetitive backcrossing using non-inbred recurrent parents that gradually increase in homozygosity?" or "How does mutational load, allelic diversity and gene dosages relate to heterosis at the diploid and the tetraploid level?".

To conclude, I would like to provide a few practical recommendations regarding the exploitation of the reproductive mutations identified in this thesis. Starting with self-compatibility, breeders can use the KASP markers developed in **Chapter 2** to assess the presence of *Sli* in their diploid germplasm. If this germplasm is composed of dihaploids, breeders can reasonably expect to identify *Sli* bearing clones. Otherwise, I recommend to introgress *Sli* from publicly available *S. tuberosum* clones such as RH89-039-16 rather than from the *S. chacoense* clones such as M6. Another important remark is that *Sli* is active gametophytically therefore only *Sli* bearing pollen will participate in self-fertilization and therefore self-compatibility will never be lost upon selfing. Regarding the *Stmsh4* mutants identified in **Chapter 4**, recommendations will depend on the targeted breeding strategy. Breeders only interested in diploid F_1 hybrid breeding should remove *Stmsh4* mutant alleles from their germplasm because homozygous plants are sterile. On the other hand, breeders interested in tetraploid TPS production would benefit from the increased level of heterozygosity and uniformity that reduced recombination provides to FDR unreduced pollen. Therefore, they are advised to maintain *Stmsh4* mutant alleles in a heterozygous condition in their diploid germplasm (using the KASP markers developed in **Chapter 4)** until fixation is required for interploidy crosses. Finally, regarding the major-effect QTLs regulating unreduced pollen production identified in **Chapter 5**, I would suggest focusing on QTLs *q2nP_CE_7* and *q2nP_CE_*12 and on their respective candidate genes *StJR1* and *StJR2* due to their association with FDR. One important aspect to keep in mind is that although unreduced pollen production is necessary for interploidy crosses, viable haploid pollen is also necessary to maintain fertility for crosses between diploids. By leveraging the recessive inheritance and additivity of the two QTLs identified, breeders have the opportunity to fine tune the level of unreduced pollen produced at different stages of their breeding program by fixating one or both QTLs. For example, different recurrent parents could be fixed for one QTL while the other QTL is kept in a heterozygous condition. Upon hybridisation of recurrent parents having complementary fixed QTLs, progenitor clones that are fixed for both QTLs and thus produce mostly unreduced pollen could be selected for interploidy crosses.

Outlook on reproductive biology research in potato

In this thesis, I addressed questions related to the genetic basis of reproductive traits in potato. While my work did not specifically focus of male fertility, this topic surfaced in all four experimental chapters, and deserves further attention. Poor male fertility is a historical problem in potato breeding (Krantz 1924) and the need to address it has only increased with the development of breeding schemes exploiting self-compatible diploids. Low male fertility can be caused by two distinct genetic phenomena: cytoplasmic male sterility (CMS) which is due to cytoplasmic nuclear interactions, and nuclear male sterility which is due to detrimental variants in nuclear genes.

CMS is observed in interspecific crosses between cultivated potato and some of their wild relatives such as *S. stoloniferum*, *S. demission* and *S. verrucosum* (Sanetomo and Gebhardt, 2015). A few years ago, a short anther phenotype associated with the interaction of a *S. tuberosum* group Andigenum cytoplasm and a *S. chacoense* nuclear genomic region was reported (Endelman and Jansky, 2016). More recently, a mitochondrial region associated with tetrad sterility in *S. stoloniferum* was identified (Sanetomo et al., 2022) Overall, research on CMS in potato seems to be gaining momentum due to its potential breeding and agronomical applications. Indeed, TPS production using CMS maternal lines can reduce hybrid seed production costs, because emasculation is not required and unintended release of the maternal inbred line can be prevented (Anisimova and Gavrilenko, 2017; de Vries et al., 2023). Moreover, the yield potential of diploid F_1 hybrids can be compromised by high berry production (Zhang et al., 2021) and the volunteer plants developing from true seeds can diminish the effect of crop rotation. Generating CMS F_1 hybrids would prevent both phenomena.

However, in this section, I think that it is more relevant to address nuclear male sterility because it will facilitate the maintenance of fertility upon inbreeding which is important for both the recurrent parents of Fixation-Restitution breeding and the inbred lines of diploid F_1 hybrid breeding. Successful male reproduction depends on a series of complex biological events that include specification of the stamen identity, anther cell division and differentiation, male meiosis, pollen grain development, anther dehiscence, pollen grain germination, pollen tube growth and double fertilisation (Gómez et al., 2015; Ma, 2005). Genetic mutations disrupting any link in this chain of tightly regulated events will compromise male fertility. The identification and removal of such major-effect mutations from the potato germplasm would facilitate innovative breeding approaches. This could be achieved using forward genetics as demonstrated by the identification of *Stmsh4* mutant alleles causing haploid pollen sterility in **Chapter 4**, QTLs for pollen shed in **Chapter 3**, and QTLs for unreduced pollen production (which imply a reduced proportion of haploid pollen) in **Chapter 5**. Other examples can be

found in the literature such as the identification via bulk segregant analysis of a mutant allele of *StDYT1* resulting in flower abortion in diploid potato (Zhang et al., 2021). However, as suggested in the earlier paragraphs of this chapter, it could be more efficient to adopt a "translational genomics" approach. Male fertility has been widely investigated in *A. thaliana* and in crops dependant on fertilisation for their harvested product (Liu et al., 2021; Ma, 2005; Wan et al., 2020; Wilson and Zhang, 2009). Moreover, pathways regulating floral organ identity, meiosis, tapetum and pollen grain development appear to be largely conserved across angiosperm (Gómez et al., 2015; Lei and Liu, 2020; Mercier and Grelon, 2008; Zheng et al., 2020). Based on the expectation that male fertility knowledge can be successfully translated from well-studied species to potato, a list of genes predicted to be involved in male fertility could be extracted from the annotation of a potato refence genome (preferably derived from a fertile inbred clone). After alignment of potato resequencing data to this reference genome and subsequent variant calling, the effect of sequence variants on gene expression and protein function involved in male fertility could be predicted. Taking the mutational load of potato into account, one can expect to identify many deleterious mutations. A strategy to prioritise their investigation and facilitate their removal from the germplasm will be necessary. Priority criteria could be based on their prevalence in the germplasm and/or on the predicted severity of their deleterious effect.

Interestingly, a by-product of this approach would be the identification of new reproductive mutants which could further contribute to innovative potato breeding strategies. For example, a non-functioning allele of the potato ortholog of *AtSPO11-1*, responsible for meiotic double strand-break formation, would more effectively abolish crossovers than *Stmsh4* alleles and thus contribute to the formation of non-recombinant unreduced gametes. In addition, some of the fertility mutations identified could underly the different mechanisms of unreduced megaspores formation reported in potato which include pseudohomotypic first meiotic division (Jongedijk et al., 1991b), omission of the second meiotic division (Werner and Peloquin, 1990) and post-meiotic genome doubling (Bastiaanssen et al., 1998). Gaining control over unreduced megaspore formation would open prospects for routine bilateral sexual polyploidisation in potato breeding.

References

Aboobucker SI, Zhou L, Lübberstedt T. 2023. Haploid male fertility is restored by parallel spindle genes in Arabidopsis thaliana. *Nat Plants* 1–5. doi:10.1038/s41477-022-01332-6

Achakkagari SR, Kyriakidou M, Gardner KM, De Koeyer D, De Jong H, Strömvik MV, Tai HH. 2022. Genome sequencing of adapted diploid potato clones. *Frontiers in Plant Science* 13:2838. doi:10.3389/fpls.2022.954933

Adams JR, de Vries ME, Zheng C, van Eeuwijk FA. 2022. Little heterosis found in diploid hybrid potato: The genetic underpinnings of a new hybrid crop. *G3 Genes|Genomes|Genetics* 12:jkac076. doi:10.1093/g3journal/jkac076

Almekinders CJM, Chujoy E, Thiele G. 2009. The Use of True Potato Seed as Pro-poor Technology: The Efforts of an International Agricultural Research Institute to Innovating Potato Production. *Potato Res* 52:275–293. doi:10.1007/s11540-009-9142-5

Alsahlany M, Enciso-Rodriguez F, Lopez-Cruz M, Coombs J, Douches DS. 2021. Developing self-compatible diploid potato germplasm through recurrent selection. *Euphytica* 217:47. doi:10.1007/s10681-021-02785-0

Amselem J, Cornut G, Choisne N, Alaux M, Alfama-Depauw F, Jamilloux V, Maumus F, Letellier T, Luyten I, Pommier C, Adam-Blondon A-F, Quesneville H. 2019. RepetDB: a unified resource for transposable element references. *Mobile DNA* 10:6. doi:10.1186/s13100-019-0150-y

Anisimova IN, Gavrilenko TA. 2017. Cytoplasmic male sterility and prospects for its utilization in potato breeding, genetic studies and hybrid seed production. *Russ J Genet Appl Res* 7:721–735. doi:10.1134/S2079059717070024

Anithakumari AM, Tang J, van Eck HJ, Visser RGF, Leunissen JAM, Vosman B, van der Linden CG. 2010. A pipeline for high throughput detection and mapping of SNPs from EST databases. *Molecular Breeding* 26:65–75. doi:10.1007/s11032-009-9377-5

Armstrong SJ, Caryl AP, Jones GH, Franklin FCH. 2002. Asy1, a protein required for meiotic chromosome synapsis, localizes to axis-associated chromatin in Arabidopsis and Brassica. *Journal of Cell Science* 115:3645–3655. doi:10.1242/jcs.00048

Aversano R, Scarano M-T, Aronne G, Caruso I, D'Amelia V, De Micco V, Fasano C, Termolino P, Carputo D. 2015. Genotype-specific changes associated to early synthesis of autotetraploids in wild potato species. *Euphytica* 202:307–316. doi:10.1007/s10681-014-1338-3

Bao Z, Li C, Li G, Wang P, Peng Z, Cheng L, Li H, Zhang Z, Li Y, Huang W, Ye M, Dong D, Cheng Z, VanderZaag P, Jacobsen E, Bachem CWB, Dong S, Zhang C, Huang S, Zhou Q. 2022. Genome architecture and tetrasomic inheritance of autotetraploid potato. *Molecular Plant* 15:1211–1226. doi:10.1016/j.molp.2022.06.009

Bastiaanssen HJM, Van Den Berg PMMM, Lindhout P, Jacobsen E, Ramanna MS. 1998. Postmeiotic restitution in 2n-egg formation of diploid potato. *Heredity* 81:20–27. doi:10.1046/j.1365-2540.1998.00370.x

Bateson W. 1909. Heredity and variation in modern lights. *Darwin and modern science*.

Berchowitz LE, Francis KE, Bey AL, Copenhaver GP. 2007. The Role of AtMUS81 in Interference-Insensitive Crossovers in A. thaliana. *PLOS Genetics* 3:e132. doi:10.1371/journal.pgen.0030132

Bethke PC, Halterman DA, Francis DM, Jiang J, Douches DS, Charkowski AO, Parsons J. 2022. Diploid Potatoes as a Catalyst for Change in the Potato Industry. *American Journal of Potato Research* 1–21. doi:10.1007/s12230-022-09888-x

Bicknell RA, Koltunow AM. 2004. Understanding Apomixis: Recent Advances and Remaining Conundrums. *The Plant Cell* 16:S228–S245. doi:10.1105/tpc.017921

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114. doi:10.1093/BIOINFORMATICS/BTU170

Bomblies K. 2022. Learning to tango with four (or more): the molecular basis of adaptation to polyploid meiosis. *Plant Reprod*. doi:10.1007/s00497-022-00448-1

Bourke PM, van Geest G, Voorrips RE, Jansen J, Kranenburg T, Shahin A, Visser RGF, Arens P, Smulders MJM, Maliepaard C. 2018. polymapR—linkage analysis and genetic map construction from F1 populations of outcrossing polyploids. *Bioinformatics* 34:3496–3502. doi:10.1093/bioinformatics/bty371

Bourke PM, Voorrips RE, Hackett CA, van Geest G, Willemsen JH, Arens P, Smulders MJM, Visser RGF, Maliepaard C. 2021. Detecting quantitative trait loci and exploring chromosomal pairing in autopolyploids using polyqtlR. *Bioinformatics* 37:3822–3829. doi:10.1093/bioinformatics/btab574

Bradshaw JE, Bryan GJ, 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

Bretagnolle F, Thompson JD. 1995. Gametes with the somatic chromosome number: mechanisms of their formation and role in the evolution of autopolyploid plants. *New Phytologist* 129:1–22. doi:10.1111/j.1469-8137.1995.tb03005.x

Brooks ME, Kristensen K, van Benthem KJ, Magnusson A, Berg CW, Nielsen A, Skaug HJ, Maechler M, Bolker BM. 2017. {glmmTMB} Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal* 9:378–400.

Brownfield L, Yi J, Jiang H, Minina EA, Twell D, Köhler C. 2015. Organelles maintain spindle position in plant meiosis. *Nature Communications* 6:1–9. doi:10.1038/ncomms7492

Cabout S, Leask MP, Varghese S, Yi J, Peters B, Conze LL, Köhler C, Brownfield L. 2017. The meiotic regulator JASON utilizes alternative translation initiation sites to produce differentially localized forms. *Journal of Experimental Botany* 68:4205–4217. doi:10.1093/jxb/erx222

Calvo-Baltanás V, De Jaeger-Braet J, Cher WY, Schönbeck N, Chae E, Schnittger A, Wijnker E. 2022. Knock-down of gene expression throughout meiosis and pollen formation by virus-induced gene silencing in Arabidopsis thaliana. *The Plant Journal* 111:19–37. doi:10.1111/tpj.15733

Calvo-Baltanás V, Wijnen CL, Yang C, Lukhovitskaya N, Snoo CB, Hohenwarter L, Keurentjes JJB, Jong H, Schnittger A, Wijnker E. 2020. Meiotic crossover reduction by virus-induced gene silencing enables the efficient generation of chromosome substitution lines and reverse breeding in *Arabidopsis thaliana*. *The Plant Journal* 104:1437–1452. doi:10.1111/tpj.14990

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:1–9. doi:10.1186/1471-2105-10-421

Cao X, Liu Xiaoyan, Wang Xiaotian, Yang M, van Giang T, Wang J, Liu Xiaolin, Sun S, Wei K, Wang Xiaoxuan, Gao J, Du Y, Qin Y, Guo Y, Huang Z. 2019. B-class MADS-box TM6 is a candidate gene for tomato male sterile-15 26.

Theoretical and Applied Genetics 132:2125–2135. doi:10.1007/s00122-019-03342-z

Capo A, Cammareri M, Rocca FD, Errico A, Zoina A, Conicella C. 2002. Evaluation for chipping and tuber soft rot (Erwinia carotovora) resistance in potato clones from unilateral sexual polyploidization (2x × 4x). *Am J Pot Res* 79:139–145. doi:10.1007/BF02881523

Celis C, Scurrah M, Cowgill S, Chumbiauca S, Green J, Franco J, Main G, Klezebrink D, Visser RGF, Atkinson HJ. 2004. Environmental biosafety and transgenic potato in a centre of diversity for this crop. *Nature* 432:222–225. doi:10.1038/nature03048

Chakravarti A. 1991. A graphical representation of genetic and physical maps: The Marey map. *Genomics* 11:219–222. doi:10.1016/0888-7543(91)90123-V

Charlesworth D, Willis JH. 2009. The genetics of inbreeding depression. *Nat Rev Genet* 10:783–796. doi:10.1038/nrg2664

Chase SS. 1963. Analytic breeding in solanum tuberosum l. – a scheme utilizing parthenotes and other diploid stocks. *Canadian Journal of Genetics and Cytology* 5:359–363. doi:10.1139/g63-049

Chelysheva L, Gendrot G, Vezon D, Doutriaux M-P, Mercier R, Grelon M. 2007. Zip4/Spo22 Is Required for Class I CO Formation but Not for Synapsis Completion in Arabidopsis thaliana. *PLOS Genetics* 3:e83. doi:10.1371/journal.pgen.0030083

Chelysheva L, Vezon D, Chambon A, Gendrot G, Pereira L, Lemhemdi A, Vrielynck N, Guin SL, Novatchkova M, Grelon M. 2012. The Arabidopsis HEI10 Is a New ZMM Protein Related to Zip3. *PLOS Genetics* 8:e1002799. doi:10.1371/journal.pgen.1002799

Chen C, Zhang W, Timofejeva L, Gerardin Y, Ma H. 2005. The Arabidopsis ROCK-N-ROLLERS gene encodes a homolog of the yeast ATP-dependent DNA helicase MER3 and is required for normal meiotic crossover formation. *The Plant Journal* 43:321–334. doi:10.1111/j.1365-313X.2005.02461.x

Chen M, Fan W, Ji F, Hua H, Liu J, Yan M, Ma Q, Fan J, Wang Q, Zhang S, Liu G, Sun Z, Tian C, Zhao F, Zheng J, Zhang Q, Chen J, Qiu J, Wei X, Chen Z, Zhang P, Pei D, Yang J, Huang X. 2021. Genome-wide identification of agronomically important genes in outcrossing crops using OutcrossSeq. *Molecular Plant* 14:556–570. doi:10.1016/j.molp.2021.01.003

Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34:i884–i890. doi:10.1093/bioinformatics/bty560

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* 6:80–92. doi:10.4161/fly.19695

Cipar MS, Peloquin SJ, Hougas RW. 1964. Haploidy and the Identification of Self-Incompatibility Alleles in Cultivated Potato Groups. *Canadian Journal of Genetics and Cytology* 9:511–518. doi:10.1139/g67-055

Clegg MT, Allard RW, Kahler AL. 1972. Is the gene the unit of selection? Evidence from two experimental plant populations. *Proceedings of the National Academy of Sciences of the United States of America* 69:2474–2478. doi:10.1073/pnas.69.9.2474

Clot CR, Klein D, Koopman J, Schuit C, Engelen CJM, Hutten RCB, Brouwer M, Visser RGF, Juranić M, Eck HJ van. 2023a. Desynapsis in potato is caused by StMSH4 mutant alleles and leads to either highly uniform unreduced pollen or sterility. doi:10.1101/2023.02.23.529759

Clot CR, Polzer C, Prodhomme C, Schuit C, Engelen CJM, Hutten RCB, van Eck HJ. 2020. The origin and widespread occurrence of Sli-based self-compatibility in potato. *Theoretical and Applied Genetics*. doi:10.1007/s00122-020-03627-8

Clot CR, Wang X, Koopman J, Navarro AT, Bucher J, Visser RGF, Finkers R, van Eck HJ. 2023b. High-Density Linkage Map Constructed from a Skim Sequenced Diploid Potato Population Reveals Transmission Distortion and QTLs for Tuber Yield and Pollen Shed. *Potato Res*. doi:10.1007/s11540-023-09627-7

Conner JA, Ozias-Akins P. 2017. Apomixis: Engineering the ability to harness hybrid vigor in crop plantsMethods in Molecular Biology. Humana Press Inc. pp. 17–34. doi:10.1007/978-1-4939-7286-9_2

Conover WJ, Iman RL. 1979. On multiple-comparisons procedures. *Los Alamos Sci Lab Tech Rep LA-7677-MS* 1:14.

Corneillie S, De Storme N, Van Acker R, Fangel JU, De Bruyne M, De Rycke R, Geelen D, Willats WGT, Vanholme B, Boerjan W. 2019. Polyploidy Affects Plant Growth and Alters Cell Wall Composition. *Plant Physiology* 179:74–87. doi:10.1104/pp.18.00967

Cronn R, Liston A, Parks M, Gernandt DS, Shen R, Mockler T. 2008. Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. *Nucleic Acids Research* 36:122. doi:10.1093/nar/gkn502

d'Erfurth I, Cromer L, Jolivet S, Girard C, Horlow C, Sun Y, To JPC, Berchowitz LE, Copenhaver GP, Mercier R. 2010. The CYCLIN-A CYCA1;2/TAM Is Required for the Meiosis I to Meiosis II Transition and Cooperates with OSD1 for the Prophase to First Meiotic Division Transition. *PLoS Genetics* 6:e1000989. doi:10.1371/journal.pgen.1000989

d'Erfurth I, Jolivet S, Froger N, Catrice O, Novatchkova M, Mercier R. 2009. Turning Meiosis into Mitosis. *PLoS Biology* 7:e1000124. doi:10.1371/journal.pbio.1000124

d'Erfurth I, Froger N, Jolivet S, Novatchkova M, Mercier R, Catrice O, Simon M, Jenczewski E. 2008. Mutations in AtPS1 (Arabidopsis thaliana Parallel Spindle 1) Lead to the Production of Diploid Pollen Grains. *PLoS Genetics* 4:e1000274. doi:10.1371/journal.pgen.1000274

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R. 2011. The variant call format and VCFtools. *Bioinformatics* 27:2156–2158. doi:10.1093/bioinformatics/btr330

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. 2021. Twelve years of SAMtools and BCFtools. *GigaScience* 10:1–4. doi:10.1093/gigascience/giab008

de Jong H, Rowe PR. 1971. Inbreeding in cultivated diploid potatoes. *Potato Research* 14:74–83. doi:10.1007/BF02355931

de Martino G, Pan I, Emmanuel E, Levy A, Irish VF. 2006. Functional Analyses of Two Tomato APETALA3 Genes Demonstrate Diversification in Their Roles in Regulating Floral Development. *The Plant Cell* 18:1833–1845. doi:10.1105/tpc.106.042978

de Massy B. 2013. Initiation of Meiotic Recombination: How and Where? Conservation and Specificities Among Eukaryotes. *Annual Review of Genetics* 47:563–599. doi:10.1146/annurev-genet-110711-155423

de Storme N, Geelen D. 2014. The impact of environmental stress on male reproductive development in plants: Biological processes and molecular mechanisms. *Plant, Cell and Environment* 37:1–18. doi:10.1111/pce.12142

de Storme N, Geelen D. 2013. Sexual polyploidization in plants - cytological mechanisms and molecular regulation. *New Phytologist* 198:670–684. doi:10.1111/nph.12184

de Storme N, Geelen D. 2011. The arabidopsis mutant jason produces unreduced first division restitution male gametes through a parallel/ fused spindle mechanism in meiosis II. *Plant Physiology* 155:1403–1415. doi:10.1104/pp.110.170415

de Vries ME, Adams JR, Eggers E, Ying S, Stockem JE, Kacheyo OC, van Dijk LCM, Khera P, Bachem CW, Lindhout P, van der Vossen EAG. 2023. Converting Hybrid Potato Breeding Science into Practice. *Plants* 12:230. doi:10.3390/plants12020230

Dempewolf H, Hodgins KA, Rummell SE, Ellstrand NC, Rieseberg LH. 2012. Reproductive Isolation during Domestication. *The Plant Cell* 24:2710–2717. doi:10.1105/tpc.112.100115

Deyhle F, Sarkar AK, Tucker EJ, Laux T. 2007. WUSCHEL regulates cell differentiation during anther development. *Developmental Biology* 302:154–159. doi:10.1016/j.ydbio.2006.09.013

Dion É, Li L, Jean M, Belzile F. 2007. An Arabidopsis MLH1 mutant exhibits reproductive defects and reveals a dual role for this gene in mitotic recombination. *The Plant Journal* 51:431–440. doi:10.1111/j.1365-313X.2007.03145.x

Dirks R, van Dun K, de Snoo CB, van den Berg M, Lelivelt CLC, Voermans W, Woudenberg L, de Wit JPC, Reinink K, Schut JW, van der Zeeuw E, Vogelaar A, Freymark G, Gutteling EW, Keppel MN, van Drongelen P, Kieny M, Ellul P, Touraev A, Ma H, de Jong H, Wijnker E. 2009. Reverse breeding: a novel breeding approach based on engineered meiosis. *Plant Biotechnology Journal* 7:837–845. doi:10.1111/j.1467-7652.2009.00450.x

Dix DJ, Allen JW, Collins BW, Poorman-Allen P, Mori C, Blizard DR, Brown PR, Goulding EH, Strong BD, Eddy EM. 1997. HSP70-2 is required for desynapsis of synaptonemal complexes during meiotic prophase in juvenile and adult mouse spermatocytes. *Development* 124:4595–4603. doi:10.1242/dev.124.22.4595

Dobzhansky TH. 1936. Studies on hybrid sterility. II. Localization of sterility factors in Drosophila pseudoobscura hybrids. *Genetics* 21:113.

Dongyu Q, Dewei Z, Ramanna MS, Jacobsen E. 1995. A comparison of progeny from diallel crosses of diploid potato with regard to the frequencies of 2n-pollen grains. *Euphytica: International Journal of Plant Breeding* 92:313–320. doi:10.1007/BF00037114

Douches DS, Maas D, Jastrzebski K, Chase RW. 1996. Assessment of potato breeding in the USA over the past century. *Crop Science* 36:1544–1552. doi:10.2135/cropsci1996.0011183X003600060024x

Douches DS, Quiros CF. 1988a. Genetic recombination in a diploid synaptic mutant and a solanum tuberosum × S. Chacoense diploid hybrid. *Heredity* 60:183–191. doi:10.1038/hdy.1988.30

Douches DS, Quiros CF. 1988b. Genetic strategies to determine the mode of 2n egg formation in diploid potatoes. *Euphytica* 38:247–260. doi:10.1007/BF00023528

Draaistra J. 2006. Genetic analyses of root-knot nematode resistance in potato.

Dukić M, Bomblies K. 2022. Male and female recombination landscapes of diploid *Arabidopsis arenosa*. *Genetics* 220. doi:10.1093/genetics/iyab236

Duvick DN. 2005. The Contribution of Breeding to Yield Advances in maize (Zea mays L.). *Advances in Agronomy* 86:83–145. doi:10.1016/S0065-2113(05)86002-X

Dwivedi SL, Heslop-Harrison P, Spillane C, McKeown PC, Edwards D, Goldman I, Ortiz R. 2023. Evolutionary dynamics and adaptive benefits of deleterious mutations in crop gene pools. *Trends in Plant Science* 28:685–697. doi:10.1016/j.tplants.2023.01.006

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797. doi:10.1093/nar/gkh340

Eggers EJ, van der Burgt A, van Heusden SAW, de Vries ME, Visser RGF, Bachem CWB, Lindhout P. 2021. Neofunctionalisation of the Sli gene leads to self-compatibility and facilitates precision breeding in potato. *Nature Communications* 12:1–9. doi:10.1038/s41467-021-24267-6

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS one* 6:e19379. doi:10.1371/journal.pone.0019379

Enciso-Rodriguez F, Manrique-Carpintero NC, Nadakuduti SS, Buell CR, Zarka D, Douches D. 2019. Overcoming Self-Incompatibility in Diploid Potato Using CRISPR-Cas9. *Frontiers in Plant Science* 10:376. doi:10.3389/fpls.2019.00376

Endelman J, Jansky SH, Butler N, Christensen G. 2019. Genetic Evidence of a Recessive Lethal Allele on Potato Chromosome 12 - Annual Report of the Potato Association of America. *American Journal of Potato Research* 96:331. doi:10.1007/s12230-019-09723-w

Endelman JB, Jansky SH. 2016. Genetic mapping with an inbred line-derived F2 population in potato. *Theoretical and Applied Genetics* 129:935–943. doi:10.1007/s00122-016-2673-7

Erilova A, Brownfield L, Exner V, Rosa M, Twell D, Scheid OM, Hennig L, Köhler C. 2009. Imprinting of the Polycomb Group Gene MEDEA Serves as a Ploidy Sensor in Arabidopsis. *PLoS Genetics* 5:e1000663. doi:10.1371/journal.pgen.1000663

Farré A, Benito IL, Cistué L, de Jong JH, Romagosa I, Jansen J. 2011. Linkage map construction involving a reciprocal translocation. *Theoretical and Applied Genetics* 122:1029–1037. doi:10.1007/s00122-010-1507-2

Fisher RA. 1918. The correlation between relatives on the supposition of Mendelian inheritance. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh* 52:399–433.

Gallego ME, Jeanneau M, Granier F, Bouchez D, Bechtold N, Charles, White I. 2001. Disruption of the Arabidopsis RAD50 gene leads to plant sterility and MMS sensitivity. *The Plant Journal* 25:31–41. doi:10.1111/j.1365-313X.2001.00928.x

Gao SM, Yang MH, Zhang F, Fan LJ, Zhou Y. 2019. The strong competitive role of 2n pollen in several polyploidy hybridizations in Rosa hybrida. *BMC Plant Biology* 19:127. doi:10.1186/s12870-019-1696-z

Gardiner L-J, Wingen LU, Bailey P, Joynson R, Brabbs T, Wright J, Higgins JD, Hall N, Griffiths S, Clavijo BJ, Hall A. 2019. Analysis of the recombination landscape of hexaploid bread wheat reveals genes controlling recombination and gene conversion frequency. *Genome Biol* 20:1–16. doi:10.1186/s13059-019-1675-6

Gardner KM, Douglass K, De Jong H, De Koeyer D, Tai HH. 2019. Genetic Mapping of Self-Compatibility in Diploid Potato Using Genotyping By SequencingPAG Conference XXVII. San Diego, CA. p. PE0940.

Gebhardt C, Ritter E, Barone A, Debener T, Walkemeier B, Schachtschabel U, Kaufmann H, Thompson RD, Bonierbale MW, Ganal MW, Tanksley SD, Salamini F. 1991. RFLP maps of potato and their alignment with the

homoeologous tomato genome. *Theoretical and Applied Genetics* 83:49–57. doi:10.1007/BF00229225

Gildemacher PR, Demo P, Barker I, Kaguongo W, Woldegiorgis G, Wagoire WW, Wakahiu M, Leeuwis C, Struik PC. 2009. A description of seed potato systems in Kenya, Uganda and Ethiopia. *American journal of potato research* 86:373–382.

Golmirzaie AM, Ortiz R, Atlin GN, Iwanaga M. 1998. Inbreeding and true seed in tetrasomic potato. I. Selfing and open pollination in Andean landraces (Solanum tuberosum Gp. Andigena). *Theor Appl Genet* 97:1125–1128. doi:10.1007/s001220051000

Gómez JF, Talle B, Wilson ZA. 2015. Anther and pollen development: A conserved developmental pathway. *Journal of Integrative Plant Biology* 57:876–891. doi:10.1111/jipb.12425

Gonzalo A, Lucas MO, Charpentier C, Sandmann G, Lloyd A, Jenczewski E. 2019. Reducing MSH4 copy number prevents meiotic crossovers between non-homologous chromosomes in Brassica napus. *Nature Communications* 10:1–9. doi:10.1038/s41467-019-10010-9

Goodrich CE. 1863. The origination and test culture of seedling potatoes. *Trans NYS Agric Soc* 23:89–134.

Gopal J. 2015. Challenges and Way-forward in Selection of Superior Parents, Crosses and Clones in Potato Breeding. *Potato Research* 58:165–188. doi:10.1007/s11540-015-9292-6

Grelon M, Vezon D, Gendrot G, Pelletier G. 2001. AtSPO11-1 is necessary for efficient meiotic recombination in plants. *EMBO Journal* 20:589–600. doi:10.1093/emboj/20.3.589

Guan Y, Leu NA, Ma J, Chmátal L, Ruthel G, Bloom JC, Lampson MA, Schimenti JC, Luo M, Wang PJ. 2020. SKP1 drives the prophase I to metaphase I transition during male meiosis. *Sci Adv* 6:eaaz2129. doi:10.1126/sciadv.aaz2129

Gutaker RM, Weiß CL, Ellis D, Anglin NL, Knapp S, Luis Fernández-Alonso J, Prat S, Burbano HA. 2019. The origins and adaptation of European potatoes reconstructed from historical genomes. *Nat Ecol Evol* 3:1093–1101. doi:10.1038/s41559-019-0921-3

Hammer Ø, Harper DAT, Ryan PD education and data analysis. PE 4(1): 9pp. http://palaeo-electronica.org/2001_1/past/issue1_01.htm. 2001. PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica* 4:9.

Hardigan MA, Laimbeer FPE, Newton L, Crisovan E, Hamilton JP, Vaillancourt B, Wiegert-Rininger K, Wood JC, Douches DS, Farré EM, Veilleux RE, Buell CR. 2017. Genome diversity of tuber-bearing Solanum uncovers complex evolutionary history and targets of domestication in the cultivated potato. *Proceedings of the National Academy of Sciences of the United States of America* 114:E9999–E10008. doi:10.1073/pnas.1714380114

Harlan JR, De Wet JMJ. 1975. On Ö. Winge and a Prayer: The origins of polyploidy. *The Botanical Review* 41:361–390. doi:10.1007/BF02860830

Hawkes JG. 1979. Genetic poverty of the potato in Europe. *Broadening the genetic base of crops Wageningen: Centre for Agricultural Publishing and Documentation* 19–27.

Hawkes JG, Francisco-Ortega J. 1993. The early history of the potato in Europe. *Euphytica* 70:1–7. doi:10.1007/BF00029633

Hermsen JGT. 1980. Breeding for apomixis in potato: Pursuing a utopian scheme. *Euphytica* 29:595–607. doi:10.1007/BF00023207

Hermsen JGT. 1978. Genetics of self-compatibility in dihaploids of Solanum tuberosum L. 3. Lethality of S-bearing translocation homozygotes. *Euphytica* 27:13–17.

Hermsen JGT, Ramanna MS, Jongedijk E. 1985. Apomictic approach to introduce uniformity and vigour into progenies from true potato seed (TPS)Present and Future Strategies for Potato Breeding and Improvement. Report 26th Planning Conf. CIP, Lima, Peru. pp. 99–114.

Hermsen JGT, Verdenius J. 1973. Selection from Solanum tuberosum group phureja of genotypes combining high-frequency haploid induction with homozygosity for embryo-spot. *Euphytica* 22:244–259. doi:10.1007/BF00022632

Higgins JD, Armstrong SJ, Franklin FCH, Jones GH. 2004. The Arabidopsis MutS homolog AtMSH4 functions at an early step in recombination: Evidence for two classes of recombination in Arabidopsis. *Genes and Development* 18:2557–2570. doi:10.1101/gad.317504

Higgins JD, Sanchez-Moran E, Armstrong SJ, Jones GH, Franklin FCH. 2005. The Arabidopsis synaptonemal complex protein ZYP1 is required for chromosome synapsis and normal fidelity of crossing over. *Genes Dev* 19:2488–2500. doi:10.1101/gad.354705

Higgins JD, Vignard J, Mercier R, Pugh AG, Franklin FCH, Jones GH. 2008. AtMSH5 partners AtMSH4 in the class I meiotic crossover pathway in Arabidopsis thaliana, but is not required for synapsis. *The Plant Journal* 55:28–39. doi:10.1111/j.1365-313X.2008.03470.x

Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. *Genetics Research* 8:269–294. doi:10.1017/S0016672300010156

Holm S. 1979. A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics* 6:65–70.

Hoopes G, Meng X, Hamilton JP, Achakkagari SR, de Alves Freitas Guesdes F, Bolger ME, Coombs JJ, Esselink D, Kaiser NR, Kodde L, Kyriakidou M, Lavrijssen B, van Lieshout N, Shereda R, Tuttle HK, Vaillancourt B, Wood JC, de Boer JM, Bornowski N, Bourke P, Douches D, van Eck HJ, Ellis D, Feldman MJ, Gardner KM, Hopman JCP, Jiang J, De Jong WS, Kuhl JC, Novy RG, Oome S, Sathuvalli V, Tan EH, Ursum RA, Vales MI, Vining K, Visser RGF, Vossen J, Yencho GC, Anglin NL, Bachem CWB, Endelman JB, Shannon LM, Strömvik MV, Tai HH, Usadel B, Buell CR, Finkers R. 2022. Phased, chromosome-scale genome assemblies of tetraploid potato reveal a complex genome, transcriptome, and predicted proteome landscape underpinning genetic diversity. *Molecular Plant* 15:520–536. doi:10.1016/j.molp.2022.01.003

Hosaka K, Hanneman RE. 1998a. Genetics of self-compatibility in a self-incompatible wild diploid potato species Solanum chacoense. 1. Detection of an S locus inhibitor (Sli) gene. *Euphytica* 99:191–197. doi:10.1023/A:1018353613431

Hosaka K, Hanneman RE. 1998b. Genetics of self-compatibility in a self-incompatible wild diploid potato species Solanum chacoense. 2. Localization of an S locus inhibitor (Sli) gene on the potato genome using DNA markers. *Euphytica* 103:265–271. doi:10.1023/A:1018380725160

Hua ZH, Fields A, Kao TH. 2008. Biochemical models for S-RNase-based self-incompatibility. *Molecular Plant* 1:575–585. doi:10.1093/mp/ssn032

Hunter JD. 2007. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* 9:90–95. doi:10.1109/MCSE.2007.55

Husband BC. 2016. Effect of inbreeding on pollen tube growth in diploid and tetraploid Chamerion angustifolium: Do polyploids mask mutational load in pollen? *American Journal of Botany* 103:532–540. doi:10.3732/ajb.1500243

Hutten RCB. 1994. Basic aspects of potato breeding via the diploid level (PhD thesis).

Hutten RCB, Schippers MGM, Hermsen JGTh, Ramanna MS. 1994a. Comparative performance of FDR and SDR progenies from reciprocal 4x-2x crosses in potato. *Theoret Appl Genetics* 89:545–550. doi:10.1007/BF00222446

Hutten RCB, Scholberg EJMM, Huigen DJ, Hermsen JGT, Jacobsen E. 1994b. Analysis of dihaploid induction and production ability and seed parent x pollinator interaction in potato. *Euphytica* 72:61–64. doi:10.1007/BF00023773

Hutten RCB, Soppe WJJ, Hermsen JGT, Jacobsen E. 1995. Evaluation of dihaploid populations from potato varieties and breeding lines. *Potato Research* 38:77–86. doi:10.1007/BF02358072

Ikeda M, Mitsuda N, Ohme-Takagi M. 2009. *Arabidopsis* WUSCHEL Is a Bifunctional Transcription Factor That Acts as a Repressor in Stem Cell Regulation and as an Activator in Floral Patterning. *The Plant Cell* 21:3493–3505. doi:10.1105/tpc.109.069997

Iwanaga M, Peloquin SJ. 1982. Origin and evolution of cultivated tetraploid potatoes via 2n gametes. *Theoretical and Applied Genetics* 61:161–169. doi:10.1007/BF00273885

Jackson N, Sanchez-Moran E, Buckling E, Armstrong SJ, Jones GH, Franklin FCH. 2006. Reduced meiotic crossovers and delayed prophase I progression in AtMLH3-deficient Arabidopsis. *The EMBO Journal* 25:1315–1323. doi:10.1038/sj.emboj.7600992

Jacobs J. M. E., Van Eck HJ, Arens P, Verkerk-Bakker B, te Lintel Hekkert B, Bastiaanssen HJM, El-Kharbotly A, Pereira A, Jacobsen E, Stiekema WJ. 1995. A genetic map of potato (Solanum tuberosum) integrating molecular markers, including transposons, and classical markers. *Theoretical and Applied Genetics* 91:289–300. doi:10.1007/BF00220891

Jacobs J. M.E., Van Eck HJ, Arens P, Verkerk-Bakker B, te Lintel Hekkert B, Bastiaanssen HJM, El-Kharbotly A, Pereira A, Jacobsen E, Stiekema WJ. 1995. A genetic map of potato (Solanum tuberosum) integrating molecular markers, including transposons, and classical markers. *Theoretical and Applied Genetics* 91:289–300. doi:10.1007/BF00220891

Jansky SH, Charkowski AO, Douches DS, Gusmini G, Richael C, Bethke PC, Spooner DM, Novy RG, De Jong H, De Jong WS, Bamberg JB, Thompson AL, Bizimungu B, Holm DG, Brown CR, Haynes KG, Sathuvalli VR, Veilleux RE, Creighton Miller J, Bradeen JM, Jiang J. 2016. Reinventing potato as a diploid inbred line-based crop. *Crop Science* 56:1412–1422. doi:10.2135/cropsci2015.12.0740

Jansky SH, Chung YS, Kittipadukal P. 2014. M6: A Diploid Potato Inbred Line for Use in Breeding and Genetics Research. *Journal of Plant Registrations* 8:195. doi:10.3198/jpr2013.05.0024crg

Jansky SH, Spooner DM. 2018. The Evolution of Potato Breeding In: Goldman I, editor. Plant Breeding Reviews. John Wiley & Sons. pp. 169–214.

Johnson MA, Harper JF, Palanivelu R. 2019. A Fruitful Journey: Pollen Tube Navigation from Germination to Fertilization. *Annual Review of Plant Biology* 70:809–837. doi:10.1146/annurev-arplant-050718-100133

Jones G.H., Heslop-Harrison J.S. 1996. Classical and Molecular Cytogenetics of Arabidopsis thaliana. Oxford University Press.

Jongedijk E. 1991. Desynapsis and FDR 2N-megaspore formation in diploid potato : potentials and limitations for breeding and for the induction of diplosporic apomixis.

Jongedijk E, Hutten RCB, Wolk JMASA van der, Stekhoven SIJS. 1991a. Synaptic mutants in potato, Solanum tuberosum L. III. Effect of the Ds-1/ds-1 locus (desynapsis) on genetic recombination in male and female meiosis. *Genome* 34:121–130. doi:10.1139/g91-019

Jongedijk E, Ramanna MS. 1989. Synaptic mutants in potato, Solanum tuberosum L. II. Concurrent reduction of chiasma frequencies in male and female meiosis of ds-1 (desynapsis) mutants. *Genome* 32:1054–1062. doi:10.1139/g89-553

Jongedijk E, Ramanna MS. 1988. Synaptic mutants in potato, Solanum tuberosum L. I. Expression and identity of genes for desynapsis. *Genome* 30:664–670. doi:10.1139/g88-112

Jongedijk E, Ramanna MS, Sawor Z, Hermsen JGT. 1991b. Formation of first division restitution (FDR) 2n-megaspores through pseudohomotypic division in ds-1 (desynapsis) mutants of diploid potato: routine production of tetraploid progeny from 2xFDR × 2xFDR crosses. *Theoret Appl Genetics* 82:645–656. doi:10.1007/BF00226804

Jongedijk E, van der Wolk JMASA, Suurs LCJM. 1990. Analysis of glutamate oxaloacetate transaminase (GOT) isozyme variants in diploid tuberous Solanum; inheritance and linkage relationships to ds1 (desynapsis), y (tuber flesh colour), cr (crumpled) and yc (yellow cotyledon). *Euphytica* 45:155–167. doi:10.1007/BF00033282

Kaiser NR, Jansky S, Coombs JJ, Collins P, Alsahlany M, Douches DS. 2021. Assessing the Contribution of Sli to Self-Compatibility in North American Diploid Potato Germplasm Using KASP™ Markers. *American Journal of Potato Research* 1–10. doi:10.1007/s12230-021-09821-8

Kang YJ, Lee T, Lee J, Shim S, Jeong H, Satyawan D, Kim MY, Lee S-H. 2016. Translational genomics for plant breeding with the genome sequence explosion. *Plant Biotechnology Journal* 14:1057–1069. doi:10.1111/pbi.12449

Kao T, Tsukamoto T. 2004. The Molecular and Genetic Bases of S-RNase-Based. *The Plant Cell* 16:72–83. doi:10.1105/tpc.016154.S-RNase-Based

Kaplinski L, Lepamets M, Remm M. 2015. GenomeTester4: a toolkit for performing basic set operations - union, intersection and complement on k-mer lists. *GigaScience* 4:58. doi:10.1186/s13742-015-0097-y

Kidane-Mariam H-M, Peloquin SJ. 1975. Method of Diplandroid Formation and Yield of Progeny from Reciprocal (4x-2x) Crosses1. *Journal of the American Society for Horticultural Science* 100:602–603. doi:10.21273/JASHS.100.6.602

Kloosterman B, Abelenda JA, Gomez MDMC, Oortwijn M, De Boer JM, Kowitwanich K, Horvath BM, Van Eck HJ, Smaczniak C, Prat S, Visser RGF, Bachem CWB. 2013. Naturally occurring allele diversity allows potato cultivation in northern latitudes. *Nature* 495:246–250. doi:10.1038/nature11912

Kokot M, Długosz M, Deorowicz S. 2017. KMC 3: counting and manipulating k-mer statistics. *Bioinformatics* 33:2759–2761. doi:10.1093/bioinformatics/btx304

Kondo K, Yamamoto M, Matton DP, Sato T, Hirai M, Norioka S, Hattori T, Kowyama Y. 2002. Cultivated tomato has defects in both S-RNase and HT genes required for stylar function of self-incompatibility. *The Plant Journal* 29:627–636. doi:10.1046/j.0960-7412.2001.01245.x

Koren S, Rhie A, Walenz BP, Dilthey AT, Bickhart DM, Kingan SB, Hiendleder S, Williams JL, Smith TPL, Phillippy AM. 2018. De novo assembly of haplotype-resolved genomes with trio binning. *Nat Biotechnol* 36:1174–1182. doi:10.1038/nbt.4277

Krantz FA. 1924. Potato breeding methods. *Technical Bulletin* 25.

Kreiner JM, Kron P, Husband BC. 2017a. Evolutionary Dynamics of Unreduced Gametes. *Trends in Genetics* 33:583–593. doi:10.1016/J.TIG.2017.06.009

Kreiner JM, Kron P, Husband BC. 2017b. Frequency and maintenance of unreduced gametes in natural plant populations: associations with reproductive mode, life history and genome size. *New Phytologist* 214:879–889. doi:10.1111/nph.14423

Kubo K, Entani T, Takara A, Wang N, Fields AM, Hua Z, Toyoda M, Kawashima S, Ando T, Isogai A, Kao T, Takayama S. 2010. Collaborative Non-Self Recognition System in S-RNase Based Self-Incompatibility. *Science* 330:796–799. doi:10.1126/science.1195243

Kubo K, Paape T, Hatakeyama M, Entani T, Takara A, Kajihara K, Tsukahara M, Shimizu-Inatsugi R, Shimizu KK, Takayama S. 2015. Gene duplication and genetic exchange drive the evolution of S-RNase-based self-incompatibility in Petunia. *Nature Plants* 1:14005. doi:10.1038/nplants.2014.5

Kumar A, Siddappa S, Bhardwaj V, Dalamu, Singh B, Sharma N, Dipta B, Kumar V, Goutam U, Sood S. 2023. Generation of Asynaptic Mutants in Potato by Disrupting StDMC1 Gene Using RNA Interference Approach. *Life* 13:174. doi:10.3390/life13010174

Kyriakidou M, Achakkagari SR, Gálvez López JH, Zhu X, Tang CY, Tai HH, Anglin NL, Ellis D, Strömvik MV. 2020. Structural genome analysis in cultivated potato taxa. *Theoretical and Applied Genetics* 133:951–966. doi:10.1007/s00122-019-03519-6

Lambing C, Franklin FCH, Wang C-JR. 2017. Understanding and Manipulating Meiotic Recombination in Plants. *Plant Physiology* 173:1530–1542. doi:10.1104/pp.16.01530

Larson G, Piperno DR, Allaby RG, Purugganan MD, Andersson L, Arroyo-Kalin M, Barton L, Climer Vigueira C, Denham T, Dobney K, Doust AN, Gepts P, Gilbert MTP, Gremillion KJ, Lucas L, Lukens L, Marshall FB, Olsen KM, Pires JC, Richerson PJ, Rubio de Casas R, Sanjur OI, Thomas MG, Fuller DQ. 2014. Current perspectives and the future of domestication studies. *Proceedings of the National Academy of Sciences* 111:6139–6146. doi:10.1073/pnas.1323964111

Lee H-S, Huang S, Kao T. 1994. S proteins control rejection of incompatible pollen in Petunia inflata. *Nature* 367:560–563. doi:10.1038/367560a0

Lei X, Liu B. 2020. Tapetum-Dependent Male Meiosis Progression in Plants: Increasing Evidence Emerges. *Frontiers in Plant Science* 10.

Leisner CP, Hamilton JP, Crisovan E, Manrique-Carpintero NC, Marand AP, Newton L, Pham GM, Jiang J, Douches DS, Jansky SH, Buell CR. 2018. Genome sequence of M6, a diploid inbred clone of the high-glycoalkaloid-producing tuber-bearing potato species Solanum chacoense, reveals residual heterozygosity. *Plant Journal* 94:562–570. doi:10.1111/tpj.13857

Leyva-Pérez M de la O, Vexler L, Byrne S, Clot CR, Meade F, Griffin D, Ruttink T, Kang J, Milbourne D. 2022. PotatoMASH—A Low Cost, Genome-Scanning Marker System for Use in Potato Genomics and Genetics Applications. *Agronomy* 12:2461. doi:10.3390/agronomy12102461

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. doi:10.1093/bioinformatics/btp324

Li Y, Li G, Li C, Qu D, Huang S. 2013. Prospects of diploid hybrid breeding in potato. *Chin Potato J* 27:96–99.

Li Y, Shen Y, Cai C, Zhong C, Zhu L, Yuan M, Ren H. 2010. The type II Arabidopsis formin14 interacts with microtubules and microfilaments to regulate cell division. *Plant Cell* 22:2710–2726. doi:10.1105/tpc.110.075507

Lian Q, Tang D, Bai Z, Qi J, Lu F, Huang S, Zhang C. 2019. Acquisition of deleterious mutations during potato polyploidization. *Journal of Integrative Plant Biology* 61:7–11. doi:10.1111/JIPB.12748

Lindhout P, de Vries M, ter Maat M, Ying S, Viquez-Zamora M, van Heusden S. 2016. Hybrid potato breeding for improved varieties. pp. 99–122. doi:10.19103/as.2016.0016.04

Lindhout P, Meijer D, Schotte T, Hutten RCB, Visser RGF, van Eck HJ. 2011. Towards F1 Hybrid Seed Potato Breeding. *Potato Research* 54:301–312. doi:10.1007/s11540-011-9196-z

Lindhout P, Struik PC. 2023. Chapter 2. Hybrid potato breeding and production systemsImpact of Hybrid Potato. Wageningen Academic Publishers. pp. 17–30. doi:10.3920/978-90-8686-946-6_2

Liu S, Li Z, Wu S, Wan X. 2021. The essential roles of sugar metabolism for pollen development and male fertility in plants. *The Crop Journal* 9:1223–1236. doi:10.1016/j.cj.2021.08.003

Love SL. 1999. Founding clones, major contributing ancestors, and exotic progenitors of prominent North American potato cultivars. *American Journal of Potato Research* 76:263–272. doi:10.1007/BF02853624

Lynn A, Soucek R, Börner GV. 2007. ZMM proteins during meiosis: Crossover artists at work. *Chromosome Research* 15:591–605. doi:10.1007/s10577-007-1150-1

Ma H. 2005. Molecular Genetic Analyses of Microsporogenesis and Microgametogenesis in Flowering Plants. *Annual Review of Plant Biology* 56:393–434. doi:10.1146/annurev.arplant.55.031903.141717

Ma L, Zhang C, Zhang B, Tang F, Li F, Liao Q, Tang D, Peng Z, Jia Y, Gao M, Guo H, Zhang J, Luo X, Yang H, Gao D, Lucas WJ, Li C, Huang S, Shang Y. 2021. A nonS-locus F-box gene breaks self-incompatibility in diploid potatoes. *Nature Communications* 12:1–8. doi:10.1038/s41467-021-24266-7

Macaisne N, Novatchkova M, Peirera L, Vezon D, Jolivet S, Froger N, Chelysheva L, Grelon M, Mercier R. 2008. SHOC1, an XPF Endonuclease-Related Protein, Is Essential for the Formation of Class I Meiotic Crossovers. *Current Biology* 18:1432–1437. doi:10.1016/j.cub.2008.08.041

Mackay IJ, Cockram J, Howell P, Powell W. 2021. Understanding the classics: the unifying concepts of transgressive segregation, inbreeding depression and heterosis and their central relevance for crop breeding. *Plant Biotechnology Journal* 19:26–34. doi:10.1111/pbi.13481

Manrique-Carpintero NC, Coombs JJ, Cui Y, Veilleux RE, Buell CR, Douches D. 2015. Genetic Map and QTL Analysis of Agronomic Traits in a Diploid Potato Population using Single Nucleotide Polymorphism Markers. *Crop Science* 55:2566–2579. doi:10.2135/cropsci2014.10.0745

Manrique-Carpintero NC, Coombs JJ, Pham GM, Laimbeer FPE, Braz GT, Jiang J, Veilleux RE, Buell CR, Douches DS. 2018. Genome Reduction in Tetraploid Potato Reveals Genetic Load, Haplotype Variation, and Loci Associated With Agronomic Traits. *Frontiers in Plant Science* 9.

Manrique-Carpintero NC, Coombs JJ, Veilleux RE, Buell CR, Douches DS. 2016. Comparative analysis of regions with distorted segregation in three diploid populations of potato. *G3: Genes, Genomes, Genetics* 6:2617–2628. doi:10.1534/g3.116.030031

Marand AP, Jansky SH, Gage JL, Hamernik AJ, de Leon N, Jiang J. 2019. Residual Heterozygosity and Epistatic Interactions Underlie the Complex Genetic Architecture of Yield in Diploid Potato. *Genetics* 212:317–332. doi:10.1534/genetics.119.302036

Mayer U, Herzog U, Berger F, Inzé D, Jürgens G. 1999. Mutations in the PILZ group genes disrupt the microtubule cytoskeleton and uncouple cell cycle progression from cell division in Arabidopsis embryo and endosperm. *European Journal of Cell Biology* 78:100–108. doi:10.1016/S0171-9335(99)80011-9

McClure B, Cruz-García F, Romero C. 2011. Compatibility and incompatibility in S-RNase-based systems. *Annals of Botany* 108:647–658. doi:10.1093/aob/mcr179

Mendel G. 1866. Versuche über pflanzen-hybriden. Verhandlungen des naturforschenden Vereines in Brünn 4: 3-47Classic Papers in Genetics. Prentice-Hall Englewood Cliffs, New Jersey.

Mendiburu AO, Peloquin SJ. 1977. The significance of 2N gametes in potato breeding. *Theoret Appl Genetics* 49:53–61. doi:10.1007/BF00275164

Mendiburu AO, Peloquin SJ. 1971. High yielding tetraploids from 4x-2x and 2x-2x matings. *Amer Potato J* 48:300–301.

Mendoza H. 1985. Preliminary results on yield and uniformity of potoes grown from true seedCIP Report of a Planning Conference on the Production of Potatoes from True Seed. Manila, Philippines. pp. 156–172.

Mendoza HA, Haynes FL. 1974. Genetic basis of heterosis for yield in the autotetraploid potato. *Theoret Appl Genetics* 45:21–25. doi:10.1007/BF00281169

Mercier R, Grelon M. 2008. Meiosis in plants: Ten years of gene discovery. *Cytogenetic and Genome Research* 120:281–290. doi:10.1159/000121077

Mercier R, Jolivet S, Vezon D, Huppe E, Chelysheva L, Giovanni M, Nogué F, Doutriaux M-P, Horlow C, Grelon M, Mézard C. 2005. Two Meiotic Crossover Classes Cohabit in Arabidopsis: One Is Dependent on MER3,whereas the Other One Is Not. *Current Biology* 15:692–701. doi:10.1016/j.cub.2005.02.056

Mercier R, Mézard C, Jenczewski E, Macaisne N, Grelon M. 2015. The Molecular Biology of Meiosis in Plants. *Annual Review of Plant Biology* 66:297–327. doi:10.1146/annurev-arplant-050213-035923

Michelmore RW, Paran I, Kesseli RV. 1991. Michelmore et al 1991. Proc Natl Acad Sci. *Proceedings of the National Academy of Sciences of the United States of America* 88:9828–9832. doi:10.1073/pnas.88.21.9828

Mieulet D, Jolivet S, Rivard M, Cromer L, Vernet A, Mayonove P, Pereira L, Droc G, Courtois B, Guiderdoni E, Mercier R. 2016. Turning rice meiosis into mitosis. *Cell Research* 26:1242–1254. doi:10.1038/cr.2016.117

Mirzaghaderi G, Hörandl E. 2016. The evolution of meiotic sex and its alternatives. *Proceedings of the Royal Society B: Biological Sciences* 283. doi:10.1098/rspb.2016.1221

Mok DWS, Peloquin SJ. 1975a. Three mechanisms of diplandroid (2n pollen) formation in diploid potatoes. *Canadian Journal of Genetics and Cytology* 17:217–225. doi:10.1139/g75-029

Mok DWS, Peloquin SJ. 1975b. The inheritance of three mechanisms of diplandroid (2n pollen) formation in diploid potatoes. *Heredity* 35:295–302. doi:10.1038/hdy.1975.100

Morgan C, White MA, Franklin FCH, Zickler D, Kleckner N, Bomblies K. 2021. Evolution of crossover interference enables stable autopolyploidy by ensuring pairwise partner connections in Arabidopsis arenosa. *Current Biology* 31:4713-4726.e4. doi:10.1016/j.cub.2021.08.028

Muller H. 1942. Isolating mechanisms, evolution, and temperatureBiol. Symp. pp. 71–125.

Murfett J, Atherton TL, Mou B, Gassert CS, McClure BA. 1994. S-RNase expressed in transgenic Nicotiana causes S-allele-specific pollen rejection. *Nature* 367:563–566. doi:10.1038/367563a0

Neale MJ, Pan J, Keeney S. 2005. Endonucleolytic processing of covalent protein-linked DNA double-strand breaks. *Nature* 436:1053–1057. doi:10.1038/nature03872

Nettancourt D de. 1997. Incompatibility in angiosperms. *Sexual Plant Reproduction* 10:185–199.
Okonechnikov K, Conesa A, García-Alcalde F. 2015. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* 32:btv566. doi:10.1093/bioinformatics/btv566
Olsder J, Hermsen JGT. 1976. Genetics of self-compatibility in dihaploids of Solanum tuberosum L. I. Breeding behaviour of two self-compatible dihaploids. *Euphytica* 25:597–607. doi:10.1007/BF00041597
Olsson O, Paik C. 2016. Long-run cultural divergence: Evidence from the Neolithic Revolution. *Journal of Development Economics* 122:197–213. doi:10.1016/j.jdeveco.2016.05.003
Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM. 2016. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biology* 17:132. doi:10.1186/s13059-016-0997-x
Ortiz R, Iwanaga M, Peloquin S. 1994. Breeding potatoes for the developing countries using wild tuber bearing Solanum species and ploidy manipulations. *J Genet Breed* 89–98.
Ortiz R, Peloquin S. 1992. Recurrent selection for 2n gamete production in 2X potatoes. *Journal of genetics & breeding* 46:383–389.
Otto SP. 2007. The Evolutionary Consequences of Polyploidy. *Cell* 131:452–462. doi:10.1016/j.cell.2007.10.022
Otto SP, Whitton J. 2000. Polyploid incidence and evolution. *Annu Rev Genet* 34:401–437. doi:10.1146/annurev.genet.34.1.401
Page SL, Hawley RS. 2003. Chromosome Choreography: The Meiotic Ballet. *Science* 301:785–789. doi:10.1126/science.1086605
Palumbo F, Pasquali E, Albertini E, Barcaccia G. 2021. A Review of Unreduced Gametes and Neopolyploids in Alfalfa: How to Fill the Gap between Well-Established Meiotic Mutants and Next-Generation Genomic Resources. *Plants* 10:999. doi:10.3390/plants10050999
Pancaldi F, van Loo EN, Senio S, Al Hassan M, van der Cruijsen K, Paulo M-J, Dolstra O, Schranz ME, Trindade LM. 2023. Syntenic Cell Wall QTLs as Versatile Breeding Tools: Intraspecific Allelic Variability and Predictability of Biomass Quality Loci in Target Plant Species. *Plants* 12:779. doi:10.3390/plants12040779
Pancaldi F, Vlegels D, Rijken H, van Loo EN, Trindade LM. 2022. Detection and Analysis of Syntenic Quantitative Trait Loci Controlling Cell Wall Quality in Angiosperms. *Frontiers in Plant Science* 13.
Park T-H, Kim J-B, Hutten RCB, van Eck HJ, Jacobsen E, Visser RGF. 2007. Genetic Positioning of Centromeres Using Half-Tetrad Analysis in a 4 *x* –2 *x* Cross Population of Potato. *Genetics* 176:85–94. doi:10.1534/genetics.107.070870
Peloquin SJ, Boiteux LS, Simon PW, Jansky SH. 2008. A Chromosome-Specific Estimate of Transmission of Heterozygosity by 2n Gametes in Potato. *Journal of Heredity* 99:177–181. doi:10.1093/jhered/esm110
Persson L, Carney Almroth BM, Collins CD, Cornell S, de Wit CA, Diamond ML, Fantke P, Hassellöv M, MacLeod M, Ryberg MW, Søgaard Jørgensen P, Villarrubia-Gómez P, Wang Z, Hauschild MZ. 2022. Outside the Safe Operating Space of the Planetary Boundary for Novel Entities. *Environ Sci Technol* 56:1510–1521. doi:10.1021/acs.est.1c04158
Peterson BA, Holt SH, Laimbeer FPE, Doulis AG, Coombs J, Douches DS, Hardigan MA, Buell CR, Veilleux RE. 2016. Self-Fertility in a Cultivated Diploid Potato Population Examined with the Infinium 8303 Potato Single-Nucleotide Polymorphism Array. *The Plant Genome* 9:0. doi:10.3835/plantgenome2016.01.0003

Peterson R, Slovin JP, Chen C. 2010. A simplified method for differential staining of aborted and non-aborted pollen grains. *International Journal of Plant Biology* 1:66–69. doi:10.4081/pb.2010.e13

Pham GM, Hamilton JP, Wood JC, Burke JT, Zhao H, Vaillancourt B, Ou S, Jiang J, Buell CR. 2020. Construction of a chromosome-scale long-read reference genome assembly for potato. *GigaScience* 9:1–11. doi:10.1093/gigascience/giaa100

Phumichai C, Hosaka K. 2006. Cryptic improvement for fertility by continuous selfing of diploid potatoes using Sli gene. *Euphytica* 149:251–258. doi:10.1007/s10681-005-9072-5

Phumichai C, Hosaka K, Mori M, Kamijima O, Kobayashi A. 2006. Toward the development of highly homozygous diploid potato lines using the self-compatibility controlling Sli gene. *Genome* 48:977–984. doi:10.1139/g05-066

Piepho H-P, Möhring J. 2007. Computing Heritability and Selection Response From Unbalanced Plant Breeding Trials. *Genetics* 177:1881–1888. doi:10.1534/genetics.107.074229

Plaisted RL, Hoopes RW. 1989. The past record and future prospects for the use of exotic potato germplasm. *American Potato Journal* 66:603–627. doi:10.1007/BF02853982

Preedy KF, Hackett CA. 2016. A rapid marker ordering approach for high-density genetic linkage maps in experimental autotetraploid populations using multidimensional scaling. *Theoretical and Applied Genetics* 129:2117–2132. doi:10.1007/s00122-016-2761-8

Prodhomme C, Esselink D, Borm T, Visser RGF, van Eck HJ, Vossen JH. 2019. Comparative Subsequence Sets Analysis (CoSSA) is a robust approach to identify haplotype specific SNPs; mapping and pedigree analysis of a potato wart disease resistance gene Sen3. *Plant Methods* 15:60. doi:10.1186/s13007-019-0445-5

PSGC. 2011. Genome sequence and analysis of the tuber crop potato. *Nature* 475:189–195. doi:10.1038/nature10158

Pucker B, Irisarri I, Vries J de, Xu B. 2022. Plant genome sequence assembly in the era of long reads: Progress, challenges and future directions. *Quantitative Plant Biology* 3:e5. doi:10.1017/qpb.2021.18

Purugganan MD, Fuller DQ. 2009. The nature of selection during plant domestication. *Nature* 457:843–848. doi:10.1038/nature07895

Pyatnitskaya A, Borde V, De Muyt A. 2019. Crossing and zipping: molecular duties of the ZMM proteins in meiosis. *Chromosoma* 128:181–198. doi:10.1007/s00412-019-00714-8

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842. doi:10.1093/bioinformatics/btq033

Rabinowitz D, Linder CR, Ortega R, Begazo D, Murguia H, Douches DS, Quiros CF. 1990. High levels of interspecific hybridization between solanum sparsipilum and s. stenotomum in experimental plots in the andes. *American Potato Journal* 67:73–81. doi:10.1007/BF02990957

Ramanna MS. 1983. First division restitution gametes through fertile desynaptic mutants of potato. *Euphytica* 32:337–350. doi:10.1007/BF00021442

Ramanna MS. 1979. A re-examination of the mechanisms of 2n gamete formation in potato and its implications for breeding. *Euphytica* 28:537–561. doi:10.1007/BF00038921

Ramanna MS, Jacobsen E. 2003. Relevance of sexual polyploidization for crop improvement - A review. *Euphytica* 133:3–8. doi:10.1023/A:1025600824483

Ramírez Gonzales L, Shi L, Bergonzi SB, Oortwijn M, Franco-Zorrilla JM, Solano-Tavira R, Visser RGF, Abelenda JA, Bachem CWB. 2021. Potato CYCLING DOF FACTOR 1 and its lncRNA counterpart *StFLORE* link tuber development and drought response. *The Plant Journal* tpj.15093. doi:10.1111/tpj.15093

Ramsey J, Schemske DW. 1998. PATHWAYS, MECHANISMS, AND RATES OF POLYPLOID FORMATION IN FLOWERING PLANTS. *Annual Review of Ecology and Systematics* 29:467–501. doi:10.1146/annurev.ecolsys.29.1.467

Renner SS, Ricklefs RE. 1995. Dioecy and its correlates in the flowering plants. *American journal of botany* 596–606. doi:10.5282/ubm/epub.14619

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

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nature biotechnology* 29:24–6. doi:10.1038/nbt.1754

Salman-Minkov A, Sabath N, Mayrose I. 2016. Whole-genome duplication as a key factor in crop domestication. *Nature Plants* 2:16115. doi:10.1038/NPLANTS.2016.115

Sanetomo R, Akai K, Nashiki A. 2022. Discovery of a novel mitochondrial DNA molecule associated with tetrad pollen sterility in potato. *BMC Plant Biology* 22:302. doi:10.1186/s12870-022-03669-8

Sanetomo R, Gebhardt C. 2015. Cytoplasmic genome types of European potatoes and their effects on complex agronomic traits. *BMC Plant Biology* 15:1–16. doi:10.1186/s12870-015-0545-y

Sanford JC, Hanneman RE. 1982. A possible heterotic threshold in the potato and its implications for breeding. *Theoret Appl Genetics* 61:151–159. doi:10.1007/BF00273884

Sattler MC, Carvalho CR, Clarindo WR. 2016. The polyploidy and its key role in plant breeding. *Planta* 243:281–296. doi:10.1007/s00425-015-2450-x

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez JY, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A. 2012. Fiji: An open-source platform for biological-image analysis. *Nature Methods* 9:676–682. doi:10.1038/nmeth.2019

Schneeberger K, Ossowski S, Lanz C, Juul T, Petersen AH, Nielsen KL, Jørgensen J-E, Weigel D, Andersen SU. 2009. SHOREmap: simultaneous mapping and mutation identification by deep sequencing. *Nature Methods* 6:550–551. doi:10.1038/nmeth0809-550

Scurrah M, Celis-Gamboa C, Chumbiauca S, Salas A, Visser RGF. 2008. Hybridization between wild and cultivated potato species in the Peruvian Andes and biosafety implications for deployment of GM potatoes. *Euphytica* 164:881–892. doi:10.1007/s10681-007-9641-x

Seibert T, Abel C, Wahl V. 2020. Flowering time and the identification of floral marker genes in Solanum tuberosum ssp. andigena. *Journal of Experimental Botany* 71:986–996. doi:10.1093/jxb/erz484

Serrentino M-E, Borde V. 2012. The spatial regulation of meiotic recombination hotspots: Are all DSB hotspots crossover hotspots? *Experimental Cell Research*, Experimental Cell Research Special Review Issue: Chromosome Biology, 2012 318:1347–1352. doi:10.1016/j.yexcr.2012.03.025

Sharma SK, Bolser D, de Boer J, Sønderkær M, Amoros W, Carboni MF, D'Ambrosio JM, de la Cruz G, Di Genova A, Douches DS, Eguiluz M, Guo X, Guzman F,

Hackett CA, Hamilton JP, Li G, Li Y, Lozano R, Maass A, Marshall D, Martinez D, McLean K, Mejía N, Milne L, Munive S, Nagy I, Ponce O, Ramirez M, Simon R, Thomson SJ, Torres Y, Waugh R, Zhang Z, Huang S, Visser RGF, Bachem CWB, Sagredo B, Feingold SE, Orjeda G, Veilleux RE, Bonierbale M, Jacobs JME, Milbourne D, Martin DMA, Bryan GJ. 2013. Construction of Reference Chromosome-Scale Pseudomolecules for Potato: Integrating the Potato Genome with Genetic and Physical Maps. *G3: Genes, Genomes, Genetics* 3:2031–2047. doi:10.1534/g3.113.007153

Shi D, Wu Jun, Tang H, Yin H, Wang H, Wang Ran, Wang Runze, Qian M, Wu Juyou, Qi K, Xie Z, Wang Z, Zhao X, Zhang S. 2019. Single-pollen-cell sequencing for gamete-based phased diploid genome assembly in plants. *Genome Res* 29:1889–1899. doi:10.1101/gr.251033.119

Shumate A, Salzberg S. 2022. LiftoffTools: a toolkit for comparing gene annotations mapped between genome assemblies. doi:10.12688/f1000research.124059.1

Shumate A, Salzberg SL. 2021. Liftoff: accurate mapping of gene annotations. *Bioinformatics* 37:1639–1643. doi:10.1093/bioinformatics/btaa1016

Sijacic P, Wang X, Skirpan AL, Wang Y, Dowd PE, McCubbin AG, Huang S, Kao T. 2004. Identification of the pollen determinant of S-RNase-mediated self-incompatibility. *Nature* 429:302–305. doi:10.1038/nature02523

Song L, Endelman JB. 2022. Using haplotype and QTL analysis to fix favorable alleles in diploid potato breeding. *The Plant Genome* 16:e20339. doi:10.1002/tpg2.20339

Spooner DM, McLean K, Ramsay G, Waugh R, Bryan GJ. 2005a. A single domestication for potato based on multilocus amplified fragment length polymorphism genotyping. *Proceedings of the National Academy of Sciences* 102:14694–14699. doi:10.1073/pnas.0507400102

Spooner DM, Nuñez J, Rodríguez F, Naik PS, Ghislain M. 2005b. Nuclear and chloroplast DNA reassessment of the origin of Indian potato varieties and its implications for the origin of the early European potato. *Theor Appl Genet* 110:1020–1026. doi:10.1007/s00122-004-1917-0

Stam P. 1993. Construction of integrated genetic linkage maps by means of a new computer package: Join Map. *The Plant Journal* 3:739–744. doi:10.1111/j.1365-313X.1993.00739.x

Steffen W, Richardson K, Rockström J, Cornell SE, Fetzer I, Bennett EM, Biggs R, Carpenter SR, de Vries W, de Wit CA, Folke C, Gerten D, Heinke J, Mace GM, Persson LM, Ramanathan V, Reyers B, Sörlin S. 2015. Planetary boundaries: Guiding human development on a changing planet. *Science* 347:1259855. doi:10.1126/science.1259855

Stokstad E. 2019. The new potato. *Science* 363:574–577. doi:10.1126/science.363.6427.574

Su X, Wang B, Geng X, Du Y, Yang Q, Liang B, Meng G, Gao Q, Yang W, Zhu Y, Lin T. 2021. A high-continuity and annotated tomato reference genome. *BMC Genomics* 22:898. doi:10.1186/s12864-021-08212-x

Sukhotu T, Hosaka K. 2006. Origin and evolution of Andigena potatoes revealed by chloroplast and nuclear DNA markers. *Genome* 49:636–647. doi:10.1139/g06-014

Sun H, Jiao W-B, Krause K, Campoy JA, Goel M, Folz-Donahue K, Kukat C, Huettel B, Schneeberger K. 2022. Chromosome-scale and haplotype-resolved genome assembly of a tetraploid potato cultivar. *Nat Genet* 54:342–348. doi:10.1038/s41588-022-01015-0

Sun L, Williams JS, Li S, Wu L, Khatri WA, Stone PG, Keebaugh MD, Kao T. 2018. S-Locus F-Box Proteins Are Solely Responsible for S-RNase-Based Self-

Incompatibility of Petunia Pollen. *The Plant Cell* 30:2959–2972. doi:10.1105/tpc.18.00615

Sun Y, Ambrose JH, Haughey BS, Webster TD, Pierrie SN, Muñoz DF, Wellman EC, Cherian S, Lewis SM, Berchowitz LE, Copenhaver GP. 2012. Deep Genome-Wide Measurement of Meiotic Gene Conversion Using Tetrad Analysis in Arabidopsis thaliana. *PLOS Genetics* 8:e1002968. doi:10.1371/journal.pgen.1002968

Takayama S, Isogai A. 2005. Self-Incompatibility in Plants. *Annual Review of Plant Biology* 56:467–489. doi:10.1146/annurev.arplant.56.032604.144249

Tang D, Jia Y, Zhang J, Li H, Cheng L, Wang P, Bao Z, Liu Z, Feng S, Zhu X, Li D, Zhu G, Wang H, Zhou Yao, Zhou Yongfeng, Bryan GJ, Buell CR, Zhang C, Huang S. 2022a. Genome evolution and diversity of wild and cultivated potatoes. *Nature* 606:535–541. doi:10.1038/s41586-022-04822-x

Tang D, Jia Y, Zhang J, Li H, Cheng L, Wang P, Bao Z, Liu Z, Feng S, Zhu X, Li D, Zhu G, Wang H, Zhou Yao, Zhou Yongfeng, Bryan GJ, Robin Buell C, Zhang C, Huang S. 2022b. Genome evolution and diversity of wild and cultivated potatoes. *Nature 2022* 1–7. doi:10.1038/s41586-022-04822-x

Tayalé A, Parisod C. 2013. Natural pathways to polyploidy in plants and consequences for genome reorganization. *Cytogenetic and Genome Research* 140:79–96. doi:10.1159/000351318

Tello J, Montemayor MI, Forneck A, Ibáñez J. 2018. A new image-based tool for the high throughput phenotyping of pollen viability: evaluation of inter- and intra-cultivar diversity in grapevine. *Plant methods* 14:3. doi:10.1186/s13007-017-0267-2

Thérèse Navarro A, Bourke PM, van de Weg E, Clot CR, Arens P, Finkers R, Maliepaard C. 2023. Smooth Descent: A ploidy-aware algorithm to improve linkage mapping in the presence of genotyping errors. *Frontiers in Genetics* 14.

Thomas-Sharma S, Abdurahman A, Ali S, Andrade-Piedra JL, Bao S, Charkowski AO, Crook D, Kadian M, Kromann P, Struik PC, Torrance L, Garrett KA, Forbes GA. 2016. Seed degeneration in potato: the need for an integrated seed health strategy to mitigate the problem in developing countries. *Plant Pathology* 65:3–16. doi:10.1111/ppa.12439

Thompson RD, Uhrig H, Hermsen JGTh, Salamini F, Kaufmann H. 1991. Investigation of a self-compatible mutation in Solanum tuberosum clones inhibiting S-allele activity in pollen differentially. *MGG Molecular & General Genetics* 226–226:283–288. doi:10.1007/BF00273614

Tribhuvan KU, Sandhya, Kumar K, Sevanthi AM, Gaikwad K. 2018. MutMap: a versatile tool for identification of mutant loci and mapping of genes. *Indian Journal of Plant Physiology* 23:612–621. doi:10.1007/s40502-018-0417-1

Uijtewaal BA, Huigen DJ, Hermsen JGTh. 1987a. Production of potato monohaploids (2n=x=12) through prickle pollination. *Theoretical and Applied Genetics* 73:751–758. doi:10.1007/BF00260786

Uijtewaal BA, Jacobsen E, Hermsen JGT. 1987b. Morphology and vigour of monohaploid potato clones, their corresponding homozygous diploids and tetraploids and their heterozygous diploid parent. *Euphytica* 36:745–753. doi:10.1007/BF00051857

Uitdewilligen JGAML, Wolters A-MA, D'hoop BB, Borm TJA, Visser RGF, van Eck HJ. 2013. A Next-Generation Sequencing Method for Genotyping-by-Sequencing of Highly Heterozygous Autotetraploid Potato. *PLoS ONE* 8:e62355. doi:10.1371/journal.pone.0062355

van Berloo R, Hutten RCB, van Eck HJ, Visser RGF. 2007. An Online Potato Pedigree Database Resource. *Potato Research* 50:45–57. doi:10.1007/s11540-007-9028-3

van de Weg WE, Voorrips RE, Finkers R, Kodde LP, Jansen J, Bink MCAM. 2004. Pedigree genotyping: A new pedigree-based approach of QTL identification and allele miningActa Horticulturae. International Society for Horticultural Science. pp. 45–50. doi:10.17660/ActaHortic.2004.663.1

van der Meer P, Angenon G, Bergmans H, Buhk HJ, Callebaut S, Chamon M, Eriksson D, Gheysen G, Harwood W, Hundleby P, Kearns P, Mcloughlin T, Zimny T. 2023. The Status under EU Law of Organisms Developed through Novel Genomic Techniques. *European Journal of Risk Regulation* 14:93–112. doi:10.1017/err.2020.105

van Dijk LCM, Kacheyo OC, de Vries ME, Lommen WJM, Struik PC. 2022. Crop Cycle Length Determines Optimal Transplanting Date for Seedlings from Hybrid True Potato Seeds. *Potato Res* 65:435–460. doi:10.1007/s11540-021-09524-x

van Dijk LCM, Lommen WJM, de Vries ME, Kacheyo OC, Struik PC. 2021. Hilling of Transplanted Seedlings from Novel Hybrid True Potato Seeds Does Not Enhance Tuber Yield but Can Affect Tuber Size Distribution. *Potato Res* 64:353–374. doi:10.1007/s11540-020-09481-x

van Eck HJ, Oortwijn, Maria E.P., Terpstra, Inez R., van Lieshout, Natascha H.M., van der Knaap,Esther, Willemsen, Johan H., Bachem, Christian W.B. 2022. Engineering of tuber shape in potato (Solanum tuberosum) with marker assisted breeding or genetic modification using StOFP20. doi:10.21203/rs.3.rs-1807189/v1

van Eck HJ, Vos PG, Valkonen JPT, Uitdewilligen JGAML, Lensing H, de Vetten N, Visser RGF. 2017. Graphical genotyping as a method to map Ny (o,n)sto and Gpa5 using a reference panel of tetraploid potato cultivars. *Theoretical and Applied Genetics* 130:515–528. doi:10.1007/s00122-016-2831-y

van Lieshout N, van der Burgt A, de Vries ME, ter Maat M, Eickholt D, Esselink D, van Kaauwen MPW, Kodde LP, Visser RGF, Lindhout P, Finkers R. 2020a. Solyntus, the New Highly Contiguous Reference Genome for Potato (Solanum tuberosum). *G3; Genes|Genomes|Genetics* g3.401550.2020. doi:10.1534/g3.120.401550

van Lieshout N, van der Burgt A, de Vries ME, ter Maat M, Eickholt D, Esselink D, van Kaauwen MPW, Kodde LP, Visser RGF, Lindhout P, Finkers R. 2020b. Solyntus, the New Highly Contiguous Reference Genome for Potato (Solanum tuberosum). *G3 Genes|Genomes|Genetics* 10:3489–3495. doi:10.1534/g3.120.401550

van Os H, Andrzejewski S, Bakker E, Barrena I, Bryan GJ, Caromel B, Ghareeb B, Isidore E, de Jong W, van Koert P, Lefebvre V, Milbourne D, Ritter E, van der Voort JNAMR, Rousselle-Bourgeois F, van Vliet J, Waugh R, Visser RGF, Bakker J, van Eck HJ. 2006. Construction of a 10,000-Marker Ultradense Genetic Recombination Map of Potato: Providing a Framework for Accelerated Gene Isolation and a Genomewide Physical Map. *Genetics* 173:1075–1087. doi:10.1534/genetics.106.055871

van Os H, Stam P, Visser RGF, van Eck HJ. 2005. RECORD: a novel method for ordering loci on a genetic linkage map. *Theor Appl Genet* 112:30–40. doi:10.1007/s00122-005-0097-x

Vos PG, Uitdewilligen JGAML, Voorrips RE, Visser RGF, van Eck HJ. 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

Wan X, Wu S, Li Z, An X, Tian Y. 2020. Lipid Metabolism: Critical Roles in Male Fertility and Other Aspects of Reproductive Development in Plants. *Molecular Plant* 13:955–983. doi:10.1016/j.molp.2020.05.009

Wang J, Hill WG, Charlesworth D, Charlesworth B. 1999. Dynamics of inbreeding depression due to deleterious mutations in small populations: mutation parameters and inbreeding rate. *Genetics Research* 74:165–178. doi:10.1017/S0016672399003900

Wang-Erlandsson L, Tobian A, van der Ent RJ, Fetzer I, te Wierik S, Porkka M, Staal A, Jaramillo F, Dahlmann H, Singh C, Greve P, Gerten D, Keys PW, Gleeson T, Cornell SE, Steffen W, Bai X, Rockström J. 2022. A planetary boundary for green water. *Nat Rev Earth Environ* 3:380–392. doi:10.1038/s43017-022-00287-8

Washburn JD, Birchler JA. 2014. Polyploids as a "model system" for the study of heterosis. *Plant Reprod* 27:1–5. doi:10.1007/s00497-013-0237-4

Washburn JD, McElfresh MJ, Birchler JA. 2019. Progressive heterosis in genetically defined tetraploid maize. *Journal of Genetics and Genomics* 46:389–396. doi:10.1016/j.jgg.2019.02.010

Waskom M, Botvinnik O, Hobson P, Cole JB, Halchenko Y, Hoyer S, Miles A, Augspurger T, Yarkoni T, Megies T, Coelho LP, Wehner D, cynddl, Ziegler E, diego0020, Zaytsev YV, Hoppe T, Seabold S, Cloud P, Koskinen M, Meyer K, Qalieh A, Allan D. 2014. seaborn: v0.5.0 (November 2014). doi:10.5281/ZENODO.12710

Watanabe K, Peloquin SJ. 1989. Occurrence of 2n pollen and ps gene frequencies in cultivated groups and their related wild species in tuber-bearing Solanums. *Theoretical and Applied Genetics* 78:329–336. doi:10.1007/BF00265292

Wenzel G, Schieder O, Przewozny T, Sopory SK, Melchers G. 1979. Comparison of single cell culture derived Solanum tuberosum L. plants and a model for their application in breeding programs. *Theoretical and Applied Genetics* 55:49–55. doi:10.1007/BF00285189

Werner JE, Peloquin SJ. 1990. Inheritance and Two Medchanisms of 2n Egg Fromation in 2x Potatoes. *Journal of Heredity* 81:371–374. doi:10.1093/oxfordjournals.jhered.a111004

Wijeratne AJ, Chen C, Zhang W, Timofejeva L, Ma H. 2006. The Arabidopsis thaliana PARTING DANCERS Gene Encoding a Novel Protein Is Required for Normal Meiotic Homologous Recombination. *MBoC* 17:1331–1343. doi:10.1091/mbc.e05-09-0902

Wilkinson MD, Dumontier M, Aalbersberg IjJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJG, Groth P, Goble C, Grethe JS, Heringa J, t Hoen PAC, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone SA, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, Van Der Lei J, Van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B. 2016. Comment: The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3. doi:10.1038/sdata.2016.18

Wilson ZA, Zhang D-B. 2009. From Arabidopsis to rice: pathways in pollen development. *Journal of Experimental Botany* 60:1479–1492. doi:10.1093/jxb/erp095

Wolters A-MA, Uitdewilligen JGAML, Kloosterman BA, Hutten RCB, Visser RGF, van Eck HJ. 2010. Identification of alleles of carotenoid pathway genes important for zeaxanthin accumulation in potato tubers. *Plant Mol Biol* 73:659–671. doi:10.1007/s11103-010-9647-y

Wu D, Wang Z, Dong W, Li G, Wu Dewei. 2018. Next-generation sequencing from bulked segregant analysis identifies a dwarfism gene in watermelon. *Scientific Reports* 8:1–7. doi:10.1038/s41598-018-21293-1

Wu S, Zhang B, Keyhaninejad N, Rodríguez GR, Kim HJ, Chakrabarti M, Illa-Berenguer E, Taitano NK, Gonzalo MJ, Díaz A, Pan Y, Leisner CP, Halterman D, Buell CR, Weng Y, Jansky SH, van Eck H, Willemsen J, Monforte AJ, Meulia T, van der Knaap E. 2018. A common genetic mechanism underlies morphological diversity in fruits and other plant organs. *Nat Commun* 9:4734. doi:10.1038/s41467-018-07216-8

Wu Y, Li D, Hu Y, Li H, Ramstein GP, Zhou S, Zhang X, Bao Z, Zhang Y, Song B, Zhou Yao, Zhou Yongfeng, Gagnon E, Särkinen T, Knapp S, Zhang C, Städler T, Buckler ES, Huang S. 2023. Phylogenomic discovery of deleterious mutations facilitates hybrid potato breeding. *Cell* 186. doi:10.1016/j.cell.2023.04.008

Yang X, Zhang L, Guo X, Xu J, Zhang K, Yang Yinqing, Yang Yu, Jian Y, Dong D, Huang S, Cheng F, Li G. 2023. The gap-free potato genome assembly reveals large tandem gene clusters of agronomical importance in highly repeated genomic regions. *Molecular Plant* 16:314–317. doi:10.1016/j.molp.2022.12.010

Yant L, Hollister JD, Wright KM, Arnold BJ, Higgins JD, Franklin FCH, Bomblies K. 2013. Meiotic Adaptation to Genome Duplication in Arabidopsis arenosa. *Current Biology* 23:2151–2156. doi:10.1016/j.cub.2013.08.059

Yates AD, Allen J, Amode RM, Azov AG, Barba M, Becerra A, Bhai J, Campbell LI, Carbajo Martinez M, Chakiachvili M, Chougule K, Christensen M, Contreras-Moreira B, Cuzick A, Da Rin Fioretto L, Davis P, De Silva NH, Diamantakis S, Dyer S, Elser J, Filippi CV, Gall A, Grigoriadis D, Guijarro-Clarke C, Gupta P, Hammond-Kosack KE, Howe KL, Jaiswal P, Kaikala V, Kumar V, Kumari S, Langridge N, Le T, Luypaert M, Maslen GL, Maurel T, Moore B, Muffato M, Mushtaq A, Naamati G, Naithani S, Olson A, Parker A, Paulini M, Pedro H, Perry E, Preece J, Quinton-Tulloch M, Rodgers F, Rosello M, Ruffier M, Seager J, Sitnik V, Szpak M, Tate J, Tello-Ruiz MK, Trevanion SJ, Urban M, Ware D, Wei S, Williams G, Winterbottom A, Zarowiecki M, Finn RD, Flicek P. 2022. Ensembl Genomes 2022: an expanding genome resource for non-vertebrates. *Nucleic Acids Research* 50:D996–D1003. doi:10.1093/nar/gkab1007

Ye M, Peng Z, Tang D, Yang Z, Li D, Xu Y, Zhang C, Huang S. 2018. Generation of self-compatible diploid potato by knockout of S-RNase. *Nature Plants* 4:651–654. doi:10.1038/s41477-018-0218-6

Younis A, Hwang YJ, Lim KB. 2014. Exploitation of induced 2n-gametes for plant breeding. *Plant Cell Reports* 33:215–223. doi:10.1007/s00299-013-1534-y

Zhang C, Shaw KM, Alsahlany MA, Manrique-Carpintero NC, Garcia-Ramirez A, Zarka D, Coombs J, Douches D. 2019. Dihaploid Potato Breeding at Michigan State University - Annual Report of the Potato Association of America. *American Journal of Potato Research* 94:360. doi:10.1007/s12230-019-09723-w

Zhang Czhi, Wang P, Tang D, Yang Z, Lu F, Qi J, Tawari NR, Shang Y, Li C, Huang S. 2019. The genetic basis of inbreeding depression in potato. *Nature Genetics* 51:374–378. doi:10.1038/s41588-018-0319-1

Zhang C, Yang Z, Tang D, Zhu Y, Wang P, Li D, Zhu G, Xiong X, Shang Y, Li C, Huang S. 2021. Genome design of hybrid potato. *Cell* 184:3873-3883.e12. doi:10.1016/j.cell.2021.06.006

Zheng C, Amadeu RR, Munoz PR, Endelman JB. 2021. Haplotype Reconstruction in Connected Tetraploid F1 Populations. *Genetics* 2020.12.18.423519. doi:10.1093/genetics/iyab106

Zheng X, He L, Liu Y, Mao Y, Wang C, Zhao B, Li Y, He H, Guo S, Zhang L, Schneider H, Tadege M, Chang F, Chen J. 2020. A study of male fertility control in Medicago truncatula uncovers an evolutionarily conserved recruitment of two tapetal bHLH subfamilies in plant sexual reproduction. *New Phytologist* 228:1115–1133. doi:10.1111/nph.16770

Zhou C, Olukolu B, Gemenet DC, Wu S, Gruneberg W, Cao MD, Fei Z, Zeng Z-B, George AW, Khan A, Yencho GC, Coin LJM. 2020. Assembly of whole-chromosome pseudomolecules for polyploid plant genomes using outbred mapping populations. *Nat Genet* 52:1256–1264. doi:10.1038/s41588-020-00717-7

Zhou Q, Tang D, Huang W, Yang Z, Zhang Y, Hamilton JP, Visser RGF, Bachem CWB, Robin Buell C, Zhang Z, Zhang C, Huang S. 2020. Haplotype-resolved genome analyses of a heterozygous diploid potato. *Nature Genetics* 52:1018–1023. doi:10.1038/s41588-020-0699-x

Zhou Y, Zhang Zhiyang, Bao Z, Li H, Lyu Y, Zan Y, Wu Y, Cheng L, Fang Y, Wu K, Zhang J, Lyu H, Lin T, Gao Q, Saha S, Mueller L, Fei Z, Städler T, Xu S, Zhang Zhiwu, Speed D, Huang S. 2022. Graph pangenome captures missing heritability and empowers tomato breeding. *Nature* 606:527–534. doi:10.1038/s41586-022-04808-9

Zickler D, Kleckner N. 2015. Recombination, Pairing, and Synapsis of Homologs during Meiosis. *Cold Spring Harb Perspect Biol* 7:a016626. doi:10.1101/cshperspect.a016626

Summary

A major aspect of plant breeding relies on gaining control over the reproductive biology of crop species to accelerate their improvement. Potato (*Solanum tuberosum*) is the 3rd most important food crop after wheat and rice and a staple food for 1.3 billion people worldwide. Despite the economic and nutritional importance of potato, its genetic improvement showed less progression compared to other major food crops. In part, this is due to restrictive reproductive characteristics such as allogamy and tetraploidy which limit potato breeding. The main objective of this thesis was to explore the genetic basis of naturally occurring variations in potato sexual reproduction. In combinations these variants can be used to increase potato breeding efficiency. The focus was set on self-compatibility which facilitates the fixation of genetic gains, crossover reduction which limits allele dispersal during meiosis, and unreduced pollen production which enables interploidy crosses.

Chapter 2 describes the genetic basis and prevalence of self-compatibility in *S. tuberosum* germplasm. In this chapter, we began by investigating the inheritance of self-compatibility in two diploid potato bi-parental populations. In both populations, we mapped self-compatibility on chromosome *12* using a *k*-mer based bulk segregant analysis. I further used those *k*-mers for haplotype mining and realised that the self-compatible haplotype we identified was indistinguishable from the *Sli* haplotype, originally described in *S. chacoense*. To our surprise, this haplotype was rather prevalent among tetraploid *S. tuberosum* cultivars where its effect was masked by the breakdown of gametophytic self-incompatibility at the tetraploid level. Finally, we developed a set of KASP markers enabling marker-assisted selection of self-compatibility.

Chapter 3 exemplifies the range of genetic analyses, from marker inference, identification of transmission ratio distortion, linkage map construction to QTL mapping, which can be used to generate new insights in diploid potato genetics. In this chapter, we skim-sequenced a large bi-parental diploid potato population and used the obtained data to create high-density linkage maps. The parental maps showed different recombination rates between the maternal and paternal parent, along with structural variations, notably a 5.8 Mb inversion on chromosome *3*. Our focus then shifted to transmission ratio distortion, which can serve as a potential indicator of significant deleterious mutations. We discovered 10 loci that showed transmission ratio distortion and we identified four pairs of loci involved in non-independent assortment of alleles. Moving forward, we conducted a QTL analysis for seedling tuber yield in pots and pollen shed. The tuber yield in pots was primarily influenced by a major-effect QTL that colocalized with *StCDF1*. Additionally, we identified five minor-effect QTLs that interacted with different *StCDF1* allele combinations. As for pollen shed, a critical

factor for male fertility, we found a total of seven QTLs, including a major-effect locus situated on chromosome *2*.

Chapter 4 describes the genetic basis of meiotic crossover reduction, a phenotype historically referred to as desynapsis, in the population genotyped in the earlier chapter. Crossovers are essential for the maintenance of homologous chromosomes pairing. A drastic reduction of their frequency leads to a chaotic first meiotic division and to the formation of unbalanced haploid gametes. However, if desynapsis is combined with another meiotic mutation that restitutes the first division, then balanced unreduced gametes are produced. Using pollen microscopy, we classified descendants with stainable haploid pollen as synaptic while descendants with unstainable haploid and stainable unreduced pollen were considered desynaptic. Next, we conducted a QTL analysis and found a single causal locus located on the short arm of chromosome *8*. Within this region, the most promising candidate gene was *StMSH4*, a crucial component of the class I crossover pathway. We further discovered that a 7 bp insertion in the exonic region of *StMSH4* was associated with desynapsis in our mapping population. Using allele mining, we found another mutant allele featuring a larger exonic insertion of 3820 bp, and we confirmed that both alleles were unable to complement each other. Subsequently, we showed with a *k*-mer based approach the prevalence of non-functional *StMSH4* alleles in the European tetraploid germplasm.

Chapter 5 investigates the mechanisms of formation and the mode of inheritance of unreduced pollen production in three diploid potato bi-parental populations. Unreduced pollen production is a valuable trait for potato breeders which allow them to link diploid and tetraploid germplasms via interploidy crosses. In all three phenotyped populations, variation in the level of unreduced pollen production was primarily attributed to genotypic effects and this trait segregated quantitatively. We then showed that five QTLs regulated this trait, one of which was shared by all three populations. The two QTLs with the largest effect displayed a recessive inheritance and an additive interaction. Notably, both QTLs colocalised with genes that showed homology to *AtIAS*, a key regulator of meiosis II spindle orientation in *Arabidopsis thaliana*. The function of these candidate genes is consistent with the cytological phenotype of mis-oriented metaphase II plates we observed in some of the parental clones. Furthermore, the alleles associated with an elevated level of unreduced pollen displayed deleterious mutations, specifically an exonic transposon insert causing a premature stop, and an amino acid change within a highly conserved domain. Overall, the knowledge acquired in this chapter offers new insights into the natural allelic variation underlying unreduced pollen production in potato. Using markers it will now be possible to select breeding clones with a desirable level of unreduced pollen production, allowing diploid crosses with haploid pollen and interploidy crosses with diploid pollen.

Chapter 6 offers a discussion on the main findings of this thesis in the context of potato breeding. In this last chapter, I highlighted the advantages and limitations of several potato breeding strategies with a particular emphasis on the challenges posed by inbreeding depression. I argue that Fixation-Restitution breeding, an approach leveraging self-compatibility, crossover reduction, and unreduced pollen production, can improve potato breeding efficiency without relying on highly homozygous inbred lines. In this chapter, I also reflect on the forward genetic strategy employed throughout this thesis. I argue that adopting a "translational genomics" approach could speed up the discovery of sequence variants affecting potato sexual reproduction.

Acknowledgements

Looking back at this PhD journey through the deterministic lens advocated in one of my propositions, it becomes evident that the completion of this thesis is attributable to a multitude of circumstances, including the invaluable support I received from family, friends, and colleagues. In this ultimate section, I would like to express my gratitude to those people who accompanied me during this adventure.

Before proceeding, I sincerely apologize if your name is absent from the following acknowledgments. At present, my mind is overloaded by potatoes, their genetics, and reproductive peculiarities. To each and every one of you, whether mentioned here or not, I am grateful for the role you have played in my journey.

First and foremost, I would like to offer my gratitude and appreciation to my supervisor **Herman van Eck**. **Herman**, you are the embodiment of the applied scientist who resolutely rejects the idea of confining oneself to academia ivory tower. I admire your sharp and inquisitive mindset and have learned a lot from your extensive knowledge of potato genetics. I am indebted to you for helping me grow as a scientist, notably by giving me the opportunity to substitute you for a talk at PAG during the first year of my PhD journey. I am glad to continue our fruitful scientific collaboration with a promising new TKI project.

To my promoter **Richard Visser**, thank you for your valuable feedback on my project and on its written output. Your insights, shared during our monthly meetings, have been precious, especially during Herman's absence. I greatly appreciated your availability and interest in this project and strive to learn from your pragmatism.

The completion of this project would have been unimaginable without the support of our academic breeders **Ronald Hutten** and **Christel Engelen**. **Ronald**, your encyclopaedic knowledge of potato pedigree and expertise in breeding across ploidy is truly unparalleled. **Christel**, your talent as a potato breeder merits my deepest appreciation, and I thank you for your support in the crossing greenhouse. To both of you, thank you pushing me out of the confinement of my microscope or computer room, immersing me in the tangible world of potato breeding. While on the topic of greenhouse work, I would also like to express my gratitude to **Dirk Jan Huigen** for his organisational support and for his overall friendliness.

To my external advisor **Erik Wijnker** I would like to express my deepest gratitude. **Erik**, thank you for guidance through this PhD journey. Your experience and insightful reflections on academia have been instrumental in shaping my research and personal growth. Your expertise in meiosis has been a delightful bonus, adding a special touch to our collaboration.

My sincere appreciation also goes to my colleagues and collaborators who both contributed to this project and to the developments of my skills. **Johan Bucher**, thank you for your support during wet lab experiments. Although some of your jokes may have flown over my head, I truly appreciate the energy you brought to the lab. **José van de Belt**, thank you for welcoming me in the laboratory of genetics and sharing with me your chromosome spread expertise. **Peter Bourke**, thank you for bearing with my questions while I was simultaneously discovering R and the packages you have developed. I wish to become a colleague as helpful as you are. **Richard Finkers**, thank you for your support in genotyping the largest CE population to date. **Martina Juraníc**, it was a pure bliss to have you involved in this project. I sincerely value your guidance during the cytological experiments, your feedback on my writing, and our insightful discussions on apomixis. Finally, I would like to extend my thanks to our partners in the Teagasc team. **Dan Milbourne**, **Maria de la O Leyva-Pérez** and **Lea Vexler**; it has been an absolute pleasure collaborating with all of you on this project. **Dan**, I am particularly thankful for your guidance and support during Herman's unfortunate absence. **Mariola** and **Lea**, I extend my thanks to both of you for your assistance with potatoMASH and for being such friendly and supportive colleagues.

I would like to extend my sincere appreciation to the exceptional staff of the private partners involved in this project: **Averis Seeds**, **Bejo Zaden**, **Danespo**, **Germicopa**, **Den Hartigh**, **Meijer** and **SaKa**. Without your collaboration and support, none of this would have been possible. A heartfelt thank you for consistently being ready to provide technical assistance and for your valuable suggestions shared during our consortium meetings. The team spirit that we have fostered throughout this project has been truly enjoyable.

During this project I had the privilege to supervise three talented Msc students. To **Petros Lioupis**, **Joey Koopman** and **Dennis Klein**, thank you for contributing to this research and to my own personal growth. It has been a pleasure working with each of you, and I have no doubt that you all have bright futures ahead.

To the wonderful secretaries of our department, **Nicole Trefflich** and **Daniëlle van der Wee-Uittenbogaard**, and the brilliant EPS coordinators **Susan Urbanus** and **Juliane Teapal**, thank you for your availability and for helping me navigating the academic bureaucracy.

To all past and present members of the **Quantitative Biology and Bioinformatics group**, I extend my profound gratitude for the insightful discussion we have shared on a weekly basis. A special thanks goes to **Martijn van Kaauwen**, **Danny Esselink**, and **Matthijs Brouwer** for their invaluable assistance in helping me navigate the initial stages of my bioinformatics endeavours and to **René Smulders** for his sharp scientific eyes and kind support.

To my **fellow PhD and colleagues** at Pant Breeding, while I will not attempt to individually name everyone, I want to express my sincere appreciation to each and every one of you. Thank you for making me feel welcome in this department, for the fruitful exchange of ideas, and for the enjoyable conversations we have had during lunch and coffee breaks. A special thanks goes to **Eleni Koseoglou** and **Daniel Moñino López** for their support during the final stretch of my PhD journey. Your presence and encouragement were invaluable, providing a sense of relief knowing that both of you had gone through similar pressures. **William de Martines**, thank you for the pizza and game nights and for helping emptying my house. **Guillermo Merino Martín**, merci de m'inviter régulièrement pour grimper (et prendre une bière), c'est toujours un plaisir !

To my office mates dwelling in E2.163 including **Cai Chengcheng**, **Elisabeth Yanez**, **Gabrielle Tjoen**, **Bettina Porta** and **Li Hongbo**, I want to thank you all for bearing with me all those years and for the friendly atmosphere you created. **Hongbo** and **Elisabeth** thank you for always being open to discuss potato genetics. **Gabi**, ce fut toujours un plaisir de parler français avec toi. **Betina** thank you for your infectious smile and positive energy. **Chengcheng**, you have been the perfect office mate. Your unwavering dedication to your work is truly inspiring, and at the same time, you have always made yourself available for a chat. You have been my bioinformatics cheat code throughout this PhD journey. Whenever I was on the verge of losing my sanity while trying to install a tool or decipher poorly documented code, I knew I could count on you for help. I also cherish our discussions on meiosis, and I will miss having you as a colleague.

To the molecular potato team including **Christian Bachem**, **Sarah Bergonzi**, **Csaba Papdi**, **Shi Li**, **Wang Xulan**, **Tu Beiyu**, **Ernst-Jan Eggers**, I extend my heartfelt appreciation for opening my eyes to the molecular intricacy of potato developmental biology. **Ernst**, in the first year of our PhDs, the partial overlap of our project did not facilitate our interaction. Once this original obstacle was overcome, I discovered a brilliant and friendly colleague in you. **Li**, I am grateful for your willingness to address all my molecular inquiries, whether during lunch breaks or when I randomly interrupted you in your office. **Xulan**, it has been an absolute pleasure collaborating with you across various settings be it behind a computer screen, in the lab, in the greenhouse, or even while sorting through hundreds of bags of sprouted tubers. Your unwavering positivity in the face of adversity is a quality that I truly admire, and I strive to learn from it. I also want to express my gratitude for accepting to be my paranymph, despite having already taken on this responsibility multiple times before.

I would like to extend my deepest gratitude to my friends outside of the department who have made life truly enjoyable. Starting with my fabulous Wageningen MSc friends: **Michele Serafini**, **Sebastian Tonn**, **Demetris Taliadoros** and **George Papaioannou**, thank you for all the fun we shared. **Michele**, it was blast spending some time with you

in Italy. **Jimmy**, you always brought a smile on my face, whether it was through your jokes around a BBQ or engaging in scientific discussions over a beer. **George** I can't wait to see you again be it in the Netherlands or in Cyprus. Last but not least **Basti**, the one who stayed. Thank you for being around all those years. It brought me great relief to be able to share with you the ups and downs of this adventure, knowing that you would understand. You are a precious friend and I truly value our walks, coffees and beers.

To **Witte Wilma** dwellers and their friends, thank you for welcoming me in your colourful community! I had a lot of fun and widen my perspective on life at your contact.

To my dear friends and ex-roommates **Vincent Spoor** and **Ambra Tosto**, I want to express my heartfelt appreciation for the wonderful times we have spent together. **Vincent**, I am envious of your artistic talent and it was always a pleasure accompanying your resonant voice and subtle guitar with my banjo. The songs we have shared, along with the glasses of wine and cigarettes, hold a special place in my heart. **Ambra**, you are not only one of my closest friends, but also an incredible person who has enriched my life in many ways. Thank you for sharing your delicious recipes, Italian songs, and botanical excitement with me. I truly admire your willingness to help others and your unwavering passion for both art and science. You have taught me a lot, not only in the culinary realm but also in the art of savouring life with friends.

To the other members of the "discussion group", **David Katzin**, **Lucie Sovová**, **Henry Payne** and **Linh Nguyen**, I would like to express my sincere gratitude for the intellectual stimulation and camaraderie we shared. Our gatherings, initially focused on discussing articles and book chapters, would often evolve into movie screenings, board game evenings, or simply food sharing. Regardless of the nature of our meetings, one thing remains certain: I always had a wonderful time in your company. I want to extend a special thanks to **David** for accepting to be my paranymph and for his intellectual wit which never fails to make me smile.

To my climbing partner and dear friend **Harmen den Braber**, thank you for the incredible adventures we have shared. **Harmen**, you taught me the importance of cultivating playfulness in adulthood and of pushing my limits on the climbing wall. I cherish our climbing trips, nature walks and long discussions. I am glad that you recently joined the potato world although it saddens me that this brought you away. I am looking forward to visit you in Kenya and discover local biodiversity, mountains and potato varieties.

To my musician friends from the **Madloot collective** and the **folk session at café De Zaaier** I want to express my deepest gratitude for the incredible tunes we have shared throughout the years. Our musical journeys have taken us from France to the Netherlands, passing through the Basque Country, Brittany, and taking beautiful

detours through Scandinavia, Turkey, Kurdistan, and Syria. These magical moments where sound has connected us across cultures and borders, hold a special place in my heart. I would like to express a special thanks to **Suzi Varga** for her positive energy and great violin skills. I want to extend my appreciation to **Wouter Kounders** for being an remarkable friend and for sharing countless stages with me. **Wouter**, you are an extremely talented player and composer and I often surprise myself humming your tunes while riding my bike. Playing alongside you has not only been a joy but has also motivated me to become a better musician. Last but certainly not least, I want to give my appreciation to the heart and soul of these music collectives, **Bert Lotz** and **Judica Lookman**. **Bert**, **Judica**, I don't even know where to begin... perhaps by acknowledging the music itself, as everything started from there. Thank you for warmly welcoming me and my guitar eight years ago at De Zaaier and for igniting my love for traditional folk music. From our very first encounter, your genuine and heart-warming generosity was palpable, as you invited me and fellow musicians for a simple yet greatly appreciated dinner. Over time, our connection grew deeper, and you became not just close friends, but also mentors and, at times, parental figures. You are truly role models to me, embodying a beautiful couple who live true to their ideals, deeply committed to their craft and community. While firmly rooted in tradition, you always maintain an open mind towards novelty and open arms towards newcomers. I am incredibly grateful to both of you for introducing me to the rich and sometimes overlooked Dutch culture. Thank you for your tireless energy in organizing musical events and for all the tunes you have taught me. Your wealth of cultural and ecological knowledge, has enriched my life in many ways. In my heart, I consider you as family.

Before turning toward my biological family, there is one final person who deserves acknowledgment: **Zou Chunmei**. **Mei**, I understand that public displays of affection may not be your preference, so I will keep it brief. Thank you for being part of my life for the past six years. Thank you for bearing with me when my physical body was present but my mind was still occupied with potatoes. I look forward to continuing our journey together, exploring life hand in hand.

Enfin, à **ma famille**, des grands-parents jusqu'aux cousins, sans oublier les oncles et tantes, à ceux qui m'ont vu grandir et qui me soutiennent depuis toujours, un très grand merci. Je tiens tout particulièrement à remercier mes parents, **Francis Clot** et **Joëlle Clot**. **Papa**, **maman**, merci pour tout ce que vous avez fait pour moi. Merci de m'avoir transmis votre amour pour la nature et votre fierté pour le travail bien accompli. Merci de m'avoir encouragé à suivre mes rêves scientifiques, même si cela m'a éloigné physiquement de vous. Merci d'avoir été là pour moi, tant sur le plan financier que moral, tout au long de mes études. Merci d'avoir toujours manifesté de la curiosité à l'égard de mon travail et de mes passions. Plus je grandis, plus je suis admiratif de ce que vous avez accompli et des valeurs que vous vous êtes efforcés de transmettre. Si

un jour je deviens moi aussi parent, j'espère être à la hauteur de votre exemple. Un grand merci également à ma sœur, **Maëlys Clot**, pour sa contribution à la fois technique et artistique à ce livre, ainsi que pour son soutien précieux, notamment durant les derniers mois de rédaction. **Maë**, merci d'avoir toujours été là pour moi quand j'en avais besoin. Merci de m'avoir convaincu de prendre du repos pour mieux repartir dans cette dernière ligne droite. Tes mots ont été véritablement libérateurs. Merci d'être présente aux côtés de papa et maman alors que je suis loin. À ma petite sœur qui a bien grandi désormais, sache que je suis fier d'être ton frère.

I would like to conclude those acknowledgements with the lyrics of "Lo companhon de l'Avairon", an Occitan dialog working song which particularly resonates with my PhD journey. This song beautifully portrays the determination to leave one's homeland to perfect his craft, as well as the invaluable support provided by fellow workers during times of hardship and homesickness. To all of you who supported me throughout this journey, I recognize you in this song and I thank you from the bottom of my heart. Although my Occitan is mediocre and my English leans toward scientific dryness, I have attempted to translate my two favourite verses, thereby ruining their original rhythm, but hopefully preserving their meaning.

- Here life is so tranquil, why would you leave your job behind?

- To travel from town to town, mastery of craft I shall find.

[...]

- Midst winter's bite and thunder's might, you'll certainly miss your homeland!

- Fear not my friend for in fellowship's warmth, I'll be welcomed with open hands.

I will now leave you on the next page with the original song in Occitan and encourage you to listen to Cocanha's version by scanning the QR code.

Thank you everyone, Merci à tous,

Corentin

Lo companhon de l'Avairon

- Aici la vida es plan tranquila, per de que quitas ton chantièr ? (bis)
- Me'n cal anar de vila en vila, me rafinar dins lo mestièr.

- Te cal d'argent per lo voiatge, tant de camin te costarà ! (bis)
- Vau a pè pòrti mon bagatge, e mon mestièr me noirirà.

- Tot coratjós, per tant que siagas, rencontraràs mai d'un dangièr. (bis)
- Ai un pistolet dins las bragas, e mon baston de codonhièr.

- Pendant l'ivèrn, pendent l'auratge, regretaràs ton Avairon ! (bis)
- Mos fraires de companhonatge, m'aculhiran dins lor foairon.

- E la Ròsa que tant t'agrada, s'anava trobar lo temps long ? (bis)
- A mon retorn s'es maridada, me'n tornarai pel camin grand.

- Te tendriam pas quand sièssem mila ! Adieu pensa a ton vièlh cloquièr, (bis)
- Adieu ! me'n vau de vila en vila, me rafinar dins lo mestièr.

About the author

Corentin Clot was born on the 6th of February 1994 in Decazeville, located in Aveyron, France. Growing up near the village of Sénergues, he spent his childhood on a small beef farm called "Le Souleilladou," where four generations lived under two roofs. Even at a young age, it became apparent to his parents that Corentin's true passion lay not in farming, as he appeared more interested in counting the cows than in tending to them. His interest in biology was fostered by this bucolic environment, and his curiosity for science was sparked by the television show "C'est pas sorcier," which he religiously followed after school.

In 2012, after graduating from high-school, he relocated to Toulouse to pursue an engineering degree in Agriculture. Over the course of this five-year program, his deep affection for botany and genetics took root. Through yearly internships, he explored various corners of the world, from an ice-cream farm in the French Alps to a potato breeding program in Michigan. His passion for plant genetics was strengthen by an Erasmus exchange at Wageningen University in 2015. Back in Toulouse and inspired by this experience, he was determined to pursue a master's degree from both universities simultaneously. He ultimately earned the two degrees in 2018, successfully defending a master thesis on potato genetics in Wageningen and another on stevia morphology in Toulouse.

After a short experience as research assistant in Toulouse, he returned to Wageningen University in 2019 to embark on a PhD journey in the department of Plant Breeding. Under the guidance of Dr Herman van Eck, he delved into the world of potato genetics focusing his research on reproductive biology. The findings of his PhD project are documented in this thesis.

Education Statement of the Graduate School
Experimental Plant Sciences

Issued to: Corentin C.R. Clot
Date: 28 August 2023
Group: Plant Breeding
University: Wageningen University

1) Start-Up Phase	date	cp
► **First presentation of your project**		
Identification of Self Compatibility - Consortium Progress Meeting	28 Jun 2019	1.5
► **Writing or rewriting a project proposal**		
A new method for potato breeding: the "Fixation-Restitution" approach	Feb 2019	2.0
► **MSc courses**		
Subtotal Start-Up Phase		3.5

2) Scientific Exposure	date	cp
► **EPS PhD days**		
EPS PhD Days 'Get2Gether' 2020, Soest (NL)	10-11 Feb 2020	0.6
EPS PhD Days 'Get2Gether' 2021, online	1-2 Feb 2021	0.4
EPS PhD Days 'Get2Gether' 2022, Soest (NL)	3-4 May 2022	0.6
► **EPS theme symposia**		
EPS theme 4 symposium 'Genome Biology', Wageningen (NL)	13 Dec 2019	0.3
EPS theme 4 symposium 'Genome Biology', online	11 Dec 2020	0.2
EPS theme 4 symposium 'Genome Biology & Gene Regulation', Amsterdam (NL)	6 Dec 2022	0.3
► **Lunteren Days and other national platforms**		
Annual Meeting 'Experimental Plant Sciences', Lunteren (NL)	8-9 Apr 2019	0.6
Annual Meeting 'Experimental Plant Sciences', online	12-13 Apr 2021	0.5
Annual Meeting 'Experimental Plant Sciences', Lunteren (NL)	11-12 Apr 2022	0.6
► **Seminars (series), workshops and symposia**		
Seminar: K. Schneeberger (MPI), Tetraploid Potato Genome Assembly and Analysis	2 Nov 2020	0.1
Seminar: P. Struik (WUR), Potato Research Futur : Food Security and Sustainability	17 May 2020	0.1
Seminar: A. Hancock (MPI), Adaptation to extreme environments in Arabidopsis thaliana	24 Feb 2023	0.1
Seminar: F. Budar (INRAE), Genetics of pollen abortion in Arabidopsis natural variants	24 Feb 2023	0.1
Seminar: KeyGene Seminar 'DNA analysis in asexually reproduced crops', Wageningen (NL)	27 Sep 2022	0.2
Seminar: Applied Plant Meiosis Webinar, online	10,17,24 Nov 2022	0.6
Workshop: Breeding for Diversity - Opportunities and Challenges, Wageningen (NL)	30 Oct 2019	0.3
Workshop: Breeding for Abiotic Stress, Study Group for Plant Breeding, Wageningen (NL)	3 Feb 2023	0.2
Workshop: Breeding for Vertical Farming, Study Group for Plant Breeding, Wageningen (NL)	31 Mar 2023	0.2
Symposium: Potato futures: Impact of hybrid varieties, online	30 Nov 2020	0.3
Symposium: EPS Mendel - 200 years Mini-Symposium, Wageningen (NL)	8 Jun 2022	0.2
Symposium: Mendel Day 2023, Wageningen (NL)	8 Mar 2023	0.2
► **Seminar plus**		
► **International symposia and congresses**		
Plant and Animal Genome Conference (PAG XXVIII), San Diego (US)	11-15 Jan 2020	1.5
Virtual International Solanacea Conference (SOL 2020), online	10-11 Nov 2020	0.6
21st EUCARPIA General Congress, online	23-26 Aug 2021	0.8
21st Triennial Conference of the European Association for Potato Research (EAPR), Kraków (PL)	4-7 Jul 2022	1.2
XVII International Solanaceae Conference (SOL 2022), Thessaloniki (GR)	1-5 Nov 2022	1.2
► **Presentations**		
Presentation: EPS theme 4 symposium "The origin and widspread occurence of *Sli* based SC in potato"	13 Dec 2019	1.0
Presentation: PAG XXVIII "The origin and widspread occurence of *Sli* based SC in potato"	13 Jan 2020	1.0
Presentation: 21st EUCARPIA General Congres "Investigating the genetic basis and cytological mechanisms of 2n pollen production in diploid potato"	26 Aug 2021	1.0
Presentation: SOL 2022 "The polygenic basis of 2n pollen production in diploid potato"	2 Nov 2022	1.0
Presentation: Final Consortium meeting "Fix-Res Breeding: Focus on restitution ingredients"	29 Mar 2023	1.0
Poster: EAPR 2022 "Marker imputation and linkage map construction using a skim sequenced diploid potato population"	4 Jul 2022	1.0
Poster: SOL 2022 "Reduced recombination in potato, causing sterility or uniform 2n gametes, is associated with mutant StMASH4 alleles"	2 Nov 2022	1.0
► **Interviews**		
► **Excursions**		
EPS PhD Company Visit to Averis, Valthermond (NL)	7 Jun 2019	0.2
EPS PhD Online Company event Bejo Zaden	14 Dec 2020	0.2
EPS PhD Company Visit to Solynta, Wageningen (NL)	24 Mar 2023	0.2
Subtotal Scientific Exposure		19.6

3) In-Depth Studies	date	cp
► **Advanced scientific courses & workshops**		
The Carpentries workshop 'Genomics Data', Wageningen (NL)	5-6 Feb 2019	0.6
SLU-WUR course 'Plant Breeding and Biotechnology', Wageningen (NL)	11-13 Jun 2019	1.6
Snakemake Workshop, online	13 May 2020	0.3
eScience Center workshop 'Software Carpentry with Python', online	6-9 Sep 2021	0.2
PE&RC course 'Generalized Linear Models', Wageningen (NL)	15-17 Jun 2022	0.9
► **Journal club**		
Participation in a Literature Discussion Group	2019-2022	0.5
► **Individual research training**		
Subtotal In-Depth Studies		4.1

4) Personal Development	date	cp
► **General skill training courses**		
Scientific Paper Writing workshop, Wageningen (NL)	24 Oct 2019	0.2
EPS Introduction Course, Wageningen (NL)	29 Oct 2019	0.3
WGS course 'Brain-friendly Working & Writing', Wageningen (NL)	19 May 2020	0.3
EPS Writing Support Group, online	Jan-Mar 2021	1.0
► **Organisation of meetings, PhD courses or outreach activities**		
Organisation of the Biodiversity and Genetic variation group meeting	2020-2021	1.5
► **Membership of EPS PhD Council**		
Subtotal Personal Development		3.3

5) Teaching & Supervision Duties	date	cp
► **Courses**		
► **Supervision of BSc/MSc students**		
Petros Lioupis (MSc thesis)	Sep 2019-Mar 2020	1.0
Joey Koopman (Msc thesis)	Aug 2020-Feb 2021	1.0
Dennis Klein (Msc thesis)	May-Dec 2021	1.0
Subtotal Teaching & Supervision Duties		3.0

TOTAL NUMBER OF CREDIT POINTS*	**33.5**

Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the Educational Committee of EPS with a minimum total of 30 ECTS credits.

** A credit represents a normative study load of 28 hours of study.*

The research described in this thesis was financially supported by the Top Consortium for Knowledge and Innovation (TKI) “A new method for potato breeding: the ‘Fixation-Restitution’ approach” (grant number: TU18075). The contribution of the following TKI partners to this project and this thesis was greatly appreciated: Averis Seeds B.V., Bejo Zaden B.V., Danespo A/S, Germicopa, Den Hartigh B.V., SaKa Pflanzenzucht GmbH & Co. KG, C. Meijer B.V., and Teagasc.

Financial support from Wageningen University for printing this thesis is gratefully acknowledged.

Cover design: Maëlys Clot
Layout design: Corentin Clot
Printing: ProefschriftMaken