WARTS WARS

THE RESISTANT POTATOES STRIKE BACK

CHARLOTTE PRODHOMME

Warts Wars

The resistant potatoes strike back

Charlotte Prodhomme

Thesis committee

Promotor

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

Co-promotors

Dr J.H. Vossen
Scientist, Plant Breeding
Wageningen University & Research

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

Other members

Prof. Dr M.G.M Aarts, Wageningen University & Research
Prof. Dr J. Śliwka, IHAR, Młochów, Poland
Prof. Dr G. van den Ackerveken, Utrecht University
Dr N. de Vetten, Dekker chrysanten B.V., Hensbroek, The Netherlands

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

Warts Wars

The resistant potatoes strike back

Charlotte Prodhomme

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 Friday 17 April 2020

at 1.30 p.m. in the Aula.

Charlotte Prodhomme

Warts Wars; The resistant potatoes strike back

PhD thesis, Wageningen University, Wageningen, the Netherlands (2020)

With references, with summary in English

DOI: 10.18174/512736

ISBN: 978-94-6395-293-4

Contents

Chapter 1	General introduction	9
Chapter 2	Distribution of P1(D1) wart disease resistance in potato germplasm and GWAS identification of haplotype-specific SNP markers	39
Chapter 3	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 <i>Sen3</i>	79
Chapter 4	<i>Sen1</i> fine-mapping and candidate gene analysis	125
Chapter 5	A Hitchhiker's Guide to the Potato Wart Disease Resistance Galaxy	171
Chapter 6	General discussion	233
	Summary	265
	Acknowledgements	269
	About the author	273
	Education statement	275

Chapter 1

General introduction

The United Nations declared 2020 as the International Year of Plant Health (IYPH) (1). This year event is the opportunity to increase the general awareness of the decisive importance of plant health to improve food security, fight poverty, eradicate hunger and preserve the environment. FAO estimated that up to 40% of the global food crops are lost due to plant pests every year (2). Plant health is therefore a critical lever that we need to exploit to reach the 60% food production increase which is necessary to feed the world by 2050. Plant pests are commonly controlled by the use of pesticides, which long term effects on the environment are unknown, and since nowadays people tend to regard them as being dangerous, their use must be limited. The most sustainable and durable way to combat pest attacks is therefore to deploy resistant varieties obtained by breeding, which is one of the different themes of the IYPH 2020 (1).

Potato is one of the major staple food crops worldwide and is a cornerstone for food security. It was recently estimated that 8 to 21% of the potato production is lost due to pests and pathogens (3). Up to 100% of a potato harvest can be lost due to the quarantine potato wart disease caused by *Synchytrium endobioticum*. Farmers cannot use fungicides to combat this fungus as none are completely efficient. The only weapon at hand is therefore the cultivation of resistant potato varieties. Resistance to the higher pathotypes of the wart disease is, however, scarce in the available commercial potato varieties. Thus, there is an urgency to breed more varieties holding a complete and durable resistance to a broad spectrum of pathotypes of *S. endobioticum*.

It is in this general context that this thesis was produced. The use of host resistance is a good illustration of how plant breeding can contribute to efficient solutions to the plant health concern. In this chapter, I summarize the current knowledge about *Synchytrium endobioticum* and the resistance of potato to this pathogen. Moreover, the scope of this thesis and the thesis outline are introduced.

***Synchytrium endobioticum*, the causal agent of potato wart disease**

Taxonomy, disease symptoms and host range

The causal agent of potato wart disease was described for the first time by Schilberszky (Budapest University) in 1896 (4). He named it *Chrysophylyctis endobiotica* and assigned it to the genus Chytridinea. Schilberszky described briefly both the summer and resting sporangia of the pathogen. Percival studied the pathogen a few years later (1910) and changed its taxonomic nomenclature: he assigned it to the *Synchytrium*

genus and gave it the name of *Synchytrium endobioticum* (5). Percival was also the first to contribute to the understanding and the knowledge about the cytology of the organism. It is now admitted that *Synchytrium endobioticum* (Schilberszky) Percival belongs to the order Chytridiales and to the phylum Chytridiomycota (6). The Chytridiomycota are true fungi that have a basal position in the fungal tree of life (7) and that can inhabit both aquatic and terrestrial habitats. Their specificity compared to other fungal phyla is the lack of mycelium and the production of flagellated zoospores. Since the late nineteen-nineties, Chytridiomycetes became infamous because of the massive amphibian decline caused by *Batrachochytrium dendrobatidis* and more recently also by *Batrachochytrium salamandrivorans* (8).

Figure 1. Potato wart disease symptoms

The susceptible variety Deodara was infected with pathotype P1(D1) of *S. endobioticum*. The warts in contact with light turn green. Warts are indicated with arrows. Picture credits: Bart van de Vossenbergh.

The potato wart disease name comes from the symptoms caused by *S. endobioticum*, which are the proliferation of meristematic tissues leading to the formation of galls, mainly on tubers, but also on stems, leaves and flowers (9,10) (Figure 1). Roots, however, never show any symptoms (11,12). Infection of the above-ground

parts is less frequent than below-ground parts (12), which makes the detection of the disease quite challenging, although infected plants can sometimes show a reduced vigour (13). The warts proliferating tissues function as sinks for nutrients used by the obligate biotrophic fungus (13). The warts, green coloured in early stages if exposed to light (depending on the chlorophyll content (10)) and turning black in later stages, can increase in size until entirely covering the tubers. After the harvest, they can continue their development during storage. Yield losses caused by *S. endobioticum* can reach 50 to 100%.

Potato (*Solanum tuberosum*) is the main host of *S. endobioticum*, but other species can be artificially inoculated. In 1916, Cotton inquired about the possibility of the survival and spread of *S. endobioticum* in the field thanks to alternative hosts such as *Solanum* weeds which would act as reservoirs. He observed that *S. nigrum* and *S. dulcanara*, which can often be found in potato fields, could be artificially infected with *S. endobioticum* (14). A few years later, Martin could infect several other *Solanum* species as well as tomato (*Lycopersicon esculentum*) (15). In the late 70s, Hampson tested many tomato cultivars which turned out to be all susceptible to *S. endobioticum*, although no galls were observed as in potato (16). He also showed that species of *Capsicastrum*, *Datura*, *Physalis* and *Schizanthus* could be infected by *S. endobioticum* (17).

***Synchytrium endobioticum* life cycle**

In 1910, Percival was the first to make a real contribution to the understanding of the life cycle and of the cytology of *S. endobioticum*, although several papers describing the different spores or the infection conditions were published previously (5). A few years later, Curtis produced a very detailed description of the full life cycle of the pathogen which is still the main reference today (18). To describe the life cycle, I will start with the haploid (*n*) zoospores released by winter sporangia in spring, at the beginning of the potato growing season (Figure 2). These zoospores are mobile for a short period (20 to 40 minutes, longer if temperatures are < 15°C) (18), their movements being possible thanks to a flagellum and to the presence of water in the soil (12). Zoospores are probably attracted by the host in response to a chemical stimulus (chemotaxis) (19,20). When they reach young potato tissues, they encyst, enter the host cell and leave the cyst wall outside (21). One host cell can be infected by one or more zoospores (21). The zoospore, from now called a prosorus (*n*), travels into the cell and starts its expansion. The host cell expands too. When the prosorus occupies the greater part of the host cell, it discharges its content into the latter through a small pore. The prosorus nucleus divides and the protoplasm segments: five to six successive mitosis combined with the segmentation of the protoplasm allow the formation of 3-7 thin-walled sporangia (*n*) which compose the sorus (*n*). Zoospores (*n*) are formed by further nuclear divisions in the sporangia. The tumours are formed by the division of the neighbouring cells of the host. Finally, the host cell breaks, the mature sporangia are liberated and can release 200-300 motile zoospores (*n*) (21). There are then two different possibilities for these released cells. i) They can act as zoospores and the life cycle described previously is repeated, leading to secondary

infections. It takes 10 to 12 days to get a new generation of soral sporangia (12). ii) They can also act as gametes and fuse in pairs to form an uninucleate biflagellate zygote ($2n$) (18). The zygote enters an epidermal host cell the same way haploid zoospores do, but it will lead to the formation of resting sporangia (winter sporangia). The zygote travels into the host cytoplasm and starts its expansion. The host cell divides which brings the resting sporangia deeper in the layer of cells. The formation of a thick cell wall starts. The host cell's dead contents form an epispore on the sporangium.

Figure 2. *Synchytrium endobioticum* life cycle

The life cycle is divided into two parts: the formation of summer sporangia (haplophase) and, in stressing conditions (drought, end of the growing season), the formation of winter/resting sporangia (haplodiplophase). **1:** Motile zoospores (n). **2:** A zoospore enters the host cell, it is now called a prosorus (n). **3:** The prosorus travels in the cell and starts its expansion, the host cell expands too. **4:** When the prosorus occupies the greater part of the host cell, it discharges its content into the cytoplasm. **5:** The prosorus nucleus divides and the protoplasm segments. The sorus (n) contains 3 to 7 sporangia (n) in which zoospores (n) are formed by further nuclear divisions. **6:** The host cell breaks and releases mature sporangia. In turn, the soral envelope breaks releasing numerous motile zoospores

(*n*). **7**: Under stressing conditions, the zoospores function as gametes and fuse in pairs to form a zygote ($2n$). **8**: The zygote enters the host cell and travels in its cytoplasm. **9**: The young sporangium ($2n$) starts forming its true membrane while the host cell divides, placing the sporangium deeper in the layer of cells. **10**: The host cell dies and its dead content forms an epispore on the sporangium. **11**: The mature sporangia are released into the soil. An extra inner wall forms which is where sporogenesis takes place. **12**: Discharge of the motile zoospores (*n*). Drawings made by (18) used and modified with permission.

In the literature, as far as I know, there is no clear information about when exactly the meiosis takes place. Curtis (18) states that the meiosis occurs just before the release of mature zoospores into the soil, whereas the cycle described by Karling (22) seems to show that meiosis occurs during the maturation of the resting sporangia. In the released mature resting sporangia, an extra inner wall is formed. In this new area takes place sporogenesis and then, mature zoospores (*n*), after a dormant phase (which duration vary in function of environmental conditions), are released into the soil and can then infect host cells. Resting sporangia can remain viable in the soil for more than forty years (23).

Curtis investigated the life cycle of *S. endobioticum* using susceptible potato varieties (18). A couple of years later, Cartwright compared the infection of *S. endobioticum* on a susceptible and a resistant variety (24). She observed zoospores fixed on the surface of the resistant tubers and observed that they were also able to penetrate the cuticle and the outer cell walls. Three days after inoculation, she compared the aspect of the pathogen inside the cells of the resistant and susceptible varieties and observed that its size was reduced and the nucleus was smaller in the resistant variety. She could never observe the later stages of the life cycle, suggesting recognition occurs shortly after entering the host cell.

Introduction of *S. endobioticum* in Europe

S. endobioticum is native of the Andes mountains of South America where it co-evolved with its main host *Solanum tuberosum* (25). At the end of the XIXth century, *S. endobioticum* spread first to Europe and then to North America (Figure 3). The introduction of the pathogen in Europe probably happened because of the use of breeding material from South America in the 1840s (19).

S. endobioticum has a very limited capacity for natural dispersion: at the first stage of its life cycle, zoospores are released from winter or summer sporangia and start movements during approximatively 20 to 40 minutes. During this brief swarm period, the maximum spread of zoospores is about 9 inches (23 cm) or 30 inches (76 cm) if there is a slope (12). Therefore, its major mean of dissemination is the trading of

Figure 3. Distribution of *S. endobioticum* in 1993 and presumed dissemination scheme of the pathogen

S. endobioticum was first introduced from the Andes region to the United Kingdom in the second part of the XIXth century. From the UK, it spread to several European countries, to North America and probably to other countries worldwide such as South Africa and New Zealand. ^a Found but not established. ^b not confirmed. () Inactive. This scheme was adapted from (26) with permission.

infected tubers (or tubers with soil holding resting sporangia) between countries. There are other ways of dispersion such as commercial exchange, irrigation, vehicles (27), infested soil adhering tools, feeding animals (10) and even water flood (28).

Schilberszky first described *S. endobioticum* in what nowadays is Slovakia, but there is no evidence of the presence of the pathogen in this area at that time. Although *S. endobioticum* was first described in 1896, there are clear evidences that it was already found in the United Kingdom in 1876 (10). During the following decades, the fungus was introduced in most of the countries where potato was grown (29) (Figures 3 and 4). These new findings probably originated from the UK and were not caused by new introductions from South America. Its first observation in Germany was in 1908 and shortly after, in 1915, it was observed for the first time in The Netherlands. In Poland, *S. endobioticum* was first observed in 1925 and it was later introduced in the former USSR in 1938 (30). Nowadays, the pathogen is present on all continents except for

Antarctica and in more than twenty countries on the European continent (Figure 4), but often restricted to certain areas thanks to strict quarantine measures. These measures achieved the eradication of *S. endobioticum* in several countries such as Austria, France, Hungary, Latvia, Portugal, Slovenia and Northern Ireland (31).

Figure 4. Distribution of *S. endobioticum* in Europe in 2019

Orange countries are countries in which *S. endobioticum* is present. Yellow dots = present, pink dot in Norway = transient. This map was generated by (31) (updated 2019.09.02).

Emergence of new pathotypes of *S. endobioticum* and pathotyping method

Before 1941, only one pathotype of *S. endobioticum*, that we call pathotype 1(D1), was found in Europe. Probably due to the rising deployment of wart resistant varieties, new pathotypes started to emerge in the 1940s. In 1941, in Germany and in Czech Republic, warts were found for the first time on resistant cultivars. The emergence of several new pathotypes followed, mostly in Germany and Czech Republic, but also in Eastern Europe (former USSR) (30). To distinguish the different isolates and to group them into pathotypes, bioassays are performed on a differential set of varieties with contrasting resistance patterns (see next §). At that time, each country and laboratory had its own set of differential varieties and there was also no consensus for the naming of the pathotypes, which made comparisons between countries impossible. Some countries used a naming system based on the first letter of the locality where the new pathotype was first found, some used an Arabic numbering system. Baayen, in 2006,

proposed an harmonization of this naming system which combines the pathotype number and the locality first letter (i.e. pathotype 1 (D1)) (30). In this thesis, I use mainly the numbering system alone for simplification. Baayen also made a detailed review of the differential sets used by the different countries and of the new pathotypes that emerged worldwide (30) (Figure 5). Some pathotypes are now extinct but new ones continue to emerge. In the past ten years, despite the strict quarantine measures, pathotype 38 (Nevşehir) (32) was identified in Turkey and a similar pathotype was identified in Georgia (33), and pathotype 39 (P1) was identified in Poland (34).

Figure 5. Emergence of the main *S. endobioticum* pathotypes in Europe

References: a: (10), b: (30), c: (35), d: (32), e: (34), f: (33).

To date, there are no molecular tools to differentiate the different pathotypes of *S. endobioticum*. The European and Mediterranean Plant Protection Organization (EPPO) developed diagnostic protocols for *S. endobioticum* in which a differential set initially composed of 11 varieties was used to distinguish pathotypes 1, 2, 6, 8 and 18, the most common pathotypes in Europe (36) (Table 1). The EPPO set is now used by the majority of the European countries. Some varieties with the same resistance pattern are included, such as Producent and Combi or Karolin, Ulme and Belita. Based on current information, these varieties appear to be exchangeable. However, the genetic basis for resistance in these varieties is unknown and might be different. Another weakness of this set is that there is only one variety which allows to distinguish between pathotypes 2 and 6 (Saphir), one variety to distinguish between pathotypes 6 and 8 (Delcora), and one variety to distinguish between pathotypes 8 and 18 (Miriam). Furthermore, variation in the *S. endobioticum* populations may be

underexposed because there may be many more resistances (and hence virulences) around that are currently not detected. So, the current differential set is only poorly characterised and its composition should be re-evaluated using knowledge about the *R* genes composition of the varieties and the corresponding *S. endobioticum* effector they specifically recognise. Adding or removing varieties from this set has an important impact on the pathogen variation that can be recognised. To retain testing efficiency and throughput, the set might need to be extended only for certain countries or specific locations which have a higher diversity of pathotypes.

Table 1. EPPO differential set of varieties to distinguish *S. endobioticum* pathotypes

Potato variety	P1	P2	P6	P8	P18	EPPO set according to (36). This set was later reduced (37). The main changes are that Talent (R to P1, middle R to P2 and P6, S to P18) and Ikar and Gawin (both R to P1, P2, P6, P18) were added. Combi, Delcora, Miriam and Ulme were removed. Because of the removal of Delcora and Miriam, there is no distinction between P6 and P8, and P8 and P18 anymore.
Deodara	S	S	S	S	S	
Tomensa	S	S	S	S	S	
Eersteling	S	S	S	S	S	
Producent	R	S	S	S	S	
Combi	R	S	S	S	S	
Saphir	R	S	R	R	R	
Delcora	R	R	R	S	S	
Miriam	R	R	R	R	S	
Karolin	R	R	R	R	R	
Ulme	R	R	R	R	R	
Belita	R	R	R	R	R	

Measures for potato wart disease control

Quite a wide array of chemicals have been tested to control *S. endobioticum* (19). The main conclusion of these studies is that some chemicals were able to reduce the infection by *S. endobioticum* but none could totally eradicate the pathogen. Moreover, many of these chemical treatments have negative impacts on the environment and on the soil health. Because of this lack of efficient chemical control, the long persistence of the resting spores in the soil and the important yield losses, *S. endobioticum* was added to the A2 list of pests by the EPPO (38). When *S. endobioticum* is identified, through direct observations, bioassays, or molecular identification (conventional PCR or real-time PCR) (37), the field in which it is present is “scheduled”. A buffer zone is set to delimit the contaminated field and avoid contamination of the surrounding areas. All the potato tubers and plant waste are destroyed to eliminate *S. endobioticum*. Potato production in scheduled fields is prohibited and only resistant varieties can be grown in the buffer zones (39). After a field has been scheduled, it can be partially or completely descheduled if specific conditions are met. A field can be completely

descheduled minimum 20 years after the first identification of *S. endobioticum*. The field should have been used for cultivation during this period and a strict soil sampling, examination and bioassay procedure is carried out. A field can also be descheduled without a soil test, minimum 50 years after its scheduling, only if no susceptible crops have been cultivated on it. A partial descheduling, which will only allow the cultivation of resistant varieties on the plot, can be done at least 10 years after the scheduling. As for a complete descheduling, the field should have been used for cultivation and soil sampling and testing should be performed to determine the number of viable spores per gram of soil (40). These strict measures were successful in eradicating *S. endobioticum* in several countries or contain it to limited areas.

***S. endobioticum* genome and identification of the first Avr gene**

Recently, the mitochondrial genomes of several pathotypes of *S. endobioticum* were sequenced and compared (41). The authors identified several mitochondrial haplotype groups and observed isolates from the same pathotype associated with the different haplotype groups. This observation suggests that pathotypes have emerged independently. A striking finding was that *S. endobioticum* isolates were not composed of a single genotype but of a community of different genotypes, stable over time. These communities can explain the fast emergence of new pathotypes when they multiply on partially resistant varieties. Also recently, the nuclear genome of several pathotypes of *S. endobioticum* was sequenced and comparative genomics were performed to learn more about its obligate biotrophic lifestyle (20). The authors observed that *S. endobioticum* lacks genes encoding indispensable enzymes for the purine and pyrimidine pathways. Interestingly, *S. microbalum*, which is not an obligate biotroph but a saprobic *Synchytrium* species, holds these genes. *S. endobioticum* probably relies on its host for its purine and pyrimidine bases supply, which can explain its obligate biotrophic lifestyle. Another interesting feature of *S. endobioticum* genome is its high repeat content which could be linked to its pathogenic lifestyle. Moreover, comparative genomics of the different pathotypes allowed the identification of the first *S. endobioticum* effector protein AvrSen1 (42). AvrSen1 triggers Hypersensitive Responses (HR) in the cytoplasm of plants containing the *Sen1* gene. Several variants of the *AvrSen1* gene were identified in the other pathotypes but none of them was recognised by *Sen1*. These findings strengthen the hypothesis that the different resistance patterns of varieties to different pathotypes is due to the specific recognition of *S. endobioticum* Avr proteins by potato R proteins.

Potato resistance to *S. endobioticum*

Potato production and impact of pathogen attacks

Potato (*Solanum tuberosum*) has a key role in global food security and in the fight against poverty. It is one of the major staple food crops with wheat, rice, maize and soybean. With the current estimated population growth, world food production should increase by not less than 60% to feed the world population in 2050 (2). However, the actual yield improvement rates of maize, rice, wheat and soybean are insufficient to reach this goal (43). Potato is the most energy productive crop, is a good source of proteins and calcium (44) and has a higher water use efficiency than cereals (45). Moreover, the potato production of developing countries has highly increased in the last decades and is now more important than in developed countries. The importance of potato to meet tomorrow's food demand is therefore growing. Yield increase can be obtained by a combination of optimized cultural practices and high yielding varieties. Another important point for improvement is to reduce yield losses by combating the negative impact of (a)biotic stresses. For potato, the pests and pathogens induced crop losses have been estimated to range between 8% and 21%, depending on the production region (3). Considerable amounts of pesticide treatments are used for pests and pathogens that can be controlled this way. For other diseases such as bacterial wilt, soft rot, or wart disease, no chemical control is possible and reduction of yield losses can only be achieved by deploying resistant varieties and/or by applying strict control measures.

The plants immune system

Plants depend on an innate immune system to defend themselves against biotic stresses. The plant immune system deploys surface and intracellular receptors that recognize pathogen factors. This immune system is a result from intricate coevolution of the pathogen with its host and is framed by the zigzag model (46).

The first defence layer of this model is called PAMP-triggered immunity (PTI). PTI relies on the recognition of conserved pathogen-associated molecular patterns (PAMPs), microbe-associated molecular patterns (MAMPs), or damage-associated molecular patterns (DAMPs) released from injured cells, by transmembrane pattern recognition receptors (PRRs) such as receptor-like kinases (RLKs) and receptor-like proteins (RLPs). PAMPs/MAMPs are defined as being conserved molecules required for the pathogen/microbe survival, such as for instance bacterial lipopolysaccharides, flagellin and harpin, fungal chitin or oomycetes transglutaminase and glucans (47). Effector

proteins of successful pathogens are capable of suppressing PTI signalling, resulting in effector-triggered susceptibility (ETS). As a counter-attack, plants evolved intracellular R proteins able to directly or indirectly recognise specific effectors. This second defence layer of the zigzag model is the effector-triggered immunity (ETI) and was early identified by Flor in the gene-for-gene concept (48). In ETI, the pathogen secretes effector proteins (Avr proteins) into the host cell (or they can be translocated into the host cell). Intracellular R proteins, for the majority, are receptors proteins from the nucleotide-binding domain, leucine rich containing (NLR) protein family. Upon recognition of their matching effector, NLR proteins trigger hypersensitive responses (HR), a form of programmed cell death. There are two major NLR protein families which are distinguishable by their N- terminal domain: CNLs hold a coiled-coil (CC) N-terminal domain whereas TNLs hold a Toll-interleukin 1 receptor (TIR) N- terminal domain.

The identification of the first *S. endobioticum* effector protein AvrSen1 revealed the involvement of ETI in potato wart disease resistance. The recognition of the AvrSen1 by Sen1 happens in the cytoplasm (42) which suggests that *Sen1* is likely an NLR protein encoding gene. *S. endobioticum* effectors might be involved in ETS as well with the suppression of PTI. This hypothesis is strengthened by the observation by (24) that the pathogen is also able to penetrate the cells of resistant varieties.

The main differences between PTI and ETI are therefore the pathogen and host components involved: roughly, PTI involves conserved pathogen molecules recognized by extracellular receptors, leading to non-specific basal resistance, whereas ETI involves the specific recognition by intracellular receptors of species- or race- or strain-specific effectors. This theoretical dichotomy between the PTI and ETI branches is however not as black and white in practice and should better be perceived as multiple shades of a grey continuum (49).

RLK and NLR resistance proteins are often leading to a qualitative type of resistance and were therefore broadly targeted in mapping and cloning studies and widely used in breeding programs, as their Mendelian segregation pattern makes them easy to study and to introgress in breeding material. It was long thought that quantitative resistance is more durable than qualitative resistance. Although, the mapping and fine-mapping of genes underlying quantitative resistance showed that the same type of innate immune receptors as involved in qualitative resistance, such as RLKs and NLRs, can bring partial resistance. For instance, many Quantitative Trait Loci (QTLs) have

been mapped to *R* gene clusters and (50) made the hypothesis that weak forms of *R* genes or defeated *R* genes could bring a quantitative type of resistance. Other hypothesis have been made about the genes underlying resistance QTLs such as genes controlling morphological and developmental traits, mutations or allelic diversity of genes involved in basal defence (i.e. RLKs), genes involved in the chemical host/pathogen warfare or genes involved in the defence signalling pathways (50). The recent finding that soil bound pathogen populations like *S. endobioticum* consist of multiple genotypes may also contribute to our understanding of partial resistance, as only parts of the pathogen population can be recognised.

***R* gene mapping and cloning strategies in potato**

In potato, more than 20 *R* genes have been cloned so far, and many more have been mapped, the majority of them being from the NLR family (51). *R* gene mapping and/or cloning studies usually involve the same major steps as described by (52): 1) choosing the parents to make a segregating population, usually one resistant parent, holding one *R* gene in one copy or several resistance QTLs, is crossed with a susceptible parent; 2) phenotyping of the descendants and estimation of the segregation ratio; 3) genotyping of the segregating population and creation of a genetic map; 4) identification of markers flanking the *R* gene(s) or QTL(s) to place them on the map. Early mapping studies often made use of diploid populations to simplify the mapping process. Depending on the segregation observed in the descendants, several strategies can be followed to genotype the population and identify markers linked with the *R* gene(s) segregating. When only one major *R* gene is segregating in a 1:1 Mendelian ratio, the preferred strategy is to perform a Bulk Segregant Analysis (BSA), which consists of pooling resistant and susceptible progeny, genotyping the two contrasting bulks and identifying markers which have contrasting allele frequencies between the two bulks (53,54). The main advantage of BSA is that genotyping efforts can be drastically reduced. When resistance is segregating in a quantitative way in the mapping population, the genotyping of all individuals with a sufficient marker density and the building of a genetic map are required. Moreover, complex QTL mapping methods involving multiple regressions, mixed models or Bayesian statistics may be required.

The first mapping studies involved the genotyping of the mapping populations with Amplified Fragment Length Polymorphism (AFLP) markers and/or Restriction Fragment Length Polymorphism (RFLP) markers. Many of the *Phytophthora infestans Rpi* genes were mapped using BSA combined with RFLP or AFLP markers. This is the case for

example of *R2* (55), the *R3* locus (56), *Rpi-mcq1.1* (57), *Rpi-ber1* (58), *R10* and *R11* (59). Later, the NBS profiling approach was developed (60). This approach is based on the design of PCR-primers targeting the NBS domain of NLR genes. NBS profiling was used for instance to map *R8* (61) and was combined with BSA to map *Rpi-vnt1* (62), *Rpi-ver1* (63), *Rpi-avl1* (64), *Rpi-cap1* and *Rpi-qum1* (65). In mapping studies for other traits, RFLP and AFLP markers could be replaced by the use of SNP arrays such as the 8.3K SolCap SNP array (66,67) or the 20K SolSTW SNP array (68). For *R* genes mapping, it is however not the preferred option because of ascertainment bias. An illustration of the ascertainment bias that the use of SNP arrays can bring for resistance mapping studies is the study of (69). A tetraploid population segregating for several pathotypes of the wart disease was genotyped with the SolCap array. Several QTLs could be mapped but when tested in a broader variety panel, the markers linked with resistance identified were not diagnostic, likely due to a lack of haplotype specificity of the SolCap markers or to a linkage with the resistance loci in repulsion phase.

With the increasing accessibility of sequencing techniques, and with the release of the potato reference genome DM1-3 516 R44 (DM) (70), the recent *R* gene mapping and/or cloning studies involve the use of genotyping-by-sequencing for the genotyping of mapping populations. Whole Genome Sequencing data or genome reduction strategies can be used to identify Single-Nucleotide Polymorphisms (SNPs) through a read alignment step to the reference genome. RenSeq, a NLR gene enrichment and sequencing method designed by (71), is currently the preferred genotyping-by-sequencing method for the mapping and cloning of *R* genes from the NLR family. RenSeq, combined with BSA, has been used to map and/or clone several NLR genes such as *Rpi-ber2* and *Rpi-rzc1* (71), *Rpi-ver1* (72), *Rpi-amr3* (73), *Rysto* (74) and *H2* (75). Gene enrichment strategies can also be designed to target other gene families such as for instance the RLP/RLK genes targeted in the RLP/KSeq method designed by (76).

The cloning of *R* genes requires a good assembly of the fine-mapped resistant haplotype. Traditionally, the cloning of *R* genes involved the construction of a Bacterial Artificial Chromosome (BAC) library and a long and labour-intensive process of BAC landing or BAC walking. The release of the potato reference genome did not replace the BAC libraries as DM is not known to be resistant to any disease and the assembly of *R* gene clusters is extremely difficult. Recently, the SMRT-RenSeq method allowed to correctly assemble the targeted *R* genes thanks to the use of PacBio long reads

(73,74). The ONT MinION sequencing device (Nanopore) has also shown to be an alternative to PacBio to correctly assemble *R* gene clusters (77).

As mentioned previously, traditional *R* gene mapping in potato involves the generation of bi-parental segregating populations. Alternatives are Genome-Wide Association Studies (GWAS) which allow to include the genetic diversity of a wider panel of genotypes and to reach a better mapping resolution thanks to a higher number of recombination events. To date, the few GWAS performed to identify resistance loci in potato were successful in identifying QTLs involved in resistance to *Verticillium* wilt (78) and late blight (79–81).

Screening for potato wart disease resistance

As no chemicals are available to fight *S. endobioticum*, the only efficient option is to breed for wart resistant potato varieties. So far, very few diagnostic molecular markers for potato wart disease resistance loci have been published. For this reason, the main breeding tool is to screen second- and third-year clones for wart disease resistance in tuber-based bioassays. The two major laboratory bioassays are the Spieckermann (82) and the Glynne-Lemmerzähl assays (83,84). The Spieckermann assay, mainly used in the Netherlands, involves the inoculation of tuber eye-pieces with a compost containing winter sporangia harvested from infected tubers. The Glynne-Lemmerzähl assay, mainly used in Germany and in Poland, uses fresh warts for the inoculation of emerging sprouts on tubers or eye-pieces. Therefore, the main difference between both methods is that the first infection in the Spieckermann assay is made by winter spores whereas it is made by summer spores in the Glynne-Lemmerzähl assay. Overall, both methods give correlated results, although the Spieckermann method is known for causing more escapes, but being closer to what happens in the field through the use of winter spores. Recently, a new bioassay involving the inoculation of lateral shoots instead of tuber sprouts was developed (85). The results matched with the Spieckermann and Glynne-Lemmerzähl assays but has the advantage of not requiring tubers. A sufficient number of repeats is however needed to avoid escapes.

Potato wart disease resistance sources

Resistance to the pathotype 1 of the wart disease can be traced back in early potato breeding ancestors. Resistance to *S. endobioticum* is probably even the first disease resistance which was intentionally bred for in early breeding programs. One of the oldest varieties in which pathotype 1 resistance was reported is Myatt's Ashleaf (86) which was released in 1804 and was substantially used in British and Dutch breeding

programs (87). According to Ross, one of the old sources of pathotype 1 resistance is the old British variety Flourball released in the late 1800's (88). Pathotype 1 resistance has also been observed in the old British variety Snowdrop (86) released before 1881 (87). The pedigree of Flourball and Snowdrop is unknown, so it is not possible to determine if the pathotype 1 resistance sources used in the early breeding programs were multiple or unique. In Germany, 30% of the pathotype 1 resistant varieties of the beginning of the 20th century would come from the variety Jubel (89). Although the pedigree of Jubel is also unknown, it is likely that its resistance source is different from the British varieties. These pathotype 1 resistant varieties turned out to be susceptible to the new pathotypes that emerged in the early 1940's.

With the emergence of these new pathotypes, it was important to determine if resistance could be found in the breeding material. The German breeding clone BRA9089 was early identified as being a donor of resistance to several pathotypes to several German varieties such as Galina, Judika, Oda, Ora, Tondra and Xenia (88,90), although the pedigree of Judika and Xenia do not seem to contain BRA9089 (87). Another source of resistance to the higher pathotypes was identified in the German variety Saphir which probably inherited its resistance from the wild species *S. acaule* through the breeding clone MPI 44.1016/24 (88). Ross also referred to another possible resistance source coming from *S. tuberosum* ssp. *andigena* through the breeding clone CPC1673 (88) which was used to introgress the *G. rostochiensis* *R* gene *H1* in the breeding genepool (91,92). Resistance to the higher pathotypes was therefore already present in the breeding genepool when these new pathotypes emerged, which suggest that these higher pathotypes resistances were introgressed by chance as linkage drag from introgressing other diseases resistances, as observed for instance for the resistance coming from CPC1673.

Wild *Solanum* species are also an important reservoir for resistance genes for many diseases. Numerous *R* genes were identified in wild species such as *S. demissum* and *S. bulbocastanum* for late blight resistance for instance. As resistance to *S. endobioticum* was already present in the breeding genepool, no targeted efforts have been made to introgress new wart resistance genes from wild species. Several screening experiments were however performed to assess the resistance to several wild species and landraces to wart disease. Pathotype 1 resistance has been assessed in a panel of 52 landraces from *S. phureja*, *S. stenotomum*, *S. tuberosum* ssp. *andigenum* and *S. tuberosum* ssp. *tuberosum* (93). Pathotype 1 resistance was observed in every taxonomic group and from every geographical origin. A wide

collection of tuber-bearing *Solanum* species has also been phenotyped for several traits by (94), including resistance to pathotypes 1, 2 and 6 of *S. endobioticum*. Strong resistance to pathotype 1 was found in more than 18 different *Solanum* species including *S. acaule*, *S. boliviense*, *S. chacoense*, *S. gourlayi*, *S. microdontum*, *S. multidissectum*, *S. polyadenum*, *S. raphanifolium*, *S. spegazzinii*, *S. stoloniferum*, *S. sucrense*, *S. tuberosum* ssp. *andigena* and *S. vernei*. Resistance to pathotype 1 is therefore very abundant in wild species which can explain its presence very early in breeding material. In the same panel, resistance to pathotype 2 was observed in *S. sucrense*, *S. tuberosum* ssp. *andigena* and *S. vernei* and resistance to pathotype 6 was observed in *S. tarijense* and *S. tuberosum* ssp. *andigena*. Resistance to the higher pathotypes was also observed by (90) in *S. acaule*, *S. demissum*, *S. vernei*, *S. simplicifolium* and *S. parodii* and by (95) in *S. andigenum*, *S. chaucha*, *S. macmillanii*, *S. phureja* and *S. stenotomum*. Recently, the *Sen2* gene which gives a strong resistance to eight different pathotypes has been identified in a complex diploid hybrid with origins from *S. tuberosum*, *S. chacoense*, *S. yungasense*, *S. gourlayi*, *S. acaule*, *S. demissum*, *S. verrucosum*, *S. microdontum* and *S. phureja* (96). Wild *Solanum* species represent therefore a rich reservoir of resistance to introduce new resistance sources in the breeding genepool.

Genetic analyses of potato wart disease resistance

Potato wart disease resistance was one of the first traits in plants studied for Mendelian inheritance (86), but no consistent genetic model was built in the early studies due to inconsistent phenotyping results, the use of small populations and the use of diploid genetic models in tetraploid material. Several hypotheses were made involving the existence of one or two genes that could give resistance to wart disease, that resistance was dominant over susceptibility and that complementary and inhibitory factors were involved (86,97,98).

The progress made in polyploid genetics made the mapping of wart resistance genes feasible. The first wart disease resistance gene mapped was *Sen1*, a dominant gene involved in pathotype 1 resistance, which is located on the north arm of chromosome 11 where also numerous other resistance genes reside (99). Another dominant gene involved in pathotype 1 resistance, *Sen1-4*, was mapped in a diploid population on the south arm of chromosome 4 (100). These two mapping studies showed that resistance to the pathotype 1 can be given by a single locus and that there are at least two pathotype 1 resistance sources in the potato breeding genepool. In 2011, the first study was performed to map resistance to the higher pathotypes. The authors used

the two tetraploid half-sib populations BNA1 and SaKa1 segregating for pathotypes 1, 2, 6 and 18 resistance. They identified a QTL conferring partial resistance to pathotype 1 on chromosome 11 where *Sen1* was previously identified, a QTL on the south arm of chromosome 9 bringing specific resistance to the pathotype 18 and a pathotypes 2, 6 and 18 resistance QTL on the south arm of chromosome 1 (101). This study showed that not only qualitative resistance loci are involved in wart resistance, and that different genes are involved in pathotype 1 and the higher pathotypes resistance. This last observation is in agreement with a previous study in which resistance to pathotypes 2 and 6 did not co-segregate with *Sen1* (102).

In 2013, AFLP and SSR markers were used to map pathotypes 1, 2, 6 and 18 resistance QTLs in the Saturna x Panda tetraploid population (103). They identified a major QTL for pathotype 1 resistance on chromosome 11 near the *Sen1* region, and other QTLs involved in one or several pathotypes on chromosomes 1, 2, 6, 7, 8, 10 and 11. All these QTLs showed a small effect on resistance. In 2015, 195 SSR markers and the 8.3K SolCAP SNP array (66,67) were used to map pathotypes 1, 2, 6 and 18 resistance QTLs in the BNA2 tetraploid population (69). A major effect locus involved in pathotypes 1, 2, 6 and 18 resistance was identified in the *Sen1* region of chromosome 11. Six other loci have been identified with smaller effects on wart resistance. During the redaction of this thesis, two new studies were released in which major effect resistance loci were mapped. In a dihaploid population with Karolin as resistant parent (104), the same locus as in the BNA2 population (69), on the north arm of chromosome 11, was identified again and fine-mapped. In another diploid population segregating for resistance to several pathotypes, a new locus, *Sen2*, was identified on chromosome 11 (96), 32 Mb south from *Sen1*. This *Sen2* locus, identified in a complex hybrid with multiple wild species origins, is involved in resistance to pathotypes 1(D1), 2(G1), 6(O1), 8(F1), 18(T1), 2(Ch1), 3(M1) and 39(P1).

In the majority of the cited potato wart disease mapping studies, the diagnostic power of the designed markers was not assessed in a broader panel of varieties. The *Sen1* associated markers NI25 (99), GP125 (101) and StI046 (101) were screened in an independent panel of German and Polish varieties but the results were not elaborately reported (104). NI25 and StI046 were also screened in a panel composed only of varieties from Russia and neighbouring countries (105). Regarding loci conferring resistance to the higher pathotypes, only the markers designed by (104), linked to pathotypes 6 and 18 resistance, were validated in a panel of 38 susceptible and resistant varieties. There is therefore a great demand for diagnostic markers to identify

unique wart disease resistance genes and QTLs to facilitate the breeding of resistant varieties in Marker Assisted Selection (MAS).

Scope and outline of the thesis

The primary aim of this thesis was to identify the loci involved in resistance to the most frequent pathotypes of *S. endobioticum* in different potato genotypes, in order to produce an inventory of resources for wart disease resistance breeding. This thesis explores the type of genes involved in resistance to wart disease and the effect of combining several genes to provide full resistance to the higher pathotypes. This thesis also addresses the development of a new set of tools for the genetic analysis of a trait of interest and for the identification of haplotype specific SNPs.

In **Chapter 2**, we analysed the distribution of the pathotype 1 resistance in a variety panel representative of the commercial genepool. Pathotype 1 resistance is widely spread in breeding material which is due to successful breeding programs from the beginning of the 20th century. A Genome-Wide Association Study (GWAS) was performed in the variety panel to identify haplotype specific SNP markers linked with pathotype 1 resistance. The markers identified with the GWAS were located on the north arm of chromosome 11, in the region where the *Sen1* locus was previously identified. We validated the significantly associated markers in two tetraploid populations segregating for pathotype 1 resistance and all were linked to pathotype 1 resistance. However, no marker with a complete diagnostic value was identified, due to the limited number of markers on the chip representing the region of the *R* gene clusters C76 and C77 on the 20K SNP array. *Sen1* was identified in approximately 70% of the resistant varieties of the panel and is, therefore, the main source of pathotype 1 resistance in the breeding genepool. Due to a lack of identity-by-descent, it was not possible to trace back *Sen1* to a common founder. Less frequent alleles or loci involved in pathotype 1 resistance are present in the breeding genepool and remain to be identified.

To fine-map *Sen1* in chapter 4 and map resistance loci involved in the higher pathotypes resistance in chapters 3 and 5, we developed novel workflows, named Comparative Subsequence Sets Analysis (CoSSA), for the genetic study of traits of interest and for the design of haplotype specific markers (**Chapter 3**). As a model, we applied CoSSA to map the *Sen3* gene involved in resistance to all tested pathotypes. A resistant and a susceptible bulk were built in a tetraploid population segregating for pathotypes 2, 6 and 18 resistance. Whole Genome Sequencing (WGS) of the two bulks

and the two parents of the population was performed. *K*-mers specific to the resistant bulk and inherited from the resistant parent were obtained by applying basic set operations to the *k*-mers sets. To increase the haplotype specificity of the resistance specific *k*-mers, we also sequenced varieties susceptible to wart disease and removed commonly occurring *k*-mers from the resistance specific *k*-mers set. Next, we mapped these *k*-mers to the reference genome or made a *de novo* assembly of the reads containing these *k*-mers. Both approaches (reference dependant and reference free) allowed the design of haplotype specific SNP markers flanking the *Sen3* gene. *Sen3* was mapped to a 260 kb interval of the north arm of chromosome 11 containing the complex C76 TNL cluster. A panel of resistant and susceptible varieties was screened to validate the diagnostic value of the *Sen3* flanking markers. The panel screening and a pedigree analysis performed with CoSSA allowed to trace back *Sen3* to the variety Ora.

Before this study, it was still unknown what type of genes are involved in *Synchytrium endobioticum* resistance in potato. In **Chapter 4**, we set out to clone the first potato wart disease *R* gene. We mapped *Sen1* in chapter 2 to a region from the north arm of chromosome 11 which contains the two *R* gene clusters C76 and C77. As these two clusters are composed of *R* genes from the TNL group, we designed a RNAi construct to silence the TNLS from these two clusters to determine if *Sen1* is a TNL. The stable RNAi plants were phenotyped with *AvrSen1*, the effector matching *Sen1*. A significant reduction of the Hypersensitive Response (HR) triggered by *Sen1* was observed in the RNAi plants which confirmed that *Sen1* is an *R* gene from the TNL group. CoSSA was applied in the two populations segregating for pathotype 1 resistance (Chapter 2) to identify *Sen1* specific *k*-mers. Next, haplotype specific SNP markers were designed using the potato reference genome DM and screened in 1,104 seedlings to identify recombinants. The recombinants were phenotyped using *AvrSen1*. *Sen1* was mapped to the C76 cluster, the TNL cluster to which *Sen3* was also mapped (Chapter 3). As DM is susceptible to the potato wart disease and contains two gaps in the *Sen1* region, we used a phased *de novo* assembly of the variety Colomba, which contains *Sen1*, to design new markers in the *R* gene cluster and to identify candidate genes. The *Sen1* interval was reduced to a 97 Kb region in which five TNLS were found. RNAseq data was used to validate the expression and the gene model of the predicted candidate genes. Three genes were discarded after the analysis of RNAseq data. The coding sequences of the two remaining candidates were cloned into plant expression vectors and co-expressed with *AvrSen1* in *N. benthamiana* and in potato.

The screening of the *Sen3* flanking markers in the panel of resistant and susceptible varieties in Chapter 3 showed that *Sen3* is the main source of pathotypes 2, 6 and 18 resistance in German and Polish breeding material. In **Chapter 5**, to identify other loci giving resistance to the higher pathotypes present in the breeding genepool, we used the same strategy as in Chapter 2 and performed a GWAS for pathotypes 2, 6 and 18 resistance in panels representative of the potato breeding genepool. The significantly associated markers were validated in three tetraploid populations. *Sen4*, a new locus involved in resistance to pathotypes 2, 6 and 18, was identified on the south arm of chromosome 12. A smaller effect QTL, also involved in resistance to the three pathotypes, was identified on the north arm of chromosome 11 in the three populations. With CoSSA, we identified another resistance locus, *Sen5*, involved in resistance to pathotypes 2, 6 and 18 and located on the south arm of chromosome 5. In this chapter, we showed that CoSSA was also efficient in identifying QTLs and not only major effect genes. CoSSA was also used to further study the population from Chapter 3 in which *Sen3* was segregating. A minor effect locus was identified which was required, alongside *Sen3*, to bring full resistance to the pathotype 18. We also applied CoSSA to develop haplotype specific markers for the QTLs segregating in the SaKa1 population from (101). Finally, broad variety and wild *Solanum* accession panels were screened with CoSSA, or with the diagnostic markers designed in the different chapters, to inquire the presence of *Sen1*, *Sen2*, *Sen3*, *Sen4* and *Sen5*. This screening brought new information about how the *Sen* genes were introgressed in commercial potato. Altogether this study provided a state-of-the-art overview of non-overlapping sources for breeding wart disease resistance.

In **Chapter 6**, the results obtained in this thesis are summarized and discussed. The focus is set on the different sources of wart disease resistance present in the breeding genepool and their origin, the type of genes involved in wart disease resistance, on the importance of stacking several genes to bring full resistance to the higher pathotypes and on the best strategies to use to identify resistance genes. Finally, future perspectives and future research lines for potato wart disease management and resistance breeding are discussed.

References

1. Iyph - International Year Plant Health [Internet]. [cited 12 november 2019]. Available from: <https://www.ippc.int/en/iyph/>
2. FAO. <http://www.fao.org/faostat/>. 2017.
3. Savary S, Willocquet L, Pethybridge SJ, Esker P, McRoberts N, Nelson A. The global burden of pathogens and pests on major food crops. *Nat Ecol Evol.* march 2019;3(3):430–9.
4. Schilberszky K. Ein Neuer Schorfparasit des Kartoffelknollen. *Ber Botan Ges.* 1896;(14):36–7.
5. Percival J. Potato wart disease: the life history and cytology of *Synchytrium endobioticum*. *Zentralblatt für Bakteriologie, Parasitenkunde, Infektionskrankheiten und Hygiene.* 1910;2(25):440–7.
6. Abdullahi I, Koerbler M, Stachewicz H, Winter S. The 18S rDNA sequence of *Synchytrium endobioticum* and its utility in microarrays for the simultaneous detection of fungal and viral pathogens of potato. *Applied Microbiology and Biotechnology.* augustus 2005;68(3):368–75.
7. Celio GJ, Padamsee M, Dentinger BTM, Bauer R, McLaughlin DJ. Assembling the fungal tree of life: constructing the structural and biochemical database. *Mycologia.* 2006;98(6):850–859.
8. Van Rooij P, Martel A, Haesebrouck F, Pasmans F. Amphibian chytridiomycosis: a review with focus on fungus-host interactions. *Veterinary Research [Internet].* December 2015 [cited 21 october 2019];46(1). Available from: <http://www.veterinaryresearch.org/content/46/1/137>
9. Borthwick AW. Warty disease of potatoes. *Notes from the Royal Botanic Garden, Edinburgh.* 1907;4:115–9.
10. Gough GC. Wart disease of potatoes (*Synchytrium endobioticum*, Perc). A study of its history, distribution, and the discovery of immunity. *Journal of the Royal Horticultural Society.* 1919;45:301–12.
11. Brierley WB. Some research aspects of the wart disease problem. *Roy HortSoc Rep Int Potato Conf.* 1921;93–102.
12. Weiss F. The Conditions of Infection in Potato Wart. *American Journal of Botany.* juli 1925;12(7):413.
13. Hampson MC, Coombes JW. Stress and stimulus modifications of disease severity in the wart disease of potato. *Phytopathology.* 1985;75:817–20.
14. Cotton AD. Host Plants of *Synchytrium Endobioticum*. *Studies from the Pathological Laboratory: IV. Bulletin of Miscellaneous Information (Royal Gardens, Kew).* 1916;1916(10):272.
15. Martin MS. Additional hosts of *Synchytrium endobioticum* (Schilb.) Perc. *The Annals of Applied Biology.* 1929;XVI(3).
16. Hampson MC. Infection of additional hosts of *Synchytrium endobioticum*, the causal agent of potato wart disease: I . Tomato. 1976;56:93–4.
17. Hampson MC. Infection of additional hosts of *Synchytrium endobioticum*, the causal agent of potato wart disease: 2. Tomato, tobacco and species of *Capsicastrum*, *Datura*, *Physalis* and *Schizanthus*. 1979;59:3–6.
18. Curtis KM. IX. The life-history and cytology of *S. endobioticum* (schilb.), perc., the cause of wart disease in potato. *Philosophical Transactions of the Royal Society of London Series B, Containing Papers of a Biological Character.* 1921;210(372–381):409.
19. Obidiegwu JE, Flath K, Gebhardt C. Managing potato wart: a review of present research status and future perspective. *Theoretical and Applied Genetics.* april 2014;127(4):763–80.

20. van de Vossen BTLH, Warris S, Nguyen HDT, van Gent-Pelzer MPE, Joly DL, van de Geest HC, e.a. Comparative genomics of chytrid fungi reveal insights into the obligate biotrophic and pathogenic lifestyle of *Synchytrium endobioticum*. Sci Rep. december 2019;9(1):8672.
21. Lange L, Olson LW. Development of the zoosporangia of *Synchytrium endobioticum*, the causal agent of potato wart disease. Protoplasma. 1981;106(1):97–108.
22. Karling JS. *Synchytrium*. Academic Press, New York. 1964;(1).
23. Przetakiewicz J. The Viability of Winter Sporangia of *Synchytrium endobioticum* (Schilb.) Perc. from Poland. American Journal of Potato Research. december 2015;92(6):704–8.
24. Cartwright K. On the Nature of the Resistance of the Potato to Wart Disease. Annals of Botany. 1926;40(158):391–5.
25. Hampson MC, Proudfoot KG. Potato wart disease, its introduction to North America, distribution and control problems in Newfoundland. FAO Plant Prot Bull. 1974;22:53–64.
26. Hampson M. History, biology, and control of potato wart disease in Canada. Canadian Journal of Plant Pathology. december 1993;15(4):223–44.
27. Hampson MC, Wood SL, Coombes JW. Detection of resting spores of *Synchytrium endobioticum* in soil from vehicles at Port-aux-Basques, Newfoundland. Canadian Journal of Plant Pathology. march 1996;18(1):59–63.
28. van de Vossen BTLH. From metagenome to gene : Identification of the first *Synchytrium endobioticum* effector through comparative genomics [Internet]. Wageningen University; 2019 [cited 3 oktober 2019]. Available from: <http://library.wur.nl/WebQuery/wurpubs/551861>
29. Przetakiewicz J. Assessment of the resistance of potato cultivars to *Synchytrium endobioticum* (Schilb.) Per. in Poland. EPPO bulletin. 2008;38(2):211–215.
30. Baayen RP, Cochius G, Hendriks H, Meffert JP, Bakker J, Bekker M, e.a. History of potato wart disease in Europe – a proposal for harmonisation in defining pathotypes. European Journal of Plant Pathology. september 2006;116(1):21–31.
31. EPPO. *Synchytrium endobioticum* distribution map. Consulted the 2019.10.18. [Internet]. 2019. Available from: <https://gd.eppo.int/taxon/SYNCEN/distribution>
32. Çakır E, van Leeuwen GCM, Flath K, Meffert JP, Janssen WAP, Maden S. Identification of pathotypes of *Synchytrium endobioticum* found in infested fields in Turkey. EPPO Bulletin. augustus 2009;39(2):175–8.
33. Sikharulidze ZV, Ghoghoberidze SY, Mentink NM, Meparishvili GV, Tsetschladze TsM, van Leeuwen GCM. Identification of the pathotype of *Synchytrium endobioticum*, causal agent of potato wart disease, present in Georgia. EPPO Bull. augustus 2019;49(2):314–20.
34. Przetakiewicz J. First Report of New Pathotype 39(P1) of *Synchytrium endobioticum* Causing Potato Wart Disease in Poland. Plant Dis. 2015;99:285.
35. Przetakiewicz J. First Report of *Synchytrium endobioticum* (Potato Wart Disease) Pathotype 18(T1) in Poland. Plant Disease. 2014;98(5):688–688.
36. EPPO. *Synchytrium endobioticum*: Diagnostic protocols. EPPO Bulletin. augustus 2004;34(2):213–8.
37. EPPO. PM 7/28 (2) *Synchytrium endobioticum*. EPPO Bulletin. december 2017;47(3):420–40.
38. EPPO. EPPO A2 List of pests recommended for regulation as quarantine pests [Internet]. 2019. Available from: https://www.eppo.int/ACTIVITIES/plant_quarantine/A2_list
39. PM 9/5 (2) *Synchytrium endobioticum*. EPPO Bulletin. december 2017;47(3):511–2.
40. PM 3/59 (3) *Synchytrium endobioticum*: descheduling of previously infested plots. EPPO Bulletin. december 2017;47(3):366–8.

41. van de Vossen BTLH, Brankovics B, Nguyen HDT, van Gent-Pelzer MPE, Smith D, Dadej K, e.a. The linear mitochondrial genome of the quarantine chytrid *Synchytrium endobioticum*; insights into the evolution and recent history of an obligate biotrophic plant pathogen. BMC Evolutionary Biology [Internet]. december 2018 [cited 5 march 2019];18(1). Available from: <https://bmcevolbiol.biomedcentral.com/articles/10.1186/s12862-018-1246-6>
42. van de Vossen BTLH, Prodhomme C, van Arkel G, van Gent-Pelzer MPE, Bergervoet M, Brankovics B, e.a. The *Synchytrium endobioticum* AvrSen1 Triggers a Hypersensitive Response in *Sen1* Potatoes While Natural Variants Evade Detection. Molecular Plant-Microbe Interactions. 17 september 2019;MPMI-05-19-0138.
43. Ray DK, Mueller ND, West PC, Foley JA. Yield Trends Are Insufficient to Double Global Crop Production by 2050. Hart JP, redacteur. PLoS ONE. 19 juni 2013;8(6):e66428.
44. Renault D, Wallender WW. Nutritional water productivity and diets. Agricultural Water Management. augustus 2000;45(3):275–96.
45. Birch PRJ, Bryan G, Fenton B, Gilroy EM, Hein I, Jones JT, e.a. Crops that feed the world 8: Potato: are the trends of increased global production sustainable? Food Security. december 2012;4(4):477–508.
46. Jones JDG, Dangl JL. The plant immune system. Nature. november 2006;444(7117):323–9.
47. Nurnberger T, Brunner F, Kemmerling B, Piater L. Innate immunity in plants and animals: striking similarities and obvious differences. Immunol Rev. april 2004;198(1):249–66.
48. Flor HH. Current Status of the Gene-For-Gene Concept. Annual Review of Phytopathology. september 1971;9(1):275–96.
49. Thomma BPHJ, Nurnberger T, Joosten MHAJ. Of PAMPs and Effectors: The Blurred PTI-ETI Dichotomy. Plant Cell. januari 2011;23(1):4–15.
50. Poland JA, Balint-Kurti PJ, Wissner RJ, Pratt RC, Nelson RJ. Shades of gray: the world of quantitative disease resistance. Trends in Plant Science. januari 2009;14(1):21–9.
51. Kourelis J, van der Hoorn RAL. Defended to the Nines: 25 Years of Resistance Gene Cloning Identifies Nine Mechanisms for R Protein Function. The Plant Cell. februari 2018;30(2):285–99.
52. Śliwka J. Genetic factors encoding resistance to late blight caused by *Phytophthora infestans* (Mont.) de Bary on the potato genetic map. Cellular and molecular biology letters. 2004;9:855–67.
53. Michelmore RW, Paran I, Kesseli RV. Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proceedings of the national academy of sciences. 1991;88(21):9828–9832.
54. Giovannoni JJ, Wing RA, Ganai MW, Tanksley SD. Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations. Nucleic Acids Research. 1991;19(23):6553–6568.
55. Li X, van Eck HJ, Rouppe van der Voort JNAM, Huigen D-J, Stam P, Jacobsen E. Autotetraploids and genetic mapping using common AFLP markers: the *R2* allele conferring resistance to *Phytophthora infestans* mapped on potato chromosome 4: Theor Appl Genet. juni 1998;96(8):1121–8.
56. Huang S, Vleeshouwers VG, Werij JS, Hutten RC, van Eck HJ, Visser RG, e.a. The *R3* resistance to *Phytophthora infestans* in potato is conferred by two closely linked *R* genes with distinct specificities. Molecular Plant-Microbe Interactions. 2004;17(4):428–435.
57. Smilde WD, Brigneti G, Jagger L, Perkins S, Jones JDG. *Solanum mochiquense* chromosome IX carries a novel late blight resistance gene *Rpi-moc1*. Theor Appl Genet. januari 2005;110(2):252–8.
58. Rauscher GM, Smart CD, Simko I, Bonierbale M, Mayton H, Greenland A, e.a. Characterization and mapping of *R Pi-ber*, a novel potato late blight resistance gene from *Solanum berthaultii*. Theor Appl Genet. februari 2006;112(4):674–87.

59. Bradshaw JE, Bryan GJ, Lees AK, McLean K, Solomon-Blackburn RM. Mapping the *R10* and *R11* genes for resistance to late blight (*Phytophthora infestans*) present in the potato (*Solanum tuberosum*) R-gene differentials of Black. Theor Appl Genet. februari 2006;112(4):744–51.
60. van der Linden CG, Wouters DCAE, Mihalka V, Kochieva EZ, Smulders MJM, Vosman B. Efficient targeting of plant disease resistance loci using NBS profiling. Theor Appl Genet. juli 2004;109(2):384–93.
61. Jo K-R, Arens M, Kim T-Y, Jongsma MA, Visser RGF, Jacobsen E, e.a. Mapping of the *S. demissum* late blight resistance gene *R8* to a new locus on chromosome *IX*. Theoretical and Applied Genetics. december 2011;123(8):1331–40.
62. Pel MA, Foster SJ, Park T-H, Rietman H, van Arkel G, Jones JDG, e.a. Mapping and Cloning of Late Blight Resistance Genes from *Solanum venturii* Using an Interspecific Candidate Gene Approach. MPMI. may 2009;22(5):601–15.
63. Jacobs MMJ, Vosman B, Vleeshouwers VGAA, Visser RGF, Henken B, van den Berg RG. A novel approach to locate *Phytophthora infestans* resistance genes on the potato genetic map. Theoretical and Applied Genetics. februari 2010;120(4):785–96.
64. Verzaux E, Budding D, de Vetten N, Niks RE, Vleeshouwers VGAA, van der Vossen EAG, e.a. High Resolution Mapping of a Novel Late Blight Resistance Gene *Rpi-av11*, from the Wild Bolivian Species *Solanum avilesii*. American Journal of Potato Research. december 2011;88(6):511–9.
65. Verzaux E, van Arkel G, Vleeshouwers VGAA, van der Vossen EAG, Niks RE, Jacobsen E, e.a. High-Resolution Mapping of Two Broad-Spectrum Late Blight Resistance Genes from Two Wild Species of the *Solanum circaefolium* Group. Potato Research. juni 2012;55(2):109–23.
66. Hamilton JP, Hansey CN, Whitty BR, Stoffel K, Massa AN, Van Deynze A, e.a. Single nucleotide polymorphism discovery in elite North American potato germplasm. BMC genomics. 2011;12(1):302.
67. Felcher KJ, Coombs JJ, Massa AN, Hansey CN, Hamilton JP, Veilleux RE, e.a. Integration of Two Diploid Potato Linkage Maps with the Potato Genome Sequence. Aerts J, redacteur. PLoS ONE. 27 april 2012;7(4):e36347.
68. Vos PG, Uitdewilligen JGAML, Voorrips RE, Visser RGF, van Eck HJ. Development and analysis of a 20K SNP array for potato (*Solanum tuberosum*): an insight into the breeding history. Theoretical and Applied Genetics. december 2015;128(12):2387–401.
69. Obidiegwu JE, Sanetomo R, Flath K, Tacke E, Hofferbert H-R, Hofmann A, e.a. Genomic architecture of potato resistance to *Synchytrium endobioticum* disentangled using SSR markers and the 8.3k SolCAP SNP genotyping array. BMC Genetics [Internet]. december 2015 [cited 3 augustus 2017];16(1). Available from: <http://www.biomedcentral.com/1471-2156/16/38>
70. Potato Genome Sequencing Consortium, Pan S, Cheng S, Zhang B, Mu D, Ni P, e.a. Genome sequence and analysis of the tuber crop potato. Nature. 10 juli 2011;475(7355):189–95.
71. Jupe F, Witek K, Verweij W, Śliwka J, Pritchard L, Etherington GJ, e.a. Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. The Plant Journal. november 2013;76(3):530–44.
72. Chen X, Lewandowska D, Armstrong MR, Baker K, Lim T-Y, Bayer M, e.a. Identification and rapid mapping of a gene conferring broad-spectrum late blight resistance in the diploid potato species *Solanum verrucosum* through DNA capture technologies. Theor Appl Genet. juni 2018;131(6):1287–97.
73. Witek K, Jupe F, Witek AI, Baker D, Clark MD, Jones JDG. Accelerated cloning of a potato late blight-resistance gene using RenSeq and SMRT sequencing. Nature Biotechnology. 25 april 2016;34(6):656–60.
74. Grech-Baran M, Witek K, Szajko K, Witek AI, Morgiewicz K, Wasilewicz-Flis I, e.a. Extreme resistance to *Potato virus Y* in potato carrying the *Ry^{sto}* gene is mediated by a TIR - NLR

- immune receptor. Plant Biotechnology Journal [Internet]. 4 september 2019 [cited 2 oktober 2019]; Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/pbi.13230>
75. Strachan SM, Armstrong MR, Kaur A, Wright KM, Lim TY, Baker K, e.a. Mapping the *H2* resistance effective against *Globodera pallida* pathotype Pa1 in tetraploid potato. Theoretical and Applied Genetics. april 2019;132(4):1283–94.
 76. Lin X. Apoplastic effectors, MAMPs and surface immune receptors: the battlefield of the *Phytophthora infestans* - potato interaction. Wageningen University; 2018.
 77. Giolai M, Paajanen P, Verweij W, Witek K, Jones JDG, Clark MD. Comparative analysis of targeted long read sequencing approaches for characterization of a plant's immune receptor repertoire. BMC Genomics [Internet]. december 2017 [cited 11 november 2019];18(1). Available from: <http://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-017-3936-7>
 78. Simko I. One potato, two potato: haplotype association mapping in autotetraploids. Trends in Plant Science. september 2004;9(9):441–8.
 79. Álvarez MF, Angarita M, Delgado MC, García C, Jiménez-Gomez J, Gebhardt C, e.a. Identification of Novel Associations of Candidate Genes with Resistance to Late Blight in *Solanum tuberosum* Group Phureja. Frontiers in Plant Science [Internet]. 15 juni 2017 [cited 11 november 2019];8. Available from: <http://journal.frontiersin.org/article/10.3389/fpls.2017.01040/full>
 80. Gebhardt C, Ballvora A, Walkemeier B, Oberhagemann P, Schüler K. Assessing genetic potential in germplasm collections of crop plants by marker-trait association: a case study for potatoes with quantitative variation of resistance to late blight and maturity type. Molecular Breeding. 2004;13(1):93–102.
 81. Pajerowska-Mukhtar K, Stich B, Achenbach U, Ballvora A, Lübeck J, Strahwald J, e.a. Single Nucleotide Polymorphisms in the *Allene Oxide Synthase 2* Gene Are Associated With Field Resistance to Late Blight in Populations of Tetraploid Potato Cultivars. Genetics. march 2009;181(3):1115–27.
 82. Spieckermann A, Kothoff P. Testing potatoes for wart resistance. Deutsche Landwirtschaftliche Presse 51:114–115. 1924;
 83. Glynn MD. Infection experiments with wart disease of potatoes. *Synchytrium endobioticum* (Schilb.) Perc. Annals of Applied Biology. 1925;12(1):34–60.
 84. Lemmerz J. A new simplified infection procedure for testing potato cultivars for wart resistance. Zücher. 1930;2(288):97.
 85. van de Vossen BTLH, van Gent-Pelzer MPE, Boerma M, van der Gouw LP, van der Lee TAJ, Vossen JH. An Alternative Bioassay for *Synchytrium endobioticum* Demonstrates the Expression of Potato Wart Resistance in Aboveground Plant Parts. Phytopathology. juni 2019;109(6):1043–52.
 86. Salaman RN, Lesley JW. Genetic studies in potatoes; the inheritance of immunity to wart disease. Journal of Genetics. 1923;13(2):177–186.
 87. van Berloo R, Hutten RCB, van Eck HJ, Visser RGF. An Online Potato Pedigree Database Resource. Potato Research. januari 2007;50(1):45–57.
 88. Ross H. Potato breeding - Problems and perspectives. Suppl. to J. Plant Breeding, Paul Parey, Berlin and Hamburg. 1986.
 89. Bukasov SM. Les Bases de la Sélection de la Pomme de terre (suite et fin). Revue de botanique appliquée et d'agriculture coloniale. 1940;20(223):179–89.
 90. Rothacker D. Arbeiten zur Zuchtung krebseresistenter Kartoffeln. Der Züchter. 1957;
 91. Toxopeus HJ, Huijsman CA. Breeding for resistance to potato root eelworm: I. Preliminary data concerning the inheritance and the nature of resistance. Euphytica. oktober 1953;2(3):180–6.

92. Ellenby C. Resistance to the Potato Root Eelworm, *Heterodera rostochiensis* Wollenweber. Nature. december 1952;170(4337):1016–1016.
93. Khiutti A, Afanasenko O, Antonova O, Shuvalov O, Novikova L, Krylova E, e.a. Characterization of resistance to *Synchytrium endobioticum* in cultivated potato accessions from the collection of Vavilov Institute of Plant Industry: Resistance to *Synchytrium endobioticum* in cultivated potato accessions. Plant Breeding. december 2012;131(6):744–50.
94. van Soest LJM, Seidewitz L. Evaluation data on tuber-bearing *Solanum* species. Inst. für Pflanzenbau und Pflanzenzüchtung der FAL-Stichting voor Plantenveredeling (SVP). 1981. 165 p.
95. Rothacker D, Muller WA. Arbeiten zur Zuchtung krebsresistenter Kartoffeln. Der Züchter. 1960;
96. Plich J, Przetakiewicz J, Śliwka J, Flis B, Wasilewicz-Flis I, Zimnoch-Guzowska E. Novel gene *Sen2* conferring broad-spectrum resistance to *Synchytrium endobioticum* mapped to potato chromosome XI. Theoretical and Applied Genetics. november 2018;131(11):2321–31.
97. Black W. Studies on the inheritance of resistance to wart disease (*Synchytrium endobioticum* (Schilb.) Perc.) in potatoes. Journal of Genetics. 1935;30(1):127–146.
98. Lunden AP, Jørstad I. Investigations on the inheritance of immunity to wart disease (*Synchytrium Endobioticum* [Schilb.] Perc.) in the potato. Journal of Genetics. 1934;29(3):375–385.
99. Hehl R, Faurie E, Hesselbach J, Salamini F, Whitham S, Baker B, e.a. TMV resistance gene *N* homologues are linked to *Synchytrium endobioticum* resistance in potato. TAG Theoretical and Applied Genetics. 1999;98(3):379–386.
100. Brugmans B, Hutten RGB, Rookmaker ANO, Visser RGF, van Eck HJ. Exploitation of a marker dense linkage map of potato for positional cloning of a wart disease resistance gene. Theoretical and Applied Genetics. januari 2006;112(2):269–77.
101. Ballvora A, Flath K, Lübeck J, Strahwald J, Tacke E, Hofferbert H-R, e.a. Multiple alleles for resistance and susceptibility modulate the defense response in the interaction of tetraploid potato (*Solanum tuberosum*) with *Synchytrium endobioticum* pathotypes 1, 2, 6 and 18. Theoretical and Applied Genetics. december 2011;123(8):1281–92.
102. Gebhardt C, Bellin D, Henselewski H, Lehmann W, Schwarzfischer J, Valkonen JPT. Marker-assisted combination of major genes for pathogen resistance in potato. Theoretical and Applied Genetics. mei 2006;112(8):1458–64.
103. Groth J, Song Y, Kellermann A, Schwarzfischer A. Molecular characterisation of resistance against potato wart races 1, 2, 6 and 18 in a tetraploid population of potato (*Solanum tuberosum* subsp. *tuberosum*). Journal of Applied Genetics. mei 2013;54(2):169–78.
104. Bartkiewicz A, Chilla F, Terefe-Ayana D, Lübeck J, Strahwald J, Tacke E, e.a. Improved genetic resolution for linkage mapping of resistance to potato wart in monoparental dihaploids with potential diagnostic value in tetraploid potato varieties. Theoretical and Applied Genetics. december 2018;131(12):2555–66.
105. Antonova OY, Shvachko NA, Novikova LY, Shuvalov OY, Kostina LI, Klimenko NS, e.a. Genetic diversity of potato varieties bred in Russia and its neighboring countries based on the polymorphism of SSR-loci and markers associated with resistance R-genes. Russ J Genet Appl Res. juli 2017;7(5):489–500.

The background of the page features a stylized illustration. On the left, there are green potato leaves. In the center and bottom, there are yellow potato tubers. Interspersed among the leaves and tubers are several DNA double helix structures in red, blue, and yellow. The overall style is artistic and scientific.

Chapter 2

Distribution of P1(D1) wart disease resistance in potato germplasm and GWAS identification of haplotype-specific SNP markers

Charlotte Prodhomme, Peter G. Vos, Maria João Paulo,
Jasper E. Tammes, Richard G.
F. Visser, Jack H. Vossen, Herman J. van Eck

This chapter was published with minor modifications in
Theoretical and Applied Genetics 2020.

Abstract

Key message

A Genome-Wide Association Study using 330 commercial potato varieties identified haplotype specific SNP markers associated with pathotype 1(D1) wart disease resistance.

Abstract

Synchytrium endobioticum is a soil borne obligate biotrophic fungus responsible for wart disease. Growing resistant varieties is the most effective way to manage the disease. This paper addresses the challenge to apply molecular markers in potato breeding. Although markers linked to *Sen1* were published before, the identification of haplotype specific single nucleotide polymorphisms (hsSNPs) may result in marker assays with high diagnostic value to allow efficient applications in breeding programs. To identify hsSNP markers, we performed a Genome-Wide Association Study (GWAS) in a panel of 330 potato varieties representative of the commercial potato gene pool. SNP markers significantly associated with pathotype 1 resistance were identified on chromosome 11, at the position of the previously identified *Sen1* locus. Haplotype specificity of the SNP markers was examined through the analysis of false positives and false negatives, and validated in two independent populations. This paper illustrates why it is not always feasible to design markers without false positives and false negatives for marker-assisted selection (MAS). In the case of *Sen1*, founders could not be traced because of a lack of identity-by-descent and because of the decay of linkage disequilibrium (LD) between *Sen1* and flanking SNP markers. *Sen1* appeared to be the main source of pathotype 1 resistance in potato varieties, but it does not explain all the resistance observed. Recombination and introgression breeding may have introduced new, albeit rare haplotypes involved in pathotype 1 resistance. The GWAS approach, in such case, is instrumental to identify SNPs with the best possible diagnostic value for marker-assisted breeding.

Keywords

Solanum tuberosum, GWAS, *Synchytrium endobioticum*, pathotype 1, tetraploid, haplotype specific SNP marker, population structure.

Introduction

Potato wart disease, caused by *Synchytrium endobioticum*, induces the formation of galls on tubers of potato (*Solanum tuberosum*). The pathogen belongs to the phylum

of the Chytridiomycota, one of the early diverging fungal lineages. It is an obligate biotrophic soil borne fungus producing winter spores presumably after a sexual phase of its life cycle. These persistent spores can remain viable in the soil for more than 40 years (1). The yield losses can reach 50 to 100% and no fungicides are available against this pathogen. Therefore, it has the status of a quarantine disease. When the presence of the fungus is recorded in a field, no potato can be cultivated until soil tests become negative for the presence of sporangia (2). For these reasons, breeding for varieties resistant to the potato wart disease is essential, especially in Europe where the presence of the pathogen has been recorded in many countries (3).

Breeding for potato varieties resistant to *S. endobioticum* is challenging. Before 1941, only pathotype 1 was known and breeding programs produced resistant varieties. After 1941, new pathotypes emerged and most varieties are susceptible to them, because resistance to the available pathotype 1 was evaded (4). The availability of haplotype specific SNP (hsSNP) markers for breeders to screen their clones as well as the potato germplasm to find resistance sources is limited. Consequently, large numbers of laborious disease assays are needed in breeding programs. Different phenotyping methods exist, consisting either of field trials or laboratory assays using winter (5) or summer spores (6,7) for inoculation. Often, these different methods do not give consistent results, they are difficult to manage and time consuming. Thus, molecular markers associated with resistance should be identified for breeding resistant potato varieties.

The dominant monogenic *Sen1* locus confers strong resistance to pathotype P1 (D1), hereafter abbreviated to P1. The *Sen1* locus has been mapped on the north arm of chromosome 11 (8), in a region containing several Nucleotide-binding Leucine-rich Repeat-encoding (NLR) gene clusters. (9) described two NLR clusters (C76 and C77) at this position of chromosome 11. Both clusters consisted of Toll/interleukin-1 receptor Nucleotide-binding Leucine-rich repeat (TNL) encoding genes. Another dominant monogenic locus conferring P1 resistance, *Sen1-4*, has been mapped to the south arm of chromosome 4 (10). Additional genetic studies have been performed using diploid and tetraploid populations to identify wart disease resistance loci. *Sen2* is a dominant monogenic locus, bringing strong resistance to a wide range of pathotypes, located on chromosome 11 approximately 32 Mb south from *Sen1* (11). *Sen3*, another dominant monogenic locus involved in strong resistance to pathotypes 2, 6 and 18, was fine-mapped to the same region as *Sen1* (12,13). Multiple QTLs involved in resistance to several pathotypes or in pathotype specific resistance were

also identified across the potato genome in different genetic backgrounds (14–16). Nevertheless, few of the markers linked to *Sen1* or these other monogenic loci and QTLs have been validated in a wider panel of distantly related varieties, representative for the entire gene pool of commercial potato.

Such a validation experiment has been performed for markers linked to P1 resistance by (16), but the markers they tested were, inconveniently, linked in repulsion phase with the resistance locus. The NI25 marker, used for the first mapping of *Sen1* (8,17), was recently validated in a limited set of twelve resistant and susceptible varieties by (18) and in a panel composed only of Russian varieties (19). NI25, GP125 and StI046, all found to be linked to *Sen1* in mapping populations, were screened in a panel of 89 Polish and German varieties in the CORNET project SynTest, but the results were not elaborately reported (12). The two DArTseq markers flanking *Sen1* designed by (11) in a diploid segregating population were not validated in tetraploid varieties.

Compared to traditional linkage mapping approaches, Genome-Wide Association Studies (GWAS) allow to include the genetic diversity of a wider panel of varieties and to reach a superior mapping resolution thanks to a higher number of recombination events. Aside from studies in genetic model organisms, GWAS has already been successfully used in several crops such as soybean (20), rice (21), wheat (22) as well as potato (23,24). GWAS has never been applied to identify loci involved in *S. endobioticum* resistance, although resistance values are well documented for varieties on National Lists motivated by the Value for Cultivation and Use (VCU) requirements.

In this study, we performed a GWAS in a panel of 330 tetraploid potato varieties representative of the gene pool of commercial potato breeding. We identified a region with SNPs strongly associated with pathotype 1 resistance on the north arm of chromosome 11. Four significantly associated SNP markers were selected and validated in two different biparental populations segregating for P1 resistance, to demonstrate their linkage in coupling phase with the *Sen1* locus. The markers do not explain all the P1 resistance present in the GWAS panel. This suggests a more complex architecture of pathotype 1 resistance involving one predominant gene with a major effect along with other less frequent haploblocks with *Sen1* or *Sen1*-like alleles, or loci beyond chromosome 11.

Materials and Methods

GWAS panel

A subset of the panel of 569 potato genotypes described in (25) was used, because phenotypic data for wart disease were available for 330 individuals. These individuals are representative for the gene pool of commercial potato, because it includes old and recent varieties, developed for different markets, widely used progenitors, as well as Dutch advanced breeding lines (Supplementary file 1). Because this panel is composed of 304 commercial varieties and only 26 progenitors and breeding clones, hereafter we use the term of varieties for this GWAS panel.

GWAS panel genotyping data

The genotyping data used for GWAS were obtained by (25) with a 20K Infinium SNP array (hereafter mentioned as the SolSTW array) comprising 4,463 attempted SNPs from the potato SolCAP array (26,27) and 16.5K attempted PotVar SNPs from a targeted enrichment study (24). For 14,530 SNPs in the panel, the assay quality allowed determination of the tetraploid scores of the markers, with a score of 0 for a nulliplex genotype call, 1 for simplex, 2 for duplex, 3 for triplex and 4 for quadruplex. Markers were removed when missing values exceeded 20%, when the minor allele frequency (hereafter MAF) was below 1.25%, and when the markers were present in less than 6 varieties, resulting in a subset of 10,968 SNPs used for the GWAS (Supplementary file 1). The MAF threshold of 1.25% implies for a tetraploid species that ~5% of the varieties have this SNP allele in simplex condition, which is equivalent to the commonly used threshold of 5% in diploids.

Population structure

The population structure of the variety panel was studied using a kinship matrix, calculated using van Raden (28), with a random subset of 1,000 markers (Supplementary file 1) to perform a Principal Coordinate Analysis (PCoA), which suggested four structure groups. STRUCTURE (29) was used to assign each variety to one of the four groups. We used the settings advised by (30): length of burn-in period 10,000, number of MCMC reps after burn-in 10,000, admixture model, allele frequencies correlated among the populations. We made 10 repetitions for $K = 4$ (Supplementary file 2).

GWAS panel phenotypic data

Phenotypic data for P1 resistance were collected from various sources, such as National Lists (Value for Cultivation and Use (VCU-data)), various websites and brochures from commercial breeders (Supplementary file 3). As the scales used by the different sources varied, we converted the scales into resistant scores ranging from 1 (highly susceptible) to 10 (highly resistant), as described in Supplementary file 3. During a preliminary GWAS, false positive varieties were identified (susceptible varieties holding P1 associated markers). For some of these varieties, the phenotypic data we gathered was scarce or inconsistent. Therefore, for 12 varieties, new disease assays were performed to assess these inconsistencies. The re-assessment was performed during spring 2016 by HLB (Wijster, the Netherlands) using both the Spieckermann (5) and Glynne-Lemmerzähl (6,7) methods with 9 and 5 tuber eye-pieces respectively. Isolate MB42 P1(D1) was used for both tests (31). The same quantitative scoring scale was used for both experiments, ranging from 10 (highly resistant, corresponding to type 1 in Germany and type “–” in the Netherlands) to 1 (highly susceptible, corresponding to type 5 in Germany and type X in the Netherlands). Varieties rated in a category lower than 8 are considered as susceptible as it corresponds to very late defence necrosis, which can be slower than the pathogen maturation (32). For each variety assessed, a mean score was calculated (Supplementary table 1). Hence, the final phenotypic dataset used in the rest of the analysis included corrected phenotypes for these 12 varieties.

Genome-Wide Association Study

The genotypic data ranged from 0 to 4 reflecting the dosage of the SNPs. For each variety in the panel, we calculated a final resistance score corrected for the source effect using restricted maximum likelihood (REML). After correction for data source, phenotypes ranged from 0.85 to 10.72. The model included the source as a random effect and the variety as a fixed effect as follows: *Pathotype1 resistance* = *source* + *variety*. We carried out the association analysis using three different association models. Firstly, a simple linear regression model without any structure correction (naive model). Secondly, the 30 first Principal Coordinates of the PCoA performed on the van Raden kinship were included as a fixed effect in a mixed linear regression model to correct for the population structure. In the third model, we included five co-factors (most significant markers from the PCoA corrected model) as fixed effects in the PCoA corrected model. Quantile-quantile (qq) plots were produced to compare the models. The GWAS models were fitted in GenStat version 18 (VSN International,

2015). The genome-wide significance threshold – $\log_{10}(p) = 4.7$ was calculated by procedure QTHRESHOLD, using the method developed by (33). GWAS results were visualised on R version 3.2.2 using the EasyStrata package (34).

Validation populations

Populations segregating for P1 resistance were used to validate markers identified by the GWAS. The P1 resistant varieties Desiree and Kuras were crossed with Aventura (susceptible to P1) giving 42 (AxD population) and 35 progeny (KxA population), respectively. The descendants were phenotyped for their resistance to P1 during spring 2016 with the Spieckermann (5) and the Glynne-Lemmerzahn (6,7) methods with 8 and 6 eye-pieces respectively. Two subsets of 9 and 17 descendants from the AxD and KxA populations were re-phenotyped in spring 2017 with 12 and 6 eye-pieces for the Spieckermann and the Glynne-Lemmerzahn methods, respectively. The Spieckermann assays were performed by HLB using isolate MB42. The Glynne-Lemmerzahn assays were performed by the Laboratory of Quarantine Organisms, Department of Plant Pathology, IAH-PIB, Poland with the isolate JKI P1(D1)-2009. Disease symptoms were rated from 1 (highly resistant, early defence necrosis) to 5 (highly susceptible). Mean scores were calculated between replicates (6-8 eye-pieces, or 6-12 tubers) within years and between the two assessment years.

DNA extraction

Total genomic DNA was isolated from freshly harvested freeze-dried tubers or fresh leaves using a modified CTAB-method (35). The DNA concentration was estimated using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific) and adjusted to a concentration of 5-50 ng/μl. The DNA quality was assessed on ethidium bromide containing agarose gels.

SNP validation

To validate linkage in coupling phase of significantly associated SNPs, Kompetitive Allele Specific Polymorphisms (KASP) markers were designed. KASP primers design was done by LGC Genomics (LGC, Hoddeston, UK) or in-house, using Primer3 (v3.0.0) (36). All KASP primers are given in Supplementary table 2. Segregations patterns and associations between markers and P1 resistance was tested using χ^2 -tests and Kruskal-Wallis tests respectively. All the statistical analyses were performed with R version 3.2.2.

Results

GWAS panel structure and P1 resistance data

A panel of 330 tetraploid potato varieties composed of old widely used clones as well as recent varieties from 14 different countries was used for GWAS ((25); Supplementary file 1; Supplementary figure 1). A Principal Coordinate Analysis and a STRUCTURE analysis were performed on a random subset of 1,000 SNPs (Figure 1). The first two principal coordinates (PCo) explained 8.8% of the variance and showed four groups in the GWAS panel. The first group ('starch') was composed of varieties bred for starch industry. The second group ('GB origins') was composed of varieties with British origins. The third group ('Agria') has varieties descending from Agria. Agria, at the edge of this cluster, was frequently used in the 1990's to breed for varieties suitable for French fries. The last group ('rest') was composed of varieties with diverse origins.

The phenotypic values for P1 resistance in the GWAS panel showed a bimodal distribution, skewed toward resistance (Supplementary figure 1). The skewed distribution reflects the historical progress in breeding wart resistant varieties. Before the 1920s, 50% of the varieties were resistant to P1 whereas since the 1980s more than 77% of the varieties are resistant. In the 'starch' structure group, most varieties were resistant (Supplementary figure 2). Only Amado, Aventura, CMK2007-042-003, Festien, Karakter and Lady Rosetta were susceptible (scores below 8). Resistant varieties were also predominant in the 'Agria' structure group, which contained only seven susceptible varieties (Endeavour, Fontane, Heroud, Hzpc06-11, Lady Felicia, Markies and Sinora). The group containing varieties with British origin showed most variation for resistance: 31% of the varieties from this group were susceptible. These observations suggested a correlation between the panel structure and potato wart resistance. Indeed, the 30 first principal coordinates of the PCoA were 51% correlated with P1 resistance and were consequently included in the GWAS model to correct for the structure confounding effect.

In view of the skewed distribution of phenotypic values, the distribution of the residuals of the PCoA corrected GWAS including the five most significant but uncorrelated markers as cofactors was examined. The close to normal distribution of residuals confirmed the legitimacy of using a linear regression model for GWAS (Supplementary figure 3).

Figure 1. Principal Coordinate Analysis of the GWAS panel

The structure of the panel of 330 potato varieties was analyzed using the first two principal coordinates of the PCoA performed on a random subset of 1,000 SNPs on the GWAS panel. Each variety is represented by a pseudo-colour that scales with the level of resistance to P1 (dark colour (P1 predicted mean < 6) indicates highly susceptible varieties). The shapes of the dots indicate the STRUCTURE group affiliation of each variety according to Supplementary file 2.

Genome-Wide Association Study of P1 resistance

Association between phenotypic values for P1 resistance of 330 tetraploid varieties and 10,968 SNP markers with the naive model identified a clear pileup of 47 SNPs on chromosome 11. In addition, 23 SNPs from several other genomic regions showed significant association with wart resistance. However, when using the PCoA corrected model, only 12 markers remained associated with P1 resistance, all located on the north arm of chromosome 11 between 0.78 Mb and 4.35 Mb (Figure 2, Table 1). This region is known to harbour the *Sen1* locus associated with P1 resistance (8).

With a QQ-plot, the corrected and uncorrected (naive) models were compared to study *p*-value inflation (Supplementary figure 4). Without PCoA correction, the distribution of the observed $-\log_{10}(p)$ values deviated from the diagonal line representing the expected values under a model of no association. The model including the 30 first PCo effectively decreased this deviation. This justifies the conclusion that other regions identified with the naive model are false positives.

Analysis of predictions of SNPs from chromosome 11

For each of the 12 associated SNPs from chromosome 11 in the PCoA corrected results, we calculated the false positive rate (FPR) and the false negative rate (FNR). The FPR is the percentage of varieties which hold the marker allele but are susceptible whereas the FNR is the percentage of resistant varieties that lack the marker allele. According to the MAF, the FPR and the FNR, we divided the associated SNPs in three groups. The first group, located between 0.78 and 3.07 Mb, contained six SNPs (PotVar0064472, solcap_snp_c2_13431, PotVar0066243, PotVar0066295, PotVar0067303 and PotVar0067483; Supplementary file 1, Table 1) with a high frequency in the GWAS panel ($0.26 \leq \text{MAF} \leq 0.49$), a high FPR between 11 and 20% and a low FNR between 1 and 15%. These SNPs descended from heirloom varieties such as Yam (<1787), Myatt's Ashleaf (1804), Jaune d'Or (<1850) and Pink Fir Apple (1850), of which Myatt's Ashleaf and Pink Fir Apple are known to be resistant. The raw data (Supplementary file 1) does not suggest that the markers from group 1 are indicative for a specific haplotype conferring resistance in spite of their significant

Figure 2. GWAS of pathotype 1 resistance

Miami plot of the GWAS of pathotype 1 resistance without the PCoA correction (in blue, lower part of the plot) and with the PCoA correction (in green, upper part of the plot) performed on 330 potato varieties. The association ($-/\log_{10}(p)$) of each SNP with pathotype 1 resistance is represented. The x axis represents the physical position of each SNP on the 12 potato chromosomes (u corresponds to markers of unanchored regions and to chloroplast markers). The red horizontal lines correspond to the significance threshold calculated with the Li and Ji method ($-\log_{10}(p) = 4.7$). The strongly associated region at the beginning of chromosome 11 corresponds to the *Sen1* locus previously identified by (8).

association with wart disease resistance, and cannot be recommended for marker-assisted selection (MAS). The second group contained one marker (PotVar0066434) with a negative effect on resistance. This marker, presumably linked in repulsion phase to P1 resistance, is not useful for MAS either. The third group had five markers (PotVar0067008, solcap_snp_c1_2314, PotVar0106272, PotVar0106247 and PotVar0105904), located more south on chromosome 11 (2.77 – 4.35 Mb), with an MAF of $\sim 18\%$. These SNPs had a low FPR ($0.04 \leq \text{FPR} \leq 0.08$) but a higher FNR ($0.23 \leq \text{FNR} \leq 0.25$). Based on the year of market introduction of the variety, all group 3 markers are historically first observed in Pink Fir Apple and absent in older varieties. The pattern of marker raw data suggest that these markers are in strong LD, and belong to a haplotype that includes resistance to P1, albeit historical recombination events led to decay of LD, proportional to the distance of the markers from the resistance gene (37).

Table 1. Significantly associated markers in the PCoA corrected GWAS model

SNP	Chromosome	Position (bp)	p-value	Effect	MAF	FPR	FNR
PotVar0064472 ^a	11	787,356	6.66E-07	1.04	0.41	0.18	0.06
solcap_snp_c2_13431 ^a	11	788,222	1.92E-07	1.08	0.41	0.18	0.06
PotVar0066243 ^a	11	2,060,973	1.28E-08	1.04	0.48	0.19	0.01
PotVar0066295 ^a	11	2,261,174	1.23E-05	0.78	0.49	0.20	0.02
PotVar0066434 ^b	11	2,704,904	1.63E-06	-0.84	0.39	0.22	0.12
PotVar0067008 ^c	11	2,773,680	1.07E-09	1.47	0.17	0.04	0.24
PotVar0067303 ^a	11	2,780,260	8.20E-08	0.97	0.36	0.15	0.06
PotVar0067483 ^a	11	3,074,626	1.84E-05	0.86	0.26	0.11	0.15
solcap_snp_c1_2314 ^c	11	3,928,601	6.78E-07	1.23	0.18	0.06	0.23
PotVar0106272 ^c	11	4,224,342	8.42E-08	1.32	0.18	0.06	0.23
PotVar0106247 ^c	11	4,227,848	8.52E-08	1.26	0.19	0.08	0.23
PotVar0105904 ^c	11	4,348,897	6.33E-06	1.12	0.18	0.06	0.25

^a SNPs with a high FPR and a high MAF: not haplotype specific

^b SNP with a negative effect

^c SNPs with a lower FPR and a lower MAF: higher haplotype specificity

Phenotypic analysis of resistance in two validation populations

To further validate the significantly associated SNPs, we used two tetraploid populations AxD and KxA segregating for P1 resistance. Both populations were tested for P1 resistance using the Spieckermann and the Glynne-Lemmerzähl methods (Supplementary figure 5, Supplementary file 4). Resistance scores obtained with the

two phenotyping methods were significantly correlated ($R^2 = 0.5$, p -value < 0.0001). With the Spieckermann method, the offspring of Desiree and Kuras were comparable in their transmission and level of resistance to P1 (Kruskal-Wallis test, p -value = 0.16). With the Glynne-Lemmerzahn method, the AxD population showed a higher level of resistance than the KxA population (Kruskal-Wallis test, p -value = 0.023). The Spieckermann method showed transmission of P1 resistance was skewed toward resistance: 81% and 68.6% of the AxD and KxA populations were resistant, respectively. The Glynne-Lemmerzahn method, however, showed a 1:1 segregation ratio (χ^2 test p -value = 0.3) suggesting a monogenic major-effect locus. The P1 segregation analysis with Spieckermann data was flawed due to escapes, and does not warrant to postulate a different genetic model which accommodates for differential recognition of factors from either summer or winter spores.

Genotypic analysis of two validation populations with SNPs markers

KASP markers were designed for SNPs that were significant in the PCoA corrected GWAS model. We chose four of the five SNPs from group 3 that showed a lower MAF, a lower FPR and a higher FNR. The minor allele of these SNPs was present in simplex in Desiree and Kuras and absent from Aventura according to the SolSTW array data (Supplementary file 1). The markers were screened in the two populations and their segregation fitted the expected 1:1 segregation ratio (χ^2 tests, p -values > 0.05) (Supplementary file 4). In the AxD population, none of the markers was associated with the phenotypic data collected with the Spieckermann assays (Table 2). However, all the markers were strongly associated with P1 resistance using the Glynne-Lemmerzahn assays. In the KxA population, the four markers co-segregated with P1 resistance collected with both assays (Figure 3).

Table 2. Marker validation in AxV and KxA offspring

Marker	PGSC ¹ chr11	Aventura × Desiree		Kuras × Aventura	
		Spiecker- mann	Glynne- Lemmerzahn	Spiecker- mann	Glynne- Lemmerzahn
PotVar0067008	2,773,680	0.06	1.14E-04***	1.67E-06***	3.93E-07***
solcap_snp_ c1_2314	3,928,601	0.20	1.87E-03**	2.70E-05***	1.28E-05***
PotVar0106272	4,224,342	0.17	2.12E-03**	1.67E-06***	7.06E-06***
PotVar0105904	4,348,897	0.12	4.53E-03**	2.50E-06***	1.16E-05***

¹ Physical coordinates (in bp) on the potato reference genome (v.4.03)

* : p -value < 0.05

** : p -value < 0.01

*** : p -value < 0.001

Graphical genotyping to identify additional SNPs from the *Sen1* haplotype

PotVar0067008 was fully co-segregating with *Sen1* in the KxA population but there were eight recombinant events between the marker and *Sen1* in the AxD population (Figure 3, Supplementary file 4). To develop KASP markers closer to the NLR clusters C76 and C77 (9), we used the 135K PotVar SNP dataset developed by (24) from which markers on the SolSTW SNP array were selected. By graphical genotyping ((38); operators used: allele frequency between 12 and 24%, absent in Black 1256, Ackersegen, Y 66-13-636 and Yam, but present in Arran Pilot), we could identify a haploblock comprising 54 SNPs including PotVar0067008 (Supplementary figure 6). As this haploblock ranged from 2.70 to 2.78 Mb, we decided to design a KASP assay for a most distal SNP, PotVar0066530 located at 2,706,497 Mb. The marker was indeed co-segregating with resistance but there were still seven recombinants (Figure 3; Supplementary file 4). With operators: MAF>30%, absent in susceptible varieties Bintje, Industrie and Arran Chief, but present in resistant Arran Pilot, 31 SNPs were identified in high LD with PotVar0066243 from coordinates ranging from 2.03 to 2.26 Mb, but no haplotype structure was observed that would allow to develop additional KASP markers to reduce the amount of false positives or false negatives (Supplementary figure 7).

Sen1 distribution and contribution to P1 resistance

In the populations, the presence of the *Sen1* markers significantly increased the resistance to P1. With the Spieckermann phenotyping method, the P1 resistance scores increased from an average of 7.25 to 9.93 in the combined populations when the minor allele of PotVar0066530 was present (Supplementary figure 8). With the Glynne-Lemmerzähl method, the average score decreased from 3.6 to 1.4. In the AxD offspring, 58.8% and 79.2% of the resistant descendants with the Spieckermann and the Glynne-Lemmerzähl methods respectively held the minor allele of PotVar0066530 (Supplementary figure 5). In the KxA population, 78.3% and 100% of the resistant varieties with the Spieckermann and Glynne-Lemmerzähl assays respectively held the minor allele of the marker. The difference between the Spieckermann and the Glynne-Lemmerzähl results observed in both populations might be explained by escapes in the Spieckermann assays. The AxD offspring showed a higher proportion of resistant progeny which did not hold the minor allele compared to the KxA population. This could be explained by a higher number of recombinants between PotVar0066530 and *Sen1* or by the fact that another allele involved in P1 resistance segregated in this population.

Figure 3. Genetic and physical maps comprising the *Sen1* locus

(a) Consensus genetic map of the *Sen1* haplotype in the combined AxD and KxA validation populations (n = 77). The Glynne-Lemmerzahn phenotypic data were used to build the map. (b) Physical map of the chromosome 11 north arm adapted from (13). The KASP markers tested in this study are anchored on the right according to their physical position on DM v4.03 (39) and on the left are the NLR clusters according to (9). In grey are the markers designed in previous mapping studies: (A) (8), (B) (14), (C) (16), (D) (11).

In the GWAS panel, the average P1 resistance score was 7 when the minor allele of PotVar0067008 was absent, while the scores were 9.5, 9.9 and 9.5 when the minor allele was present in simplex, duplex and triplex respectively (Supplementary figure 8), which is in agreement with a dominant effect without additivity. The number of susceptible varieties and the variance of the P1 resistance scores decreased with the higher dosages, which can be explained by the fact that a variety is less likely to inherit

several recombined haplotypes. Overall, the PotVar0067008 minor allele was present in 68% and 16% of the resistant and susceptible varieties of the GWAS panel, respectively. The false negatives were probably not only due to recombination events between *Sen1* and PotVar0067008 but also to other alleles and/or other loci involved in P1 resistance present in the GWAS panel.

The minor allele of PotVar0067008 could be observed in every structure group of the GWAS panel (Supplementary figure 9). It was highly frequent in the 'Agria' and in the 'starch' structure groups: 74.4% and 74.5% of the two groups respectively hold the minor allele. In contrast, only 56.4% and 35.1% of the 'Rest' and the 'GB' groups respectively hold the minor allele of PotVar0067008. Selection for *Sen1* containing varieties was very successful in the starch structure group as almost 20% of the varieties hold the minor allele in duplex.

Other P1 resistance sources are available in breeding germplasm

To identify putative other P1 resistance loci present in the gene pool, we added five uncorrelated, but significantly associated markers from the chromosome 11 peak as a cofactor with fixed effect in the PCoA corrected model (PotVar0067008, PotVar0067303, PotVar0066243, SolCAP_snp_c2_13431 and Potvar0066434). The signal of the chromosome 11 peak was reduced due to the inclusion of these cofactors (Supplementary figure 10). Only one marker, PotVar0067876, remained significant considering the Li and Ji association threshold (Supplementary table 3). This marker was not significant in the two previously applied models ($-\log_{10}(p) = 0.61$ in the naive model, $-\log_{10}(p) = 0.009$ in the PCoA corrected model) and had a relatively high FPR and FNR of 16.36% and 20.3%, respectively. This marker might reveal the existence of another haplotype on chromosome 11 involved in quantitative P1 resistance. This hypothesis would require validation.

We performed a new PCoA on the 144 varieties from the GWAS panel which do not contain the minor allele of PotVar0067008 (Supplementary figure 11). The 'Agria' structure group observed in the entire panel is underrepresented in this PCoA as most (74.4%) of the 'Agria' varieties hold the minor allele of PotVar0067008 (Supplementary figure 7). A separate cluster of 'starch' varieties is observed without PotVar0067008 but resistant to P1 (except for Aventra). This could imply the presence of other wart resistance sources than *Sen1* in the 'starch' group. Likewise, the 'GB' varieties are a clear group, composed of 27 resistant and 23 susceptible varieties. Again, this could imply the presence of other wart resistance sources in this structure group. In the 71

varieties from the 'rest' group, 39 showed resistance to P1. For this group, it is less obvious to make assumptions about other wart resistance sources.

Discussion

Successful identification of *Sen1* using GWAS

Pre-existing phenotypic data on resistance to potato wart disease P1 have been collected from various public and private sources for 330 potato varieties. The phenotypes were associated with previously generated SNP data (25). Analysis of the panel showed four groups: starch varieties, processing varieties descending from Agria, varieties of British origin and a miscellaneous group. Because the panel structure correlated with P1 resistance, we applied a PCoA correction in our GWAS model to correct for the structure confounding effect. The GWAS resulted in one peak of 12 strongly associated SNPs on the north arm of chromosome 11 where the *Sen1* locus was mapped previously. We could divide these associated SNPs into three groups: (1) common SNPs with a high FPR lacking haplotype specificity, (2) one marker with a negative effect on resistance, probably linked in repulsion phase with *Sen1*, and (3) markers showing a low FPR and lower MAF in the panel. The markers from the last group belonged to a resistant haplotype first observed in Pink Fir Apple. We designed KASP markers for four SNPs from the last group and screened them in two independent bi-parental populations. In the AxD population, a 10cM distance was found between *Sen1* and PotVar0066530 (at PGSC coordinate Chr11: 2.706.497), but no recombinant was found in the KxA population. As all recombinants were found among resistant offspring, escapes from the disease assays are likely erroneously regarded as recombination events.

The Glynne-Lemmerzahn method is less ambiguous than the Spieckermann method

Our mapping populations were phenotyped with the two commonly used methods recommended by the EPPO (32). Their main difference is the material used as inoculum: compost with winter sporangia (5) versus fresh warts containing both winter and summer sporangia (6,7). Although both methods correlate well, the phenotypic distribution obtained with the Spieckermann method was skewed toward resistance, suggesting 15 escapes among 77 descendants. Few studies have compared both methods. (40) reported that the Spieckermann method is cheaper and less laborious than the Glynne-Lemmerzahn method, but the lower infection pressure and can lead to

escapes. Escapes can be avoided in the Spieckermann method when more tubers are tested.

***Sen1* probably resides in the C76 or C77 TNL gene cluster**

Our results showed that *Sen1* is located north to PotVar0066530 (2.7 Mb), where the C77 and C76 TNL clusters are located (9). Here, the *Sen3* gene has been mapped as well (12,13). However, speculations about the physical position of *Sen1* relative to the DM reference genome should be taken carefully. The reference genome has two gaps, one of which should contain the unanchored scaffold DMB734 (13). In a first report, (8) mapped *Sen1* distal to NI25 and NI27 which is in agreement with our results (Figure 3). In subsequent reports (14–16), no markers flanking *Sen1* were designed but markers associated with P1 resistance were identified. When the physical position of these markers on the potato reference genome was extrapolated, they pointed to the same region as identified in the current study (Figure 3). In the diploid population in which the gene *Sen2* was identified, *Sen1* was mapped between 1.64 Mb and 3.08 Mb which means *Sen1* would more likely belong to the C77 TNL cluster than to the C76 TNL cluster (9). However, the mapping results of this paper are different from the mapping results of (8) as they mapped *Sen1* proximal to NI25 and NI27. According to (11), this different position could be explained by the different genetic background or by the type of markers which were used in the two different studies.

A historical interpretation of *Sen1* distribution in potato germplasm

The oldest variety from the panel in which PotVar0067008 was observed was Pink Fir Apple (released in 1850, pedigree unknown), but this variety has hardly contributed to the contemporary gene pool. Pink Fir Apple is duplex for PotVar0067008, but this does not imply that this variety is duplex for two identical haplotypes, or two copies of *Sen1*. This study largely illustrates that such historical reconstructions of the breeding history of haploblocks with the *Sen1* resistance locus are complicated. We assume that related varieties or unnamed landraces with resistance should have existed, as PotVar0067008 is also present in the resistant varieties Champion (1876) and Belle De Fontenay (1885), both unrelated to Pink Fir Apple. These findings suggest that *Sen1* was already present in the gene pool before *S. endobioticum* was first described in Europe in the 1880's (41,42). These observations would argue for at least one minor allele being present in early material grown in Europe. However, breeders may have introgressed additional *Sen1* allele(s) or paralogs of *Sen1* along with other *R* genes on the north arm of chromosome 11 such as *Solanum andigena* (*Ry_{adg}*) (43) or *Solanum stoloniferum* CPC 2093 (*Ny(o,n)_{sto}*) (38). Unfortunately, pedigree analysis did not allow

to identify a common ancestral donor for the *Sen1* locus. This is in contrast to the more recently introgressed *Sen3* (13) where all cases of resistance are identical-by-descent.

About 37% of the varieties from the GWAS panel, registered before the 1960's, hold *Sen1*, against 60% of the varieties registered after 1960. These numbers illustrate well how successful the breeders were in selecting for P1 resistance in their breeding programs. Indeed, varieties such as Abundance, Flourball, Irish Cobbler, Jubel and Majestic were used in many crosses as P1 resistant parent (44–46), albeit Irish Cobbler does not contain PotVar0067008. In the GWAS panel, we observed that *Sen1* was especially selected in the starch and processing potatoes programs, where ~75% of the varieties from the 'starch' structure group have *Sen1*, including 20% in two copies. These observations make sense, as starch potatoes are cultivated in regions with narrow crop rotations. *S. endobioticum* is often found in such areas and therefore cultivation of susceptible varieties is prohibited to contain the spreading of the disease (4). In the 'Agria' structure group, 75% of the varieties have *Sen1*: Agria is donor of *Sen1*, and breeders selected P1 resistant progeny. Only ten Agria descendants did not hold the minor allele of PotVar0067008, four of which were resistant (Dione, Excellent, Lady Blanca and Vr-93-146), so we can hypothesize recombination between PotVar0067008 and *Sen1*. Although we showed that PotVar0067008 explained around 69% of the P1 resistance in the GWAS panel, its false negative rate indicates that there are likely other P1 resistance sources already present in the gene pool that remain to be discovered.

No other P1 resistance loci identified with the GWAS

To identify other genes involved in P1 resistance, we performed a GWAS with five *Sen1* associated markers as co-factor. This new GWAS model did not identify new markers outside the *Sen1* region, suggesting that no other locus is present in the potato genome. Neither could we re-discover the *Sen1-4* locus on chromosome 4 (10). One new chromosome 11 marker PotVar0067876 (at coordinate 3,268,823) emerged as significantly associated with wart resistance. The joint analysis of false positives and false negatives showed nevertheless that PotVar0067876 could not be used to recover false positives or negative predictions from our best marker PotVar0067008. The minor allele of PotVar0067876 was present in simplex in Desiree and Avenra and in duplex in Kuras. The expected segregation ratio would be 3:1 and 11:1 in the AxD and KxA populations respectively. Therefore, the size of our full-sib populations was too small to (in)validate the significance of the effect of PotVar0067876 on P1 resistance. Another strategy to identify new resistances in the GWAS panel could have been to

perform a GWAS in each of the four structure groups. However, the size of these groups did not allow us to follow such a strategy.

Other P1 resistance loci remain to be identified in the gene pool

The large proportion of false negatives was explored by visual interpretation of a PCoA of varieties without the minor allele of PotVar0067008 (Supplementary figure 11). These groups share a specific breeding history along with being frequently resistant to P1. One easy conclusion is to assume that within each group, another recombined *Sen1* haplotype or source (founder) of resistance was used, having an insufficient allele frequency to pass the MAF threshold or to achieve the statistical power for discovery by GWAS. Specifically, the group of starch varieties was distinguished from the sub-panel. This group contained only Aventura, Festien and Karakter as P1 susceptible varieties. Some of the varieties from this group are known to contain also resistance to other wart pathotypes. This is the case of Allure, Astarte, Avaya, Aviala, and Merano (47,48). Possibly, these varieties are resistant to P1 due to other sources than *Sen1*. E.g. Merano has the *Sen3* gene which gives resistance to a broader range of pathotypes (12).

Another interesting observation was that among the 77 varieties from the 'GB origin' structure group, only 27 held the minor allele of PotVar0067008. In the remaining 50 varieties, 27 were resistant to P1. It is very likely that (an)other source(s) of P1 resistance, not yet described, are present in the 'GB' structure group.

For the structure group of varieties coming from diverse origins (rest), it is difficult to make assumptions about other resistance sources to P1. For Escort, Panda, Royal, and Ulme, which are known to be resistant to other pathotypes (48), we can again propose that resistance to P1 is due to another gene conferring resistance to a broader range of pathotypes (such as *Sen3* in Ulme, (12,13)). For other varieties, which are only resistant to P1, the resistance could still come from *Sen1* if they recombined between PotVar0067008 and *Sen1* or from another locus not yet identified.

Limitations to the identification of markers with complete diagnostic value

PotVar0067008 is the most closely linked marker to *Sen1* that we identified with the GWAS. However, its diagnostic value was not optimal as it showed a false positive rate of 4% and a false negative rate of 24%. An explanation for false positives can be recombinations which occurred in some varieties between the marker and *Sen1* or background dependency of the *Sen1* gene. The explanation for false negatives can be

recombinations between the marker and *Sen1* as well, but also the presence of other alleles or other loci bringing resistance to P1 in the panel. With the genotypic data we used in the GWAS, we could not identify markers with a better diagnostic value than PotVar0067008 or PotVar0066530 for which recombinations with *Sen1* have been observed in the bi-parental populations.

When comparing the physical coordinates of the most significant markers from our GWAS study with genetic positions of *Sen1* from literature and from our validation population, we noticed that we did not achieve to identify strongly associated SNPs north of PotVar0066530 and *Sen1*. Was our GWAS analysis hampered by ascertainment bias or by insufficient numbers of markers, or a strong decay of LD? Ascertainment bias is excluded because the SNP discovery panel includes 83 varieties of which many have *Sen1*. When evaluating marker sufficiency and their distribution, it appeared that 3,104 SNP markers have been described in the first 4 Mb of chromosome 11 (24) and that a subset of 288 SNPs were placed on the SolSTW array (25). Their distribution depends on the baits designed for targeted resequencing and is shown in Supplementary figure 12. It appears that no markers from the dataset of (24) were designed between 814,888 bp and 1,416,133 bp meaning that we could not identify markers flanking north of the TNL cluster C76 (located between 1,333,729 bp and 1,589,104 bp, (9)). No markers were designed either between 1,417,043 bp and 1,952,061 bp hampering the design of diagnostic markers closer to *Sen1* (Supplementary figure 12). Moreover, from a region at 1.42 Mb, comprising 146 SNPs, only four SNPs were selected for the SNP array. This uneven coverage of designed SNP markers in the *Sen1* region in the dataset from (24) could be the primary reason that we did not identify SNP markers with better association than PotVar0067008. By graphical genotyping, we also observed that no marker specific to the *Sen1* haplotype could be identified north to 2,7 Mb. The lack of haplotype specificity of the markers north to PotVar0066530 hampered the identification of markers with complete diagnostic value.

Conclusion

This study illustrates that pre-existing genotypic data and a collection of historical phenotypic data, allowed to screen potato germplasm for genes involved in P1 resistance and to identify the haplotype specific marker PotVar0067008 linked to *Sen1*, with a false positive rate of only 4%. Because this SNP was identified using a GWAS approach, we know that the utility of this KASP marker can be extrapolated to a wide genetic background. We conclude that the *Sen1* locus, first published by (8), is the

most frequent source of P1 resistance, but other less frequent alleles or loci remain to be identified. In general, marker-assisted breeding demands for SNPs without false positives and false negatives. This paper shows reasons why this is not always feasible. First, we could not trace founders of *Sen1* and suspect a lack of identity-by-descent of *Sen1* bearing haplotypes. Second, the raw marker data describes more than a century of potato breeding and clones with recombinant haplotypes were identified, which illustrates the process of decay of LD between flanking markers and *Sen1*.

Author contribution statement

CP, PGV, JHV and HJvE conceived and designed the experiment. CP performed the experiments and data analyses with the help of JET. MJP wrote the Genstat scripts and provided statistical support. CP interpreted all the results with the help of JHV and HJvE. CP wrote the manuscript. JHV, RGFV and HJvE helped to write, improve and review the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank the members of this public-private partnership (Averis seeds BV, Böhm-Nordkartoffel Agrarproduktion GmbH & Co. OHG, Danespo, HLB BV, HZPC Holland BV, C Meijer BV, SaKa Pflanzenzucht GmbH & Co. KG and Teagasc) for providing phenotypic data, KASP marker analysis and scientific discussions. Jarosław Przetakiewicz (IHAR) and Margriet Boerma (HLB BV) are gratefully acknowledged for phenotyping the AxD and the AxK population and several varieties.

Funding

This project “An integrated genomics and effectoromics impulse for potato wart resistance management and breeding” was financially supported by the Dutch Topsector Horticulture & Starting Materials. Within the Topsector, private industry, knowledge institutes and the government are working together on innovations for sustainable production of safe and healthy food and the development of a healthy green environment. CP was supported by a Walsh fellowship provided by Teagasc.

Conflict of interest

The authors declare they have no conflicts of interest.

References

1. Przetakiewicz J. The Viability of Winter Sporangia of *Synchytrium endobioticum* (Schilb.) Perc. from Poland. *Am J Potato Res.* 2015 Dec;92(6):704–8.
2. EPPO. Specific quarantine requirements. EPPO Technical Documents. EPPO Technical Documents No. 1008. 1990;
3. Obidiegwu JE, Flath K, Gebhardt C. Managing potato wart: a review of present research status and future perspective. *Theor Appl Genet.* 2014 Apr;127(4):763–80.
4. Baayen RP, Cochius G, Hendriks H, Meffert JP, Bakker J, Bekker M, *et al.* History of potato wart disease in Europe – a proposal for harmonisation in defining pathotypes. *Eur J Plant Pathol.* 2006 Sep;116(1):21–31.
5. Spieckermann A, Kothoff P. Testing potatoes for wart resistance. *Deutsche Landwirtschaftliche Presse* 51:114–115. 1924;
6. Glynne MD. Infection experiments with wart disease of potatoes. *Synchytrium endobioticum* (Schilb.) Perc. *Ann Appl Biol.* 1925;12(1):34–60.
7. Lemmerz J. A new simplified infection procedure for testing potato cultivars for wart resistance. *Zücher.* 1930;2(288):97.
8. Hehl R, Faurie E, Hesselbach J, Salamini F, Whitham S, Baker B, *et al.* TMV resistance gene *N* homologues are linked to *Synchytrium endobioticum* resistance in potato. *Theor Appl Genet.* 1999;98(3):379–386.
9. Jupe F, Witek K, Verweij W, Śliwka J, Pritchard L, Etherington GJ, *et al.* Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. *Plant J.* 2013 Nov;76(3):530–44.
10. Brugmans B, Hutten RGB, Rookmaker ANO, Visser RGF, van Eck HJ. Exploitation of a marker dense linkage map of potato for positional cloning of a wart disease resistance gene. *Theor Appl Genet.* 2006 Jan;112(2):269–77.
11. Plich, J., Przetakiewicz, J., Śliwka, J., Flis, B., Wasilewicz-Flis, I., Zimnoch-Guzowska, E., 2018. Novel gene *Sen2* conferring broad-spectrum resistance to *Synchytrium endobioticum* mapped to potato chromosome XI. *Theor. Appl. Genet.* <https://doi.org/10.1007/s00122-018-3154-y>
12. Bartkiewicz, A., Chilla, F., Terefe-Ayana, D., Lübeck, J., Strahwald, J., Tacke, E., Hofferbert, H.-R., Flath, K., Linde, M., Debener, T., 2018. Improved genetic resolution for linkage mapping of resistance to potato wart in monoparental dihaploids with potential diagnostic value in tetraploid potato varieties. *Theor. Appl. Genet.* 131, 2555–2566. <https://doi.org/10.1007/s00122-018-3172-9>
13. Prodhomme C, Esselink D, Borm T, Visser RGF, van Eck HJ, Vossen JH. 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* [Internet]. 2019 Dec [cited 2019 Jun 12];15(1). Available from: <https://plantmethods.biomedcentral.com/articles/10.1186/s13007-019-0445-5>
14. Ballvora A, Flath K, Lübeck J, Strahwald J, Tacke E, Hofferbert H-R, *et al.* Multiple alleles for resistance and susceptibility modulate the defense response in the interaction of tetraploid potato (*Solanum tuberosum*) with *Synchytrium endobioticum* pathotypes 1, 2, 6 and 18. *Theor Appl Genet.* 2011 Dec;123(8):1281–92.
15. Groth J, Song Y, Kellermann A, Schwarzfischer A. Molecular characterisation of resistance against potato wart races 1, 2, 6 and 18 in a tetraploid population of potato (*Solanum tuberosum* subsp. *tuberosum*). *J Appl Genet.* 2013 May;54(2):169–78.
16. Obidiegwu JE, Sanetomo R, Flath K, Tacke E, Hofferbert H-R, Hofmann A, *et al.* Genomic architecture of potato resistance to *Synchytrium endobioticum* disentangled using SSR

- markers and the 8.3k SolCAP SNP genotyping array. BMC Genet [Internet]. 2015 Dec [cited 2017 Aug 3];16(1). Available from: <http://www.biomedcentral.com/1471-2156/16/38>
17. Gebhardt C, Bellin D, Henselewski H, Lehmann W, Schwarzfischer J, Valkonen JPT. Marker-assisted combination of major genes for pathogen resistance in potato. Theor Appl Genet. 2006 May;112(8):1458–64.
 18. Przetakiewicz J, Plich J. Assessment of Potato Resistance to *Synchytrium Endobioticum*. Plant Breed Seed Sci. 2017 Dec 20;76(1):37–43.
 19. Antonova OY, Shvachko NA, Novikova LY, Shuvalov OY, Kostina LI, Klimenko NS, *et al*. Genetic diversity of potato varieties bred in Russia and its neighboring countries based on the polymorphism of SSR-loci and markers associated with resistance *R*-genes. Russ J Genet Appl Res. 2017 Jul;7(5):489–500.
 20. Bandillo N, Jarquin D, Song Q, Nelson R, Cregan P, Specht J, *et al*. A Population Structure and Genome-Wide Association Analysis on the USDA Soybean Germplasm Collection. Plant Genome. 2015;8(3):0.
 21. Zhao K, Tung C-W, Eizenga GC, Wright MH, Ali ML, Price AH, *et al*. Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. Nat Commun. 2011 Sep 13;2:467.
 22. Sukumaran S, Dreisigacker S, Lopes M, Chavez P, Reynolds MP. Genome-wide association study for grain yield and related traits in an elite spring wheat population grown in temperate irrigated environments. Theor Appl Genet. 2015 Feb;128(2):353–63.
 23. D'hoop BB, Keizer PLC, Paulo MJ, Visser RGF, van Eeuwijk FA, van Eck HJ. Identification of agronomically important QTL in tetraploid potato cultivars using a marker–trait association analysis. Theor Appl Genet. 2014 Mar;127(3):731–48.
 24. Uitdewilligen JGAML, Wolters A-MA, D'hoop BB, Borm TJA, Visser RGF, van Eck HJ. A Next-Generation Sequencing Method for Genotyping-by-Sequencing of Highly Heterozygous Autotetraploid Potato. Lukens L, editor. PLoS ONE. 2013 May 8;8(5):e62355.
 25. Vos PG, Uitdewilligen JGAML, Voorrips RE, Visser RGF, van Eck HJ. Development and analysis of a 20K SNP array for potato (*Solanum tuberosum*): an insight into the breeding history. Theor Appl Genet. 2015 Dec;128(12):2387–401.
 26. Felcher KJ, Coombs JJ, Massa AN, Hansey CN, Hamilton JP, Veilleux RE, *et al*. Integration of Two Diploid Potato Linkage Maps with the Potato Genome Sequence. Aerts J, editor. PLoS ONE. 2012 Apr 27;7(4):e36347.
 27. Hamilton JP, Hansey CN, Whitty BR, Stoffel K, Massa AN, Van Deynze A, *et al*. Single nucleotide polymorphism discovery in elite North American potato germplasm. BMC Genomics. 2011;12(1):302.
 28. VanRaden PM. Genomic measures of relationship and inbreeding. INTERBULL Bull. 2007;(37):33.
 29. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155(2):945–959.
 30. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol. 2005 Jul;14(8):2611–20.
 31. van de Vossen BTLH, Brankovics B, Nguyen HDT, van Gent-Pelzer MPE, Smith D, Dadej K, *et al*. The linear mitochondrial genome of the quarantine chytrid *Synchytrium endobioticum*; insights into the evolution and recent history of an obligate biotrophic plant pathogen. BMC Evol Biol [Internet]. 2018 Dec [cited 2019 Apr 25];18(1). Available from: <https://bmcevolbiol.biomedcentral.com/articles/10.1186/s12862-018-1246-6>
 32. EPPO. *Synchytrium endobioticum*: Diagnostic protocols. EPPO Bull. 2004 Aug;34(2):213–8.
 33. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity. 2005;95(3):221.

34. Winkler TW, Kutalik Z, Gorski M, Lottaz C, Kronenberg F, Heid IM. EasyStrata: evaluation and visualization of stratified genome-wide association meta-analysis data. *Bioinformatics*. 2015 Jan 15;31(2):259–61.
35. Fulton TM, Chunwongse J, Tanksley SD. Microprep protocol for extraction of DNA from tomato and other herbaceous plants. *Plant Mol Biol Report*. 1995;13(3):207–209.
36. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press Totowa NJ Pp365–386. 2000;
37. Vos PG, Paulo MJ, Voorrips RE, Visser RGF, van Eck HJ, van Eeuwijk FA. Evaluation of LD decay and various LD-decay estimators in simulated and SNP-array data of tetraploid potato. *Theor Appl Genet*. 2017 Jan;130(1):123–35.
38. van Eck HJ, Vos PG, Valkonen JPT, Uitdewilligen JGAML, Lensing H, de Vetten N, *et al*. Graphical genotyping as a method to map *Ny(o,n)sto* and *Gpa5* using a reference panel of tetraploid potato cultivars. *Theor Appl Genet*. 2017 Mar;130(3):515–28.
39. Potato Genome Sequencing Consortium, Pan S, Cheng S, Zhang B, Mu D, Ni P, *et al*. Genome sequence and analysis of the tuber crop potato. *Nature*. 2011 Jul 10;475(7355):189–95.
40. Przetakiewicz J, Kopera K. The comparison of usefulness the Glynne-Lemmerzähl and the Spieckermann methods to assess resistance of potato (*Solanum tuberosum* L.) to *Synchytrium endobioticum* (Schilb.) Perc. pathotype 11(D1) in mass tests. (in Polish). *Biuletyn IHAR*. 2007;243:235–44.
41. Gough GC. Wart disease of potatoes (*Synchytrium endobioticum* Perc.). A study of its history, distribution and the discovery of immunity. *J R Hortic Soc*. 1919;45:310–2.
42. Hampson M. History, biology, and control of potato wart disease in Canada. *Can J Plant Pathol*. 1993 Dec;15(4):223–44.
43. Hämäläinen JH, Watanabe KN, Valkonen JPT, Arihara A, Plaisted RL, Pehu E, *et al*. Mapping and marker-assisted selection for a gene for extreme resistance to potato virus Y: *Theor Appl Genet*. 1997 Feb;94(2):192–7.
44. Black W. Studies on the inheritance of resistance to wart disease (*Synchytrium endobioticum* (Schilb.) Perc.) in potatoes. *J Genet*. 1935;30(1):127–146.
45. Lunden AP, Jørstad I. Investigations on the inheritance of immunity to wart disease (*Synchytrium Endobioticum* [Schilb.] Perc.) in the potato. *J Genet*. 1934;29(3):375–385.
46. Weiss F. The Conditions of Infection in Potato Wart. *Am J Bot*. 1925 Jul;12(7):413.
47. Baayen RP, Bonthuis H, Withagen JCM, Wander JGN, Lamers JL, Meffert JP, *et al*. Resistance of potato cultivars to *Synchytrium endobioticum* in field and laboratory tests, risk of secondary infection, and implications for phytosanitary regulations. *EPPO Bull*. 2005;35(1):9–23.
48. NVWA. Teeltvoorschrift wratziekte [Internet]. 2018 [cited 2019 Mar 27]. Available from: <https://www.nvwa.nl/onderwerpen/teeltvoorschriften-akkerbouw-en-tuinbouw/teeltvoorschrift-wratziekte>
49. Maris B. Races of the potato wart causing fungus *Synchytrium endobioticum* (Schilb.) Perc. and some data on the inheritance of resistance to race 6. *Euphytica*. 1961;10(3):269–276.
50. Borthwick AW. Note On Warty Disease Of The Potato. *Trans Bot Soc Edinb*. 1908 Jan;23(1–4):279–279.
51. Salaman RN, Lesley JW. Genetic studies in potatoes; the inheritance of immunity to wart disease. *J Genet*. 1923;13(2):177–186.

Supplementary figures

Supplementary figure 1. Distribution of pathotype 1 resistance in the GWAS panel

Distribution of the pathotype 1 resistance scores in the GWAS panel (right stacked histogram). Colours are given in function of the year of release of the variety. The scores ranged from 0 to 11 as we calculated for each variety a mean score corrected for the source effect. For each resistance category ([0-2], [2-4], [4-6], [6-8], [8-11]), the varieties origin is given (left table). AUT = Austria, BRD = Bundesrepublik Deutschland (West Germany), CAN = Canada, DDR = Deutsche Demokratische Republik (East Germany), DEN = Denmark, ESP = Spain, FRA = France, GB = Great Britain, GER = Germany, HOL = Netherlands, IRL = Ireland, POL = Poland, SWE = Sweden.

Supplementary figure 2. Distribution of pathotype 1 resistance in the structure groups

Boxplots of the distribution of the pathotype 1 resistance scores in function of the four structure groups identified with the PCoA and delimited with the Structure analysis.

Supplementary figure 3. Distribution of the residues of the GWAS models

Three GWAS models have been tested: a naive model with no correction (1), a corrected model with 30 PCOs included as a fixed effect (2) and a corrected model with 30 PCOs and 5 co-factors (PotVar0067008, PotVar0067303, PotVar0066243, solcap_snp_c2_13431, PotVar0066434) as fixed effects (3). In (a) are shown the histograms of the residues distribution and in (b) the qq-plots. When including the chromosome 11 uncorrelated most significant markers, the distribution of the residues is very close to normality.

Supplementary figure 4. QQ-plot of the GWAS results

Quantile-quantile plot (qq-plot) of the GWAS p -values ($-\log_{10}(p)$) with (green) or without (blue) PCoA correction. The qq-plot shows the observed distribution of association test statistics compared to the expected distribution of associations among the 10,968 SNPs used in the analysis. Including a PCoA correction as a fixed effect in the GWAS model decreases the deviation from the $X = Y$ line (red) for the lower part of the curve, decreasing the chance to find false associations due to the structure confounding factor.

Supplementary figure 5. Distribution of P1 resistance in the AxD and KxA populations

Distribution of the resistance means (2016 + 2017) scored with the Glycine-Lemmerzah method (a and b) and with the Spieckermann method (c and d) in the AxD (a and c) and the KxA (b and d) populations. Black bars represent the number of individuals holding the minor allele of PotVar0066530. The threshold between R and S descendants is represented with a dashed line.

Supplementary figure 8. Effect of the *Sen1* markers on pathotype 1 resistance

(a) Effect of PotVar0067008 on P1 resistance in the GWAS panel. (b) Effect of PotVar0066530 in the validation populations on P1 resistance assessed with the Spieckermann method (average 2016 + 2017) and (c) with the Glynne-Lemmerzähl method (average 2016 + 2017). 0, 1, 2 and 3 represent the dosage of the marker (absent, simplex, duplex and triplex respectively).

Supplementary figure 9. Presence of PotVar0067008 in the structure groups

(a) Dosage of PotVar0067008 minor allele in function of the structure groups and (b) proportion of each structure group with the minor allele of PotVar0067008 in nulliplex, simplex, duplex or triplex.

Supplementary figure 10. GWAS of pathotype 1 resistance including 5 co-factors

Miami plot of the GWAS of pathotype 1 resistance using 330 potato varieties. The PCoA correction and five chromosome 11 markers as cofactors is shown in the upper part of the plot (in orange). The PCoA correction alone is shown in the lower part of the plot (in green). The association ($-/+ \log_{10}(p)$) of each SNP with pathotype 1 resistance is represented. The x axis represents the physical position of each SNP on the 12 potato chromosomes (u corresponds to unanchored and chloroplastic markers). The red horizontal lines correspond to the significance threshold calculated with the Li and Ji method ($-\log_{10}(p) = 4.7$).

Supplementary figure 11. PCoA performed on the sub-panel of varieties which do not hold *Sen1*

The structure of the sub-panel of 144 varieties, which do not hold the minor allele of PotVar0067008. Still a large proportion of the varieties are resistant to pathotype 1. The PCoA was performed with a random subset of 1,000 SNPs. Each variety is represented by a pseudo-colour that scales with the level of resistance to pathotype 1 (dark colour (P1 predicted mean < 6) indicates highly susceptible varieties). The shapes of the dots indicate the structure group affiliation of each variety.

Supplementary figure 12. Physical distribution of PotVar SNPs in the first 4 Mb of chromosome 11

The volume of 3,104 SNPs identified by targeted resequencing is shown in blue (24). Only 288 SNPs (shown in red) were deployed on the SolSTW SNP array (25). The position of candidate *R* gene clusters are indicated with green boxes (9).

Supplementary tables

Supplementary table 1. Re-phenotyping of 12 varieties for which preliminary analysis showed inconsistencies between the marker’s prediction and the phenotype

Variety	Spieckermann	Glynne-Lemmerzahl
Agria	10	9.2
Belana	10	8.8
Bellanova	10	8.8
Berber	10	8.2
Caruso	10	9.2
Cherie	10	9.8
Elisabeth	10	9.6
Forza	9.8	9.8
Kastelli	10	9.7
Lady Jo	10	9
Laura	10	9.2
Starga	10	8.6
Deodara	6.9	2.2

Supplementary table 2. Primers of the KASP markers used to validate the GWAS results in the two populations

Marker	Forward primer (R allele) (5'- 3')	Forward primer (S allele) (5'- 3')	Reverse primer (5'- 3')
PotVar0066530	ATATTTTAGTGTGTTTC TCTGTGCTGC	ATATTTTAGTGTGTTT CTCTGTGCTGG	GACAAAGAAAGCAGTAG CACTTTCTCATT
PotVar0067008	GGCATTAAATTCCTTAA AACTCTCAGCG	GCATTAAATTCCTTAA AACTCTCAGCA	GGAAAGCTTTTTGTTAC AAGGTGGAGAT
solcap_snp_c1_2314	CTTCATCTCTTTCTTTCA TAACTCTAGC	TCATCTCTTTCTTTTCAT AACTCTAGT	GATGAGGAGATTTTGCAA CTTAAGGATCTA
PotVar0106272	CCAATTCACCAGGAAG AACCAATCT	CAATTCACCAGGAAGA ACCAATCG	ACTTCACGGGAAATTCAA ACTGCTGTT
PotVar0105904	GAGCAAGTAAGCCAAT GCCAATTCAT	AGCAAGTAAGCCAAT GCCAATTCAG	GCCTAGTACTTCTCGTGT TGTTTTCAAAA

Supplementary table 3. Significantly associated marker in the GWAS structure and cofactors corrected model results

SNP	Chr	Pos (bp)	p-value	Effect	MAF	FPR	FNR
PotVar0067876	11	3,268,823	9.14E-06	1.08	0.23	0.16	0.20

Supplementary files list

Supplementary files can be accessed at <https://doi.org/10.17026/dans-2zf-pnd3>

> Supplementary file 1 (Chapter2_S_file_1.xlsx)

First sheet: Overview of the 330 potato varieties of our GWAS panel with their year of market release, their country of origin, the structure group to which the variety was assigned, the phenotypic score obtained from the data in Supplementary file 3. This score was used to classify varieties into resistant or susceptible. The significantly associated markers are shown with their coordinates (PGSC v4.03 (39)) and their allele dosage. Our interpretation on concordance between marker PotVar0067008 and resistance is shown in the last columns where false positives are susceptible varieties which hold the minor allele of PotVar0067008 and false negatives are resistant varieties which do not hold the minor allele of PotVar0067008.

Second sheet: Genotypic data input used in the GWAS. Markers randomly selected with replacement for the van Raden kinship calculation are indicated in column B.

> Supplementary file 2 (Chapter2_S_file_2.xlsx)

Structure group membership of each variety according to the STRUCTURE results for $K = 4$. These results were used to describe the distribution of P1 resistance and the distribution of Sen1 in the different groups.

> Supplementary file 3 (Chapter2_S_file_3.xlsx)

Overview of the historical sources that provided data on wart disease resistance in the 330 potato varieties used in our GWAS panel. These historical sources include documents for National listing where VCU data were recorded, marketing brochures from commercial breeders as well as scientific publications. Many sources use different scales to describe the level of wart disease resistance. Qualitative data R and S were converted into 10 and 1 respectively. This allowed compilation of the next tab sheet showing all records used to calculate a single phenotypic score using REML.

> Supplementary file 4 (Chapter2_S_file_4.xlsx)

Phenotypic and genotypic observations collected across years and assays from the offspring of the AxD and KxA validation populations, along with the Mendelian classification into R and S for linkage analysis and map construction.

Chapter 3

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*

Charlotte Prodhomme, Danny Esselink, Theo Borm,
Richard G. F. Visser, Herman J.
van Eck, Jack H. Vossen

This chapter was published with minor modifications in
Plant Methods 2019.

Abstract

Background

Standard strategies to identify genomic regions involved in a specific trait variation are often limited by time and resource consuming genotyping methods. Other limiting prerequisites are the phenotyping of large segregating populations or of diversity panels and the availability and quality of a closely related reference genome. To overcome these limitations, we designed efficient Comparative Subsequence Sets Analysis (CoSSA) workflows to identify haplotype specific SNPs linked to a trait of interest from Whole Genome Sequencing data.

Results

As a model, we used the resistance to *Synchytrium endobioticum* pathotypes 2, 6 and 18 that co-segregated in a tetraploid full sib population. Genomic DNA from both parents, pedigree genotypes, unrelated potato varieties lacking the wart resistance traits and pools of resistant and susceptible siblings were sequenced. Set algebra and depth filtering of subsequences (k -mers) were used to delete unlinked and common SNPs and to enrich for SNPs from the haplotype(s) harboring the resistance gene(s). Using CoSSA, we identified a major and a minor effect locus. Upon comparison to the reference genome, it was inferred that the major resistance locus, referred to as *Sen3*, was located on the north arm of chromosome 11 between 1,259,552 bp and 1,519,485 bp. Furthermore, we could anchor the unanchored superscaffold DMB734 from the potato reference genome to a syntenous interval. CoSSA was also successful in identifying *Sen3* in a reference genome independent way thanks to the *de novo* assembly of paired end reads matching haplotype specific k -mers. The *de novo* assembly provided more R haplotype specific polymorphisms than the reference genome corresponding region. CoSSA also offers possibilities for pedigree analysis. The origin of *Sen3* was traced back until Ora. Finally, the diagnostic power of the haplotype specific markers was shown using a panel of 56 tetraploid varieties.

Conclusion

CoSSA is an efficient, robust and versatile set of workflows for the genetic analysis of a trait of interest using WGS data. Because the WGS data are used without intermediate reads mapping, CoSSA does not require the use of a reference genome. This approach allowed the identification of *Sen3* and the design of haplotype specific, diagnostic markers.

Keywords

Bulked Segregant Analysis (BSA), Whole Genome Sequencing (WGS), reference genome, *k*-mers, diagnostic markers, potato wart disease.

Introduction

The identification of genetic loci involved in important traits variation is a fundamental step in genetic research and breeding. QTL mapping strategies are often time-consuming and laborious. Often, large biparental populations ($n > 300$) segregating for the trait of interest are used, which necessitates considerable efforts for genotyping and phenotyping. Genome-Wide Association Studies (GWAS) are alternatives but equally require genotyping and phenotyping of a broad panel of individuals. Bulk Segregant Analysis (BSA) (1,2), simplify the genotyping effort as pools of individuals with contrasting phenotypes are used, which drastically reduces the number of samples to be genotyped. BSA consists of estimating and comparing allelic frequencies between the pools. The frequency of non-linked loci for the trait of interest are expected to be equivalent between the pools whereas a bias in the frequency of the loci linked to the causal genes is expected. This way, DNA sequence variants linked to the trait of interest are identified, which allows the development of markers for genetic mapping and Marker Assisted Selection (MAS) purposes.

A conventional BSA approach uses a set of formerly identified markers (2,3) such as SNP-arrays (4). Using markers not specifically developed for the population of interest can lead to ascertainment bias resulting in a limited number of markers, which may result in low mapping resolution. In polyploids, ascertainment bias is even a more severe problem, as markers in linkage disequilibrium are less informative. As sequencing became more affordable, genotyping-by-sequencing (GBS) methods, such as whole genome sequencing (WGS) (5), RNA-Seq (6) or different complexity reduction strategies (7,8) have been applied in combination with BSA. These BSA methods, collectively referred to as mapping by sequencing (MBS) (9), involve the successive extraction, pooling and sequencing of the DNA of the pooled individuals. The reads obtained from these bulks are mapped to a reference genome and variants are called in each bulk. The allelic frequencies of the variants are compared between the bulks and variants in linkage disequilibrium with the locus encoding the trait of interest are identified. When WGS data are used, the resolution is increased and it becomes possible to not only identify polymorphisms in linkage disequilibrium but also in the causal locus itself (5).

Methods using read mapping rely on the availability of a reference genome. Nowadays, many crops have reference genomes available, but this is not yet the case for some non-model species or complex polyploid species. Even if a reference genome is available for the crop of interest, the assembly quality or completeness of the reference can be sub-optimal. Moreover, the genotype used as a reference can be considerably different from the genetic background of the population of interest. The wheat pan-genome was studied by comparing whole genome sequencing data of 18 different cultivars (10). Interestingly, only 64% of the genes were present in all the cultivars and the rest showed presence-absence variations between them. This study illustrates why reference genome based approaches for genetic analysis and crop improvement can cause limited and incomplete results. This limitation is even more prominent in the case of disease resistance genes, which reside in clusters of highly similar paralogous sequences inherently difficult to assemble and highly variable in organization. Moreover, the genotype used as a reference is often susceptible to the disease and lacks the causal *R* gene paralog segregating in a mapping population. This is exemplified by the case of the fine-mapping and candidate gene analysis of a powdery mildew resistance QTL in barley (11). After an exome capture experiment was performed using homozygous recombinants, reads were mapped to the barley reference genome. Surprisingly, heterozygous sites were observed in genomic regions containing *R* genes. This could be explained by assuming the reference genome is missing some paralogs of an *R* gene cluster. The reads from the recombinant lines coming from these missing paralogs wrongly mapped to similar paralogs. In this case, a *de novo* assembly of the wrongly mapped reads was needed to identify the candidate genes.

Working directly with WGS reads without alignment to a reference offers several advantages. The use of subsequences (*k*-mers of *k* nucleotides) instead of the full reads allows to bypass sequencing errors in the reads and allows unambiguous comparisons between samples thanks to their invariable length. A notable example of the use of *k*-mers is provided in a study aiming to identify variants associated with human diseases (12). Here, the authors identified several disease-associated sequences that were not present in the human reference genome. A similar approach was applied in a panel of wheat accessions (13). The authors developed a new method, AgRenSeq, which combines resistance gene enrichment (RenSeq) and a *k*-mer based GWAS. They phenotyped an *Aegilops tauschii* panel with different stem rust races and intended to identify stem rust resistance (*Sr*) genes. AgRenSeq was successfully

applied to identify and clone four functional *R* genes independently from a reference genome in a diversity panel.

In this paper, we describe a set of workflows including a BSA approach using *k*-mers instead of variants. The Comparative Subsequence Sets Analysis (CoSSA) is suitable to quickly identify haplotype specific SNPs to develop diagnostic markers linked to traits of interest, with or without a reference genome. As an example, we used a tetraploid potato (*Solanum tuberosum*) population segregating for potato wart disease resistance.

Potato wart disease is caused by the obligate biotrophic soil-borne *Synchytrium endobioticum* Chytridiomycete fungus, which can result in dramatic yield losses. This fungus produces spores that can persist in the soil for more than 40 years (14). In addition, no chemicals are available to control the pathogen. For these reasons, quarantine regulations are imposed (15), forbidding the cultivation of potatoes on infested fields. Moreover, in protection zones around the infested fields, only resistant varieties may be grown. Therefore, resistant varieties are an essential key to manage the disease. The *Sen1* locus providing resistance to pathotype 1 isolates has been mapped previously on the north arm of chromosome 11 (16). Most potato varieties of the European gene pool are resistant to pathotype 1, but only a few are resistant to pathotypes 2, 6 and 18, all of which are occasionally detected in most European countries. Until recently, other QTLs giving resistance to various patterns of pathotypes have been mapped in diploid (17,18) and tetraploid populations (19–21). However, markers with diagnostic value were not found frequently. Only recently, a combined approach of linkage mapping, BSA and RNAseq data in a dihaploid potato population derived from the tetraploid variety Karolin allowed the identification of markers with diagnostic value linked to pathotypes 6 and 18 resistance on the north arm of chromosome 11 (22).

In this study, we applied the CoSSA workflows to map the *Synchytrium endobioticum* pathotype 2, 6 and 18 resistance from the variety Kuba, and refer to it as *Sen3*. CoSSA was implemented with and without the use of the potato reference genome and was successful in both cases. *Sen3* was fine-mapped and haplotype specific markers close to the resistance gene were developed and validated for Marker Assisted Selection. CoSSA also allowed us to improve the potato reference genome that is poorly assembled in the region of interest. Different aspects and further applications of CoSSA, such as pedigree analysis, are described in this paper and compared to earlier

approaches. We found that *Sen3* descended from the variety Ora and was indistinguishable from the resistance in the variety Karolin (21,22).

Materials and Methods

Plant material and wart disease resistance phenotyping

The potato variety Kuba (resistant to the pathotypes 1, 2, 6, and 18 (23)) was crossed to the susceptible variety Ludmilla (resistant to the pathotype 1 only (23)), resulting in a full-sib tetraploid population (KxL) of 328 clones segregating for the pathotypes 2, 6 and 18 resistance (abbreviated as P2, P6 and P18 respectively).

Progeny clones and both parents were phenotyped in spring 2016 for potato wart disease resistance with the Glynne-Lemmerzähl (24,25) method. For each progeny clone, five tubers were tested for pathotypes 2, 6 and 18 resistance. The isolates used for inoculation were respectively JKI P2(G1)-2009, JKI P6(O1)-2009 and JKI P18(T1)-2009. Resistance levels were scored from 1 (highly resistant, early defense necrosis) to 5 (highly susceptible, formation of small to large warts). Mean scores were calculated for each progeny clone for each phenotypic season (Supplementary file 1). Chi-square tests were performed to analyze the resistance segregation ratio to the three tested pathotypes in the population.

To compose the susceptible bulk (S-bulk), 17 progeny clones fully susceptible to P2(G1), P6(O1) and P18(T1) in 2016 were selected. To compose the resistant bulk (R-bulk), 17 progeny clones that were fully resistant to the three pathotypes in 2016 were selected and were re-phenotyped (six tubers) in 2017. The three isolates and phenotyping method used were the same as the previous year. All the phenotypic assays were performed in the Laboratory of Quarantine Organisms, Department of Plant Pathology, IHAR-PIB, Poland.

DNA extraction, pooling and sequencing

Genomic DNA of the bulked progeny clones, the two parents, the mother (Bzura), the suspected great-great-grand-father (BRA9089) of Kuba and of four susceptible varieties (Alegria, Desiree, Kuras and VR808) was extracted from freshly harvested leaves according to (26). DNA concentration was assessed using a Qubit Fluorometer (Invitrogen). For the S-bulk and the R-bulk, 59 ng of DNA of each individual was pooled. For all WGS experiments, 1 µg of (pooled) genomic DNA was used for library preparation and sequenced on an Illumina HiSeq2000 platform producing 151 bp paired end (PE) reads (Hartwig Medical foundation, Amsterdam, The Netherlands).

Genomic DNA of the entire full-sib population ($n = 328$) and of a panel of 56 resistant and susceptible potato varieties was extracted from freshly harvested leaves using a modified CTAB-method (27). These varieties were selected to embody the different wart resistance sources present in the European breeding germplasm. The DNA concentration was estimated using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific) and adjusted to a concentration of 5-50 ng/ μ l. The DNA quantity was confirmed and the quality was assessed on ethidium bromide containing agarose gels.

Comparative Subsequence Sets Analysis workflows

The CoSSA workflow used in this study is illustrated in Figure 1. All the scripts, the software and the code used in CoSSA are available on GitHub (<https://github.com/cprodhom/CoSSA-workflows>). The computational requirements of each of the main steps of the workflows are given in Supplementary file 2. The forward and reverse reads were quality trimmed and Illumina adapters were removed using Trimmomatic version 0.32 (28) (settings ILLUMINACLIP:TruSeq3-PE.fa:2:30:10, LEADING:3, TRAILING:3, SLIDINGWINDOW:4:15 and MINLEN:70). *K*-mer tables were built for each sequenced sample by using the GlistMaker program of the GenomeTester4 toolkit (29). GlistMaker generates *k*-mers lists from the trimmed reads and counts the frequency ($\approx$ sequencing depth) of each unique *k*-mer. We used a *k*-mer size of 31 nucleotides. This *k*-mers size was chosen as a compromise between sequence uniqueness (increase the *k*-mer size increases the chances that each *k*-mer comes from a unique region of the genome) and sequence correctness (decrease the *k*-mer size allows to reduce the PE reads errors). The *k*-mers that were observed only once were removed from the dataset as they are likely caused by sequencing errors and would unnecessarily increase the required computational power and memory if retained. The GlistCompare program of GenomeTester4 was then used to perform basic set operations such as unions, intersections or differences.

To obtain *k*-mers exclusively present in the R-bulk (R-bulk specific *k*-mers) we subtracted the S-bulk *k*-mers from the R-bulk. Additional filtering steps were applied to refine the R-bulk specific *k*-mers selection. The sequencing yield was 52 Gb for the R bulk. Therefore, the sequencing depth of the R-bulk was approximately 62x considering a haploid genome of 840Mb. Under the assumption of a uniform sequence coverage, the expected sequencing depth of the R haplotype is $\sim 16x$, as potato has a tetraploid genome ($2n = 4x$). Paralogous and other repetitive sequences are expected to be present at a multitude of this depth.

Figure 1. Comparative Subsequence Sets Analysis (CoSSA) workflow example

1. The R-bulk, the S-bulk and (optionally) the parents of the full-sib population are sequenced with Illumina 151 bp PE. The Illumina reads are quality trimmed and *k*-mer tables are built for each of the samples. The *k*-mer tables contain all the possible *k*-mers of the sample and their frequency (~sequencing depth). *K*-mers with a frequency of 1 are considered as sequencing errors and are removed from the set. The difference between the R-bulk and S-bulk *k*-mer tables is performed to keep only the R-bulk specific *k*-mers. 2. R-bulk specific *k*-mers are filtered upon their depth in function of the expected depth of the R haplotype. In our situation, we worked with a tetraploid population in which the resistance segregated with a 1:1 segregation ratio. Each of the progeny clones pooled in the R-bulk held the R-haplotype in one copy (simplex) which means that 25% of the reads cover the R-haplotype. The depth of each haploid genome in our R-bulk was ~16x so we kept the *k*-

mers with a depth from 10x to 22x. If the parents have been sequenced, additional set operations can be used to divide the k -mers in function of their parental origin: k -mers inherited from the resistant parent (Kuba, subset A: resistance specific k -mers), k -mers inherited from the susceptible parent (Ludmilla, subset B), k -mers inherited from both parents (subset C) and the k -mers inherited from none of the parents (subset D). 3a. If a reference genome is available, the k -mers are mapped to it to identify the resistance locus. 3b. If no reference genome is available, the read pairs containing the R-bulk specific k -mers are extracted from the raw reads data and used to re-assemble the resistant haplotype. Nu: Number of unique k -mers, Nt: Total number of k -mers.

In order to select for single copy SNPs specific for the resistant haplotype, we decided to keep k -mers with a depth from 10x to 22x. These filtered R-bulk specific k -mers were further divided into four different sets after comparison with the k -mers from the resistant and susceptible parent: inherited from the resistant parent (referred hereafter as “resistance specific k -mers”), inherited from the susceptible parent, inherited from both parents or deriving from none of the parents. This last set, probably representing contamination with unrelated biological material, was discarded for subsequent analyses.

In order to select SNPs with high diagnostic value, prior to the development of PCR markers, k -mer sets operations were used to eliminate k -mers present in different varieties not sharing the same trait. K -mers tables were produced for Alegria, Desiree, Kuras and VR808 which are all resistant to the pathotype 1 but susceptible to P2, P6 and P18. The union of the k -mers from these varieties was subtracted from the resistance specific k -mers set.

For reference genome based downstream analyses, the resistance specific k -mers minus S varieties, the R-bulk specific k -mers inherited from the susceptible parent and the R-bulk specific k -mers inherited from both parents were mapped to identify the genomic region harboring the trait of interest. We used BWA aln (version 0.7.12) (30) to map the R-bulk specific k -mers to version v4.03 of the *Solanum phureja* DM1-3 (DM) genome (31,32). After mapping, k -mers per 1Mb bins, starting at the beginning of each chromosome, were counted using the bedtools suite (v2.25) (33). The resulting numbers of k -mers per 1Mb bin were plotted using Microsoft Excel.

For reference genome independent downstream analysis, the paired end reads containing resistance specific k -mers were retrieved from the R-bulk and the resistant parent and assembled *de novo*. We used an in-house developed script to retrieve the PE reads containing at least one of the resistance specific k -mers. The *de novo* assembly of the extracted reads was done using SPADes (v3.11.1) (34) (command line options: -k 21, 33, 55, 77 --careful --pe1-1 --pe1-2 --pe2-1 --pe2-2). The

corresponding scaffolds are mainly derived from the haplotype bearing the resistance and are referred to as the R-haplotype scaffolds. To compare the CoSSA workflows with and without using the reference genome, we mapped the resistance specific *k*-mers to the R-haplotype scaffolds using BWA aln and allowing no mismatches (command line option: -n 0). The *k*-mers mapping to the three biggest R-haplotype scaffolds were extracted using SAMtools view (v1.3) (35) and mapped to the potato reference genome using BWA aln and allowing 0, 1 and 2 mismatches. The number of reads mapping to DM was determined using SAMtools flagstat.

K-mer sets algebra was also used for pedigree analyses. To check if different varieties sharing the same trait were identical by descent, the number of *k*-mers shared with the resistance specific *k*-mers were counted. For this purpose, the intersections between the resistance specific *k*-mers not present in the four susceptible varieties and the *k*-mers from the two pedigree clones (Bzura and BRA9089) were made. Successively, the resulting *k*-mers were mapped to the reference genome or to the R-haplotype scaffolds.

Minimum input required by CoSSA

To assess the influence of the sequencing depth on the CoSSA results, we randomly sampled subsets of reads from the R-bulk, the S-bulk, and the two parents using seqtk (v1.2) to simulate sequencing depths of 10x and 5x of the haploid genome. CoSSA was applied to these two datasets, changing only the R-bulk specific *k*-mers depth cut-off which was proportionally adapted to *k*-mers with a depth comprised between 6x and 14x and *k*-mers with a depth between 3x and 7x for the simulations with a depth of 10x and 5x respectively. To assess if we could identify the same loci with the lower depth datasets as with the full dataset, we calculated the signal to noise ratio (SNR) for the 10x and 5x simulations, with or without intersections from the parents. Without the parents' intersections, the SNR was calculated as being the average number of R-bulk specific *k*-mers mapping under the peak (the peak boundaries are defined as identified with the full dataset) divided by the number of R-bulk specific *k*-mers mapping to the rest of the chromosome. If the parent's intersections were included, the SNR was calculated as being the ratio between the average of R-bulk specific *k*-mers coming from the specific donor parent under the peak and the average of R-bulk specific *k*-mers coming from the donor parent mapping to the rest of the chromosome.

To test if sequence input for the CoSSA workflows could be minimised further, we subtracted the S-bulk *k*-mers from the resistant parent (Kuba). We kept the *k*-mers

from the difference with a depth comprised between 8x and 18x and mapped them to the reference genome. The number of mapped k -mers per 1Mb was counted and compared with the CoSSA results obtained with the two bulks and the two parents.

Design of PCR markers for validation and diagnostic purposes

A haplotype specific SNP without flanking SNPs within 30 bp on each side has a maximum of k ($=31$) resistance specific k -mers mapped to it. To select these “isolated haplotype specific SNPs” for the design of markers, we mapped the resistance specific k -mers minus the S varieties k -mers to the potato reference genome DM (v4.03). To select SNPs coming from Ludmilla, we mapped the R-bulk specific k -mers inherited from Ludmilla to the reference genome. From the resulting .bam files, the position of the mapped k -mers was inferred using SAMtools and exported to Microsoft Excel. This way, we could pinpoint isolated haplotype specific SNPs because 31 k -mers were mapped in a row and because there were no other k -mers mapped to the flanking 61 bp sequences. The candidate SNPs were visualized in the .bam file using the Integrative Genomics Viewer (IGV, v2.3.72) (Supplementary figure 1).

To identify resistant haplotype specific SNPs without using the reference genome, we mapped the resistance specific k -mers minus the S varieties k -mers to the *de novo* assembled R-haplotype scaffolds and isolated SNPs were pinpointed as described previously. As we mapped the resistance specific k -mers without the S varieties k -mers to scaffolds belonging to the resistant haplotype, the susceptible variant(s) of the SNPs were not known. Knowing the susceptible variant(s) of the identified SNPs is necessary to design KASP markers. Therefore, we also mapped the R-bulk reads, which contain reads from the resistant haplotype and the seven possible susceptible haplotypes, to the *de novo* assembly using BWA mem (version 0.7.17). Duplicates were removed using sambamba (v0.5.1) (36). The susceptible variant(s) of each selected SNP could be visualized in IGV.

KASP markers genotyping

A first set of Kompetitive Allele Specific Polymorphisms (KASP) markers (Supplementary file 3), that were used to validate the different CoSSA peaks, was designed by LGC Genomics (LGC, Hoddeston, UK) and tested in a subset ($n = 83$) of the KxL population. The clones from this subset were selected for historical reasons (first clones of the KxL population). The KASP markers name reflects the physical coordinate of the SNP, relative to DM. Allele specific forward primers and common reverse primers were designed based on the 61 bp sequences of the 31 k -mers from

CoSSA mapped to the reference genome. One KASP marker (chr11_1259552) flanking the resistance locus and two new KASP markers designed for further fine-mapping (chr11_1519485 and chr11_1666090; Supplementary file 3) were used to genotype the full population ($n = 328$). Two haplotype specific markers (chr11_1259552, chr11_1772869; Supplementary file 3) flanking the resistance locus were used to genotype an independent panel of 56 susceptible and resistant potato varieties. The KASP genotyping assays were performed by C. Meijer BV (Rilland, The Netherlands) according to the manufacturer's instructions (LGC Genomics). In brief, 1.5 μL DNA of a concentration of 5-50 ng/ μL in a 1536-well plate was dried in a fan oven for one hour at 55°C and 1 μL of the PCR mix was added (containing the 2x KASP Master mix, the primer mix and the water). The PCR was performed in a Hydrocycler 16 and the data reading was performed on a BMG PHERAstar®

A second set of three KASP markers (NODE1_7193, NODE2_18410, NODE3_17594; Supplementary file 3) that were used to validate the "CoSSA without reference genome" workflow was designed in house using Primer3 (37). The three markers were designed for SNPs from the three longest R-haplotype scaffolds and tested in the subset of 83 KxL progeny clones. These KASP assays were performed according to the manufacturer's guidelines (LGC Genomics, Hoddesdon, UK) in a total volume of 10 μL with 5 to 50 ng of genomic DNA on a BioRad CFX96™ Real-Time System machine.

Chi² tests were used to test the goodness of fit of the segregation of the tested markers with the expected 1:1 segregation ratio. Kruskal-Wallis tests were performed using R v3.2.3 to validate the association of the tested markers with wart disease resistance.

Results

Distribution of resistance in the KxL population

The resistance of the entire full-sib population ($n = 328$) and the two parents to pathotypes 2, 6 and 18 was assessed (Supplementary file 1). The distribution of resistance to the pathotypes 2 and 6 segregated in a bimodal fashion in a 1:1 ratio (χ^2 test p -value > 0.05 ; Supplementary file 4A, B, E). The pathotype 18 scores were skewed towards susceptibility (χ^2 test p -value < 0.001 ; Supplementary file 4C) which can be explained by a weaker resistance of Kuba to pathotype 18. Nevertheless, the Pearson correlations between the resistance scores for the three pathotypes were very high (Supplementary file 4D): 0.93 between P2 and P6, 0.91 between P6 and P18 and 0.89 between P2 and P18. These observations suggested the presence of a single major resistance locus for all three pathotypes and possibly a minor effect locus improving

resistance to pathotype 18. When transforming the quantitative scores into a qualitative trait (Resistant or Susceptible, Supplementary file 1), we observed a perfect co-segregation between the three pathotypes. Only a few exceptions were found where the phenotypic classification for resistance against P2, P6 and P18 did not co-segregate. Most of these exceptions might be explained by the choice of the threshold to call a clone resistant or susceptible.

Design of a CoSSA workflows for gene mapping

To identify sequence variants linked to the wart disease resistance from the KxL population, we pursued a Bulk Segregant Analysis combined with Whole Genome Sequencing. We selected 17 fully resistant and 17 fully susceptible plants to the three pathotypes to compose a resistant bulk (R-Bulk) and a susceptible bulk (S-Bulk), respectively. Selected clones are highlighted in Supplementary file 1. The R-bulk, the S-bulk, Kuba (R parent) and Ludmilla (S parent) were sequenced with Illumina 151 bp PE reads. The calculated sequencing depth obtained for each sample was of 62x, 82x, 72x and 70x, respectively, assuming an 840 Mb genome (Supplementary file 5). It was decided not to use complete reads but subsequences (*k*-mers) of 31 nucleotides instead. Indeed, sequences with an invariable length can be unambiguously compared using set algebra. These basic set operations are less computationally demanding than methods requiring reads alignment to a reference and allow to work independently from a reference. *K*-mer tables ($k = 31$), containing the unique *k*-mers and their frequency, for the four samples were produced and the *k*-mers with a depth of one were removed (Supplementary file 5). For Kuba and Ludmilla, we observed peaks of *k*-mers at four different *k*-mer frequencies. They corresponded to sequencing errors and PCR duplicates (low depth), the *k*-mers present in one haplotype (simplex; 13 for both Kuba and Ludmilla), two haplotypes (duplex), and the *k*-mers present in three haplotypes (triplex) or four haplotypes (quadruplex) (Supplementary figure 2). The *k*-mers with a depth >52 ($4 \times 13x$) were assumed to be originating from multi-copy sequences. Except for the first two peaks (errors and single copy), no distinct frequency peaks could be observed in the bulks. Most likely, the remaining peaks are diffuse because the bulk samples were composed of 17 different DNA samples from individuals containing genetic variants from eight haplotypes. From these eight haplotypes, only the resistant haplotype was present in each R-bulk individual which, theoretically, would be twice the frequency of the single copy peak ($2 \times 7 = 14$ for the R haplotype).

In order to select *k*-mers linked to resistance, we subtracted the S-bulk *k*-mers (Nu (number of unique *k*-mers) = 1,669,156,268; Nt (total number of *k*-mers) =

45,467,875,801) from the R-bulk *k*-mers set ($N_u = 1,601,951,563$; $N_t = 34,475,435,516$) to obtain the R-bulk specific *k*-mers ($N_u = 88,774,024$; $N_t = 260,595,070$) (Supplementary file 5). After enriching for simplex *k*-mers using a *k*-mer depth threshold between $10 < N_t \text{ } k\text{-mers} < 23$, we retained 1,979,520 unique *k*-mers ($N_t = 23,219,641$). These unique *k*-mers were then divided into four different subgroups: 624,608 unique *k*-mers inherited from Kuba (resistance specific *k*-mers), 197,014 unique *k*-mers inherited from Ludmilla, 329,529 unique *k*-mers inherited from both parents and 828,369 unique *k*-mers inherited from none of the parents. From this last set of *k*-mers, 69% could be mapped to the potato reference genome. This set was not used for further analysis because they were likely due to sequencing errors and DNA contamination from an unrelated potato genotype.

To further increase the haplotype specificity of the resistance specific *k*-mers, we removed from the resistance specific *k*-mers the *k*-mers that were also found in the susceptible varieties Alegria, Desiree, Kuras and VR808 (S varieties). The sequencing depth obtained respectively for each tetraploid variety was 46x, 46x, 43x and 45x (Supplementary file 5). From 624,608 resistance specific *k*-mers, 319,443 *k*-mers remained after subtracting the S varieties *k*-mers (Supplementary files 6 and 8). The unique resistance specific *k*-mers minus the S varieties and the unique R-bulk specific *k*-mers inherited from Ludmilla and from both parents (845,986 unique *k*-mers in total) were mapped to 1Mb bins of the potato reference genome DM (Figure 2, Supplementary file 6). The highest *k*-mers peak was identified on the first 5 Mb of chromosome 11, where 13% of all the R-bulk specific *k*-mers sets mapped. From these *k*-mers, 84.3% were inherited from Kuba, the resistant parent (resistance specific *k*-mers minus S varieties). Interestingly, we observed another peak of resistance specific *k*-mers mapping to the 20-21Mb bin of the chromosome 0, and specifically to the unanchored superscaffold PGSC0003DMB000000734 (referred hereafter as DMB734). Several smaller peaks of *k*-mers could be observed in other genomic regions (Figure 2). The larger and smaller *k*-mers peaks might represent major and minor effect loci, a hypothesis that required validation. We decided to design markers to validate the two major peaks and five minor peaks of R-bulk specific *k*-mers coming from Ludmilla: on chromosome 3 (14 - 17Mb; 0.39% of the mapped three R-bulk specific *k*-mers sets), chromosome 4 (64 - 68Mb; 0.61%), chromosome 5 (0 - 3Mb; 1.46%; 7 - 10 Mb; 2.1%), and chromosome 9 (55 - 57Mb; 0.62%) - (Supplementary file 6).

Validation of the CoSSA results

To validate the hypothesised resistance loci positions as identified with CoSSA, we set out to design haplotype specific KASP markers for the two larger and five of the smaller *k*-mers peaks identified. To increase the chances to select SNPs with diagnostic value, we removed from the resistance specific *k*-mers the *k*-mers that were also found in the susceptible varieties Alegria, Desiree, Kuras and VR808. When comparing the peak of resistance specific *k*-mers mapping to chromosome 11 before and after the subtraction of the S varieties *k*-mers (Supplementary file 6F), we observed that the highest peak shifted from the 3-4Mb bin to the 1-2Mb bin. This suggested that a high proportion of frequently occurring SNPs from multiple ancestral origin mapped to the 3-4Mb bin (Supplementary file 6C).

From these *k*-mer sets, “isolated SNPs” were selected to design 24 KASP markers, covering all 7 CoSSA peaks, which were used to genotype a subset of 83 progeny clones from the KxL population. Chi² tests were used to validate the expected 1:1 segregation pattern of the markers (Table 1). All the markers followed the expected ratio except chr11_5505183 for which the segregation was skewed toward the alternative allele. (Table 1; Supplementary file 1). The KASP markers from chromosome 11 and DMB734 largely co-segregated and were strongly (*p*-values < 0.001) associated with pathotypes 2, 6 and 18 resistance as determined using Kruskal Wallis tests. Upon construction of a linkage map, it was found that DMB734 markers were positioned between markers corresponding to 1,259,552 bp and 1,772,869 bp of chromosome 11. So most likely, DMB734 must be positioned inside one of the two gaps between DMB148 and DMB505. Despite this misassembly, the order of the chromosome 11 markers and the order of the markers on DMB734 were conserved within the *Sen3* locus in comparison with the reference genome.

Markers designed in four of the smaller CoSSA peaks were not associated with resistance (*p*-values > 0.05). Only the markers chr09_55113777 and chr09_55354021 were associated with pathotypes 2, 6 and 18 resistance (*p*-values < 0.01; Table 1). This second resistance locus was inherited from the susceptible parent Ludmilla and shows a smaller effect on resistance than the chromosome 11 locus (Supplementary file 7). This shows that SNPs identified using CoSSA can identify major and minor loci, but also false positives. Hence, the SNPs identified in the peaks require validation through marker analysis.

Figure 2. Comparative Subsequence Sets Analysis in the KxL population

The R-bulk specific k -mers from the four different sets are mapped to the 12 chromosomes of DM and to the unanchored scaffolds (chromosome 0, v4.03). For each chromosome, the average of the number of R-bulk specific k -mers mapped has been calculated and for each 1Mb bin, the number of k -mers mapped from each set has been transformed to a percentage of the chromosome average. (a) Number of NLR genes per bin (1Mb) from the DM genome according to (7). (b) R-bulk specific k -mers minus k -mers present in the susceptible varieties ($y_{\max} = 1,208\%$) inherited from Kuba, (c) the k -mers inherited from Ludmilla ($y_{\max} = 469\%$), (d) the k -mers inherited from both parents ($y_{\max} = 516\%$), (e) intersection of the k -mers from (b) with the k -mers from BRA9089 ($y_{\max} = 428\%$), (f) intersection of the k -mers from (b) with the k -mers from Bzura ($y_{\max} = 1,205\%$). The arrows indicate the validated resistance loci and the position of the DMB734 scaffold.

Table 1. Validation of the CoSSA SNPs

SNP	R allele donor	Association with P2 resistance*	Association with P6 resistance*	Association with P18 resistance*	χ^2	χ^2 test significance
chr00_20792298	Kuba	2.1E-14	1.2E-14	8.3E-15	1.5	ns
chr00_20801212	Kuba	1.3E-14	7.3E-15	4.8E-15	1.2	ns
chr00_20858042	Kuba	2.1E-14	1.2E-14	8.3E-15	1.5	ns
chr00_20872536	Kuba	1.3E-14	7.3E-15	4.8E-15	1.2	ns
chr03_15683229	Ludmilla	6.8E-01	6.9E-01	3.0E-01	0.1	ns
chr04_64488861	Ludmilla	6.6E-01	9.7E-01	3.8E-01	0.1	ns
chr04_65204393	Ludmilla	4.5E-01	1.1E-01	2.0E-01	1.5	ns
chr05_2117034	Ludmilla	3.8E-01	9.4E-02	1.4E-01	0.0	ns
chr05_8795719	Ludmilla	2.0E-01	9.0E-02	1.6E-01	0.0	ns
chr05_9254557	Ludmilla	2.3E-01	8.4E-02	1.8E-01	0.0	ns
chr09_55113777	Ludmilla	2.4E-03	9.0E-04	4.2E-03	0.0	ns
chr09_55354021	Ludmilla	4.6E-03	1.3E-03	5.9E-03	0.1	ns
chr11_799292	Kuba	7.3E-13	6.7E-14	1.5E-13	0.4	ns
chr11_1259552	Kuba	2.0E-12	1.8E-13	4.5E-13	0.8	ns
chr11_1680269	Kuba	2.2E-14	1.2E-14	7.7E-15	1.5	ns
chr11_1772869	Kuba	2.9E-13	1.8E-13	1.4E-13	0.6	ns
chr11_1919971	Kuba	2.2E-13	1.4E-13	1.1E-13	0.8	ns
chr11_2040481	Kuba	2.2E-13	1.4E-13	1.1E-13	0.8	ns
chr11_2298154	Kuba	2.1E-12	9.7E-12	1.1E-12	1.8	ns
chr11_2687396	Kuba	3.6E-11	6.0E-11	1.9E-11	2.1	ns
chr11_2815640	Kuba	2.5E-11	5.5E-11	1.4E-11	1.5	ns
chr11_3325292	Kuba	1.2E-09	4.6E-09	7.7E-10	2.1	ns
chr11_4362386	Kuba	1.0E-09	3.0E-08	3.4E-09	1.0	ns
chr11_5505183	Kuba	4.8E-06	6.4E-04	4.3E-04	6.5	< 0.01

χ^2 tests: null hypothesis: the marker follows a 1:1 segregation pattern, ns: not significant.

* p -values of association between KASP marker results and Glynne-Lemmerzahl tests of 2016 (Kruskal-Wallis tests)

CoSSA data as a source for sequence variants in fine-mapping

In the subset of 83 progeny clones, we could identify 20 recombinants between markers chr11_799292 and chr11_5505183 on the chromosome 11 haplotype (Figure 3B; Supplementary file 1). Three informative recombinants indicated that the locus giving resistance to all the tested pathotypes, which we will refer to as *Sen3*, is within a 513,317 bp window relative to the reference genome (excluding DMB734; flanked by chr11_1259552 and chr11_1772869). We used the CoSSA resistance specific k -mers without the S varieties k -mers to design two more KASP markers inside the resistance interval (chr11_1519485 and chr11_1666090). Screening of the full KxL

Figure 3. Genetic and physical maps of the resistant haplotype

(A) Physical map of the three longest scaffolds of the R-haplotype *de novo* assembly. (B) Genetic map of the resistance region in a subset ($n = 83$) of the Kuba x Ludmilla population. The chromosome 0 scaffold could be anchored to the resistance interval. The analysis of recombinants allowed to fine-map the region flanked by chr11_1259552 and chr11_1772869. The markers developed without using the potato reference genome are underlined. (C) Genetic map of the resistance region in the full Kuba x Ludmilla population ($n = 328$). The resistance locus was fine-mapped to an interval between chr11_1259552 and chr11_1519485. (D) Physical map of the chromosome 11 first 6Mb region. The KASP markers developed to screen the Kuba x Ludmilla population are in black on the right of the haplotype and on the left are the NLR clusters (7). The markers in grey are the markers used to map *Sen1* (NI25 and NI27) (16,18) and to map Karolin's resistance (solcap markers and SSCP13) (22).

population ($n = 328$) with chr11_1259552, chr11_1519485 and chr11_1666090 narrowed *Sen3* to a 260kb region between PGSC coordinates 1,259,552 bp and 1,519,485 bp (Figure 3C). Interestingly, this genomic position contains many TNL (TIR Nucleotide-binding site Leucine-rich repeat) sequences (Figure 3D) and matched both the C76 (7) cluster which is syntenous to the XIa-TNL cluster (38) and RH11.1a clusters (39).

Only two false positives (KxL9006 and KxL9201) and one false negative (KxL9128) were observed in the full population ($n = 328$). Because of the scarcity of these observations, we can reasonably assume that they were due to mismatches between the samples used for the phenotyping and genotyping. None of the clones which showed a non-perfect co-segregation of the qualitative resistance traits to P2, P6 and P18 showed a recombination event between chr11_1259552 and chr11_1519485 (Supplementary file 1).

CoSSA based gene mapping without a reference genome

Identification of sequence variants by mapping reads to a reference genome has its limitations. The genes or regions of interest may be absent or overly different from the reference. Especially *R* gene clusters are difficult to assemble (as evident from the unanchored DMB734 contig; this study), and divergent, even among closely related haplotypes. We were interested to test if CoSSA could also be used without a reference genome. We extracted 408,281 and 469,251 read pairs from the R-bulk and from Kuba, respectively, which contained at least one of the resistance specific *k*-mers and assembled them. The obtained *de novo* assembly was composed of 21,306 scaffolds ($N_{50} = 1,115$) ranging in size from 100 until 46,690 bp (Supplementary table 1). To locate the R haplotype specific variants, we mapped the resistance specific *k*-mers minus the S varieties *k*-mers to the *de novo* scaffolds. "Isolated SNPs" were selected

to avoid KASP assay failure due to (unknown) flanking variants. To know the susceptible allele(s) of each DNA sequence variant, we also mapped the R-bulk reads to the *de novo* scaffolds. Next, we designed one KASP marker on each of the three longest scaffolds (node1 = 46,690 kb, node2 = 44,756 kb, and node3 = 38,969 kb; Figure 3A) and tested them in the subset of 83 offspring. The three markers were indeed linked to resistance and could be placed in the linkage map (Figure 3B). The three markers were specific to the resistant haplotype and the markers present on node3 and node1 were flanking the resistance gene, while the node2 marker fully co-segregated.

To further inquire the benefit of the CoSSA workflow without a reference genome, we mapped the resistance specific *k*-mers to the three longest R-haplotype scaffolds (node1, node2 and node3) allowing no mismatches. The mapped *k*-mers were extracted and mapped to the potato reference genome allowing 0, 1 and 2 mismatches (Supplementary table 2). When no mismatches were allowed, only 2.95%, 0.58% and 41.85% of the *k*-mers mapping to the nodes 1, 2 and 3 respectively could map to the reference genome. When 1 mismatch was allowed (the mismatch corresponds to the resistance specific SNP present in each *k*-mer), 25.6%, 33.11% and 66.63% of the *k*-mers mapping to nodes 1, 2 and 3 respectively mapped to the reference genome. We released even more the mapping stringency by allowing 2 mismatches during the mapping, which allowed *k*-mers containing 2 SNPs to be mapped. With these settings, 45.17%, 60.78% and 81.85% of the *k*-mers mapping to nodes 1, 2 and 3 mapped to the reference. Overall, the nodes 1 and 2 R-haplotype scaffolds contained only a small portion of *k*-mers that mapped to the reference genome, which can be explained by poor assembly of the reference genome or by huge differences between the DM haplotype compared to the *Sen3* haplotype. The node3 scaffold contained more *k*-mers, which could map to DM. According to our genetic map (Figure 3B), the DM assembly deviates strongly from the *Sen3* haplotype in the region between the flanking markers, which corresponds to the *R* genes clusters. These results clearly show the advantage of working with a reference genome independent approach.

CoSSA-based pedigree analysis

To assess which clone in the pedigree of Kuba has been the donor of resistance (Supplementary figure 3), we sequenced BRA9089 (suggested great, great grandparent of Kuba) and Bzura (grandparent of Kuba) and obtained sequencing depths of 51x and 47x respectively. We used the resistance specific *k*-mers without the S varieties *k*-mers and made the intersection with Bzura and BRA9089 *k*-mers sets.

Bzura and BRA9089 shared respectively 75.7% and 6.8% of the resistance specific *k*-mers without the S varieties *k*-mers. These shared *k*-mers were mapped to the reference genome (Figure 2). Along the seven first bins of chromosome 11 (Supplementary figure 4A), Bzura shared 99.44% of the *k*-mers with Kuba whereas BRA9089 shared only 12.9% of the *k*-mers with Kuba. These results showed that Bzura shared the full resistant haplotype with Kuba but it was not the case for BRA9089. In order to rule out that crossover events on both sides of the causal gene had resulted in lack of similarity in the *k*-mers lists of Kuba and BRA9089, we focussed on the fine-mapped interval (1,259,552 bp - 1,519,485 bp). BRA9089 shared 1.7% of the Kuba resistance specific *k*-mers whereas Bzura hold 99.4% of them. In the *de novo* assembled R-haplotype scaffolds we described previously, on scaffolds node2 and node1 containing the markers flanking *Sen3*, Bzura shared 99.9% of the R-bulk specific *k*-mers with Kuba and BRA9089 0.02% only (Supplementary figure 4B). These observations strongly suggested that the BRA9089 clone we sequenced was not the donor of resistance ^a.

Robustness of CoSSA based gene mapping

In the CoSSA workflows, we applied a depth cut-off on the R-bulk specific *k*-mers (10 to 22x). We wanted to assess how these upper and lower depth cut-offs affected the CoSSA peaks identification. If the parents were sequenced and no lower cut-off was applied on the R-bulk specific *k*-mers (2 – 22x; Supplementary figure 5A), the signal-to-noise (SNR) ratio of *Sen3* (chromosome 11: 0-5Mb) was drastically reduced from 45.15 to 6.04 (Supplementary figure 5D). The signal from the peak was still 6 times higher than the noise from the rest of the chromosome so the peak was still identifiable. In the case of the minor effect locus from Ludmilla on chromosome 9 (55-57Mb), that we will refer as *Ludmilla_chr9*, however, the SNR was reduced from 10.04 to 1.75 which was too low to identify the peak. Therefore, the lower depth cut-off was not necessary to identify the major effect locus but necessary to identify the minor effect one. When no upper cut-off was applied (Supplementary figure 5C), the SNR of both the major effect and the minor effect loci were unaffected: the SNR of *Sen3* was of 45.56 instead of 45.15 when an upper cut-off was applied and the SNR of *Ludmilla_chr9* was of 10.11 instead of 10.04 (Supplementary figure 5D). Therefore, the upper cut-off threshold does not affect the detection of the peaks but impacts the selection against multi copy SNPs.

To determine the minimum input to supply for CoSSA to be efficient, we analysed the CoSSA results when the parents' sequences were not included in the workflow. When

sequencing only the two bulks, the SNR of the *Sen3* peak was halved (45.15 to 22.19; Supplementary figure 5D) but it was still sufficient to clearly identify the peak. For the minor effect locus, however, the SNR was reduced from 10.04 to 1.88 when the parents were not sequenced which was too low to identify the peak. Therefore, sequencing only the two bulks was sufficient to identify the major effect locus but not to identify the minor effect one. When the parents were not sequenced, the *k*-mers lower depth cut-off became vital, as the SNR of *Sen3* was reduced from 22.19 (10 to 22x) to 1.38 (2 to 22x) which made the peak not identifiable. Moreover, new peaks arose in other positions of the genome (Supplementary figure 5A). These peaks were composed of *k*-mers that were not inherited from either of the parents, likely due to DNA contamination from unrelated biological material as discussed previously. This contamination would not have been removed if the parents had not been sequenced, making the lower depth cut-off the only way to discard it.

To further minimize the sequence data input for CoSSA, we reduced the sequencing depth of the samples (two bulks and parents) to a depth per haploid genome of 10x and 5x instead of 16x with the full dataset used previously (Supplementary table 3 ; Figure 4). The SNR of *Sen3* was reduced from 22.2 (16x) to 3.02 (10x) and 1.3 (5x) if the parental origin of the *k*-mers was not included in the CoSSA workflow, (Figure 4A and D). Therefore, a minimum sequencing depth of 10x per haploid genome was sufficient to identify *Sen3* when no parents were sequenced. However, new peaks arose on chromosomes 2, 4 and 5, which were false positives, as they were not identified with a depth of 16x per haploid genome (Figure 4A). For the minor effect locus on chromosome 9, 16x per haploid genome was already not sufficient for its identification when no parents were sequenced. When the parental origin of the *k*-mers was included in the workflow, *Sen3* SNR was reduced from 45.3 (16x) to 5.5 (10x) and 2 (5x) (Figure 4B and E). Therefore, even a sequencing depth of 5x per haploid genome was sufficient to identify the major locus, *Sen3*, if the parents were sequenced. However, as observed previously, new peaks arose on other chromosomes but the chromosome 11 peak was still the highest (Figure 4B). Regarding the minor QTL (*Ludmilla_chr9*), the SNR was reduced from 10 (16x) to 4.2 (10x) and 2.3 (5x) for 10x and 5x (Figure 4C and F). However, many new peaks arose for the 10x and 5x datasets (Figure 4C). The identification of this locus would then require the design of many more markers for each of these peaks to discard false positives.

Finally, to reduce the CoSSA input by sequencing fewer samples, we simulated the situation where only the resistant parent (Kuba) and the susceptible bulk were

Figure 4. Assessment of the CoSSA minimum required sequencing depth

(A) Total R-bulk specific *k*-mers, (B) R-bulk specific *k*-mers inherited from Kuba (resistance specific *k*-mers) and (C) R-bulk specific *k*-mers inherited from Ludmilla mapped to the potato reference genome using the full sequencing depth ($\sim 16\times$ / haploid) (red), a depth of $10\times$ / haploid (green) and a depth of $5\times$ / haploid (yellow). (D) Signal to noise ratio (SNR) of the different peaks in the case the parents are not sequenced, (E) and (F) SNR of the different validated peaks in the case parents are sequenced.

sequenced. The difference between Kuba's *k*-mers sets and the S-bulk's *k*-mers was made. The *k*-mers with a depth comprised between 8 and $18\times$, according to the peak ($n=13$) of simplex *k*-mers in Kuba (Supplementary figure 2), were mapped to the reference genome (Supplementary figure 6). With this reduced input, CoSSA was successful to identify *Sen3*, which had an SNR of 7.39. Obviously, the minor QTL *Ludmilla_chr9* was not found because it is absent from Kuba. However, one smaller peak arose on chromosome 2 (0-13Mb) with an SNR of 4.28. This peak might be an additional minor effect locus. Again, a validation would be required to rule-out the fact that it is a false positive peak.

Diagnostic value of the haplotype specific markers for breeding

To assess the diagnostic value of the resistant haplotype specific KASP markers to detect *Sen3* in breeding clones and varieties, we genotyped an independent panel of 56 tetraploid resistant and susceptible varieties with the flanking markers chr11_1259552 and chr11_1772869 (Table 2). The resistant allele of the markers was absent in the tested varieties susceptible to the pathotypes 2, 6 and 18, including the BRA9089 clone which we showed to lack the *Sen3* haplotype using CoSSA. In 17 of the tested varieties, both flanking markers scored positive. From these varieties, 14 (including Bzura that contained *Sen3* as determined using CoSSA) are known to be fully resistant to pathotypes 1, 2, 6 and 18. For the other three varieties (Gawin, Rudawa and Sonda), only a partial pathotype resistance pattern was known. We observed several false negative varieties in the dataset, which were resistant to the pathotypes 2, 6 and/or 18 but did not score positive to the tested markers. These varieties likely got their resistance from another source than *Sen3*.

Table 2. *Sen3* markers validation in a panel of 56 varieties

Variety	Registration		Flanking makers		Resistant to				Clone in pedigree		
	Country	Year	FM1	FM2	P1	P2	P6	P18	Ora	BRA9089	Capella
Actaro	HOL	2013	0	0	R	R	R	R			
Adam	POL	2005	1	1	R	R	R	R	?	?	?
Alegria	BRD	2003	0	0	S	S	S	S			
AM78-3704	HOL		0	0	R	R	R	R			
Andante	BRD	2003	0	0	R	R	R	R			
Antares	DDR	1961	0	0	R	R		R			
Apollo	DDR	1956	0	0	R	R		R			
Aventra	HOL	2005	0	0	S	R	R	R			
Axion	HOL	2008	0	0	R	R	R	R			
Belita	HOL	1986	0	0	R	R	R	R			
Bintje	HOL	1910	0	0	S	S	S	S			
BRA9089	GER		0	0	S	S	S	S			
Bzura	POL	1983	1	1	R	R	R	R	y	y	y
Capella	GER	1943	0	0							
Carrera	HOL	1999	0	0	R	S	R	S			
Cekin	POL	2004	1	1	R	R	R	R	?	?	?
Celandine	HOL	2013	0	0	S	S	R				
Constantina	BRD		1	1	R	R	R	R	?	?	?
Delcora	HOL	1988	0	0	R	R	R	S			
Deodara	GER	1913	0	0	S	S	S	S			
Desiree	HOL	1962	0	0	R	S	S	S			
Eurotonda	BRD		1	1	R	R	R	R	?	?	?
Euroviva	BRD		0	0	R	R	R	R			
Gandawa	POL	2004	1	1	R	R	R	R	?	?	?
Gawin	POL	2010	1	1	R			R	?	?	?
Giewont	POL	1955	0	0	R	S	S				
Ibis	POL	1987	1	1	R	R	R	R	n	y	y
Ikar	POL	1996	2	2	R	R	R	R	n	y	y
Karlana	DDR	1988	0	0	R	S	S	S			
Karolin	BRD	1990	1	1	R	R	R	R	y	y	y
Kuba	POL	1999	1	1	R	R	R	R	y	y	y
Kuras	HOL	1996	0	0	R	S	S	S			
Lorch	SU	1931	0	0	S						
Ludmilla	BRD	2008	0	0	R	S	S	S			
Miriam	BRD	1988	0	0	R	S	S	S			
MPI44-1016-10	GER		1	0	R	S	S	S	n	n	n
MPI44-1016-24	GER		0	0	R	R	R	R			
MonaLisa	HOL	1982	0	0	R						
Ora	DDR	1952	1	1	R	R	R	R	-	y	y
Otolia	BRD		0	0	R	S	R	R			
Parella	HOL		0	0	R	R	R	S			
Polessky Rosovy	SU	1978	0	0							
Producent	HOL	1984	0	0	R	S	S	S			
Rudawa	POL	2002	1	2	R	S	R	S	n	y	y
Saphir	BRD	1960	0	0	R	R	R	R			
Smart	HOL	2008	0	0	R	S	S	R			
Sonda	POL	2002	1	1	R	R	S	S	?	?	?
Spectra	HOL	2016	0	0	R		R				
Talent	BRD	2006	0	0	R	R	R	S			
Tivoli	DEN	2002	0	0	R	R	R	R			
Tomensa	BRD	1989	0	0	S	S	S	S			
Transit	BRD	2009	1	2	R	R	R	R	y	y	y

Ulme	BRD	1991	1	1	R	R	R	R	y	y	y
Ultra	HOL	1999	0	0	S		R				
VR808	HOL	2009	0	0	R	S	S	S			
Zagloba	POL	2007	1	1	R	R	R	R	y	y	y

Flanking markers of *Sen3*: FM1 = chr11_1259552 and FM2 = chr11_1772869. The dosage of the FM1 and FM2 is given (0: absent, 1: simplex, 2: duplex). Pedigree information is only shown if the variety was positive for the *Sen3* markers. ? : No pedigree information is available, n: This clone is not in the pedigree, y: This clone is in the pedigree.

Discussion

In this study, we designed CoSSA workflows for gene mapping and related applications in which sequence variants are deduced directly from subsequences (*k*-mers), without the necessity of variant calling software. CoSSA is a quick and robust method to identify a locus involved in the variation of a particular trait. When the biggest datasets in this study were used, the jobs parallelized and run on only one node of 10.5 Gb memory, the CoSSA workflow using a reference genome took half a day. The reference independent workflow required around 30 hours. The read pairs extraction step, currently consuming 23 hours CPU, could be improved to be quicker (Supplementary file 2). In this case study, we mapped a major and a minor locus for resistance against pathotypes 2, 6 and 18 of *Synchytrium endobioticum*, the causal agent of potato wart disease. The CoSSA workflows also enabled to identify the major locus independently from a reference genome. Furthermore, CoSSA enabled to select for haplotype specific SNPs to fine-map the *Sen3* locus corresponding to a 260 kb interval in the reference genome (between 1,259,552 bp and 1,519,485 bp). The haplotype specific SNPs identified in this way proved to have diagnostic power in potato breeding material and pedigree analysis.

S. endobioticum pathotypes 2, 6 and 18 resistance had previously been identified in a similar region as the *Sen3* locus as identified in this study (21,22). In the first study, the authors used 195 SSR markers to genotype the resistant parent Karolin and a susceptible parent which identity was not disclosed, as well as two DNA bulks of a tetraploid offspring segregating for pathotypes 1, 2, 6 and 18 resistance (21). Furthermore, the 8.3k SolCAP potato SNP array was used to identify markers significantly associated with resistance. The authors could not validate these markers with resistance in a broader panel of 91 distantly related varieties, probably because none of the markers was in coupling phase with the resistance gene. Such ascertainment bias can be circumvented only with haplotype specific SNPs. The failure to extrapolate SolCap markers associated with resistance from a biparental study to a GWAS panel should be expected because the SNP discovery panel of SolCap SNPs is not resistant to wart disease. In a later study, a dihaploid population was generated

from the resistant variety Karolin and screened with the 12.8k SolCAP potato genotyping array (22). The authors compensated the ascertainment bias caused by the SolCAP array by developing new haplotype specific markers using the RNAseq data obtained from three resistant descendants. The resistant locus was mapped to a region corresponding to a 777 kb interval of the reference genome (939,581 bp and 1,716,722 bp) which overlaps with the 260 kb interval we identified in this study.

In our study, we included Bzura, the resistant grand-parent of the studied population, and BRA9089, a speculated donor for a resistance locus found in the German and Polish potato breeding material (40–42). Among the 17 varieties from the tested panel holding the *Sen3* markers, eight were bred in Germany and nine in Poland. Unfortunately, we could only trace back the resistant haplotype in Kuba's pedigree until Ora, a daughter of BRA9089 and Capella. Neither Capella, nor BRA9089 contained *Sen3*. We re-phenotyped the BRA9089 clone that we sequenced and it turned out to be susceptible to the four pathotypes. The impossibility to trace back further than Ora the haplotype in the pedigree could be due to pedigree errors. Another likely explanation is that the Capella and/or BRA9089 clones referred in literature are different from the clones we call Capella and BRA9089 today. Three out of the varieties that harbour the *Sen3* markers do not contain Ora in their pedigree but do have Capella and BRA9089 (Table 2).

It is tempting to affirm that Kuba and Karolin share their resistance Identity-By-Descent. Unfortunately, insufficient pedigree information is available from Karolin. Kuba was tested with the flanking markers Kc8103 and RK36 and turned out to be positive (22). We tested Karolin with chr11_1259552 and chr11_1772869 from the current study and it was positive as well. Moreover, the authors tested their markers in a panel of potato varieties, 15 of which were included in our panel as well. Gandawa, Gawin, Ibis, Ikar, Karolin, Kuba, Rudawa and Ulme are positive to our markers as well as the markers developed by (22). However, the pedigree of these varieties offers many opportunities to conceive alternative routes of transmission of resistance genes from different pedigree branches. Our study enabled to trace back the *Sen3* locus until Ora. However, Ora is not present in the pedigree of all of the above-mentioned positive varieties (Table 2). With the current information, we are not able to call the wart resistances from Karolin and Kuba are identical-by-descent. CoSSA analysis would be suitable to prove sequence identity but genomic reads from Karolin are unfortunately not available. However, based on the markers produced in this study and on the *S.*

endobioticum pathotype resistance pattern, the Kuba and Karolin resistances are indistinguishable.

Since its development almost 30 years ago (1,2), Bulk Segregant Analysis strategies are used as an alternative to linkage map based QTL analysis and to identify new markers in unsaturated regions of genetic maps. BSA strategies have the main advantage of reducing the number of samples to be genotyped. In the case of CoSSA, we showed that the minimum requirement to identify *Sen3* was to sequence two samples with contrasting phenotypes (R-bulk and S-bulk) with a sequencing depth for each pool of 10x per haploid genome (i.e. 40x for a tetraploid genome), reducing radically the genotyping costs. Indeed, we observed that even in the case of a relatively high number of *k*-mers which were not inherited from any of the parents (likely due to DNA contamination), applying a filter on the resistant bulk specific *k*-mers depth drastically improved the signal-to-noise ratio and allowed to identify the resistance locus segregating (Supplementary figure 5A). If the two parents are included in the CoSSA workflow, the sequencing depth of the pools can be reduced further to 5x per haploid genome. Sequencing the two parents along with the bulks allowed to remove the previously described set of *k*-mers inherited from none of the parents. It has the other benefit of enabling the identification of minor effect loci. For instance, the minor effect locus inherited from Ludmilla on chromosome 9 would have not been identified without using the sequences of the two parents in the pipeline. Indeed, the SNR was too low, even with a sequencing depth of 16x per haploid genome, to be identified if the parents were not included in the analysis. Another alternative in CoSSA is to replace the resistant bulk by the resistant parent. This approach has the limitation that it will not identify minor QTLs from the susceptible parent. An advantage could be that it is more sensitive to minor QTLs from the resistant parent. A chromosome 2 peak (0-13Mb) was observed at the same position when using the full sequencing dataset (Figure 2). This peak was ignored initially due to low SNR. In future research, we might be able to validate this hypothesis.

The set algebra applied in the current CoSSA workflow is made possible by the fixed length of the subsequences. Such an approach based on set algebra has been implemented in the NIKS algorithm (43) to identify the causal mutations in populations derived from EMS mutagenized plants. The beginning of the NIKS workflow is similar to CoSSA: pools of mutants and wild types are sequenced, *k*-mer lists are produced for each bulk and *k*-mers with a low frequency are removed. Subsequent steps in the CoSSA workflow are clearly different and serve different goals. NIKS workflow is

designed to identify the mutation causal for a specific phenotype that segregates. Therefore, the NIKS workflow is a comparative subsequence sets analysis that identifies a “needle in the k -stack”. The CoSSA workflows presented in our study are designed to identify haplotype specific “ k -packs from the k -stack”. Moreover, NIKS workflow is not suited to identify k -stacks linked to dominant genes like *Sen3* because NIKS requires the haplotype of interest to be homozygous. Therefore, comparative subsequence set analysis is a powerful approach to address different genetic questions, but each question will require a different workflow.

Although the set algebra allows quick and unambiguous comparisons between samples, it may be stringent as it discards all the first bulk k -mers that occur at least 2 times in the second bulk. Therefore, CoSSA works best for major QTLs that can be phenotyped unambiguously. If the phenotyping is not straightforward, linked k -mers will be present both in the susceptible and resistant bulks. The use of the set algebra might also have the consequence of missing minor effect loci that will be present in both bulks. A way to reduce the stringency is to increase the number of individuals that compose the bulks. CoSSA offers another way to reduce stringency by increasing the S-bulk k -mers depth cut-off, which would not require a bigger population. Instead of removing from the R-bulk all the k -mers that occur at least 2 times in the S-bulk, we could remove only the k -mers which occur more than $\frac{x}{n} x m$ with x the depth of the S-bulk, n the number of individuals in the S-bulk, and m the number of phenotyping mistakes allowed (> 0).

The selection of haplotype specific variants was improved by including four potato varieties susceptible to the wart disease. Including these varieties in the CoSSA workflow halved the number of resistance specific k -mers. In other words, without this extra-step, the markers we designed had only 50% of chance to be haplotype specific. Moreover, we showed that the peak shifted from 3-4Mb to 1-2Mb when removing common k -mers present in the four susceptible varieties. This suggested that the 1-2Mb region contained SNPs with a higher haplotype specificity than the 3-4Mb region. This might be due to a recombination of the original *Sen3* haplotype, which replaced the centromeric part with a more common haplotype. Therefore, after the removal of the susceptible varieties k -mers, only the k -mers derived from the *Sen3* introgression remained. We improved the SNPs selection as well by applying a depth cut-off on the resistance specific k -mers. The lower depth cut-off allowed to remove noise coming from k -mers that were not linked to resistance but were, by chance, not present in the S-bulk or noise coming from PCR duplicated reads containing sequencing errors. The

upper cut-off, on the other hand, allowed to select for SNPs coming from single copy regions. The thresholds can be chosen by estimating the depth of the resistant haplotype according to the sequencing depth of the R-bulk, or by determining the k -mers frequency peaks such as in Supplementary figure 2.

The markers specific to the *Sen3* haplotype were tested in an independent panel of 56 tetraploid varieties (Table 2) and their diagnostic value was tested. Indeed, none of the susceptible varieties tested were positive to the two flanking markers. Among the 17 varieties that were positive, 14 are known to be resistant to pathotypes 1, 2, 6 and 18. Gawin, Rudawa and Sonda are only known to be resistant to P1 in addition to P18, P6 and P2 respectively. These incomplete resistance patterns could be explained by the fact that: 1. not all pathotypes were tested, 2. the isolates of the tested pathotypes used for phenotyping were derived from different sources and that their different genetic composition causes they are not recognised by *Sen3*, or 3. thresholds for resistance and susceptibility were not sufficiently uniform. We found in our KxL population plants that harboured both flanking markers but that were only partially resistant to P18. Future analyses will learn what the genetic basis for the partial or complete P18 resistance conferred by *Sen3* is.

The previously discussed strategy to include more varieties without the trait of interest in the CoSSA workflow to remove common SNPs could be expanded to an association mapping type of approach. Instead of applying the method on a segregating population, CoSSA could as well be applied on bulks from a broader panel of genotypes with opposite phenotypes. This strategy would be successful only when the causal genes are identical-by-descent. A similar method, called AgRenSeq, has recently been developed (13) and proved to be efficient to map *R* genes in *Aegilops tauschii*. The authors performed RenSeq, which requires prior knowledge about Nucleotide binding Leucine rich Repeat (NLR) genes repertoire and also introduces a bias to NLR rich regions of the genome and prevent the identification of other genes (non-NLR). CoSSA with WGS data is an unbiased genome wide approach that involves moderate sequencing costs and computation extensive k -mer counting instead of computation intensive reference genome dependent variant calling. An untargeted association mapping method based on the use of k -mers (HAWK) has been developed (12). CoSSA differs from HAWK by its simplicity as it does not require advanced statistical methods or strong bioinformatics skills. Moreover, the last step of the CoSSA workflow consisting of assembling the read pairs containing the resistance specific k -mers instead of the k -mers themselves as in the HAWK pipeline should give a better quality of the

assembly. Indeed, by using the k -mers, one loses the advantage of the paired end reads that can help resolving repetitive regions as well as duplicated or regions harbouring low complexity levels. We overcame this drawback by selecting the haplotype specific read pairs using the resistance specific k -mers.

The *de novo* assembly of the reads containing the resistance specific k -mers and the mapping of the resistance specific k -mers to the *de novo* scaffolds allows to reduce the number of candidate SNPs for which to design markers. Indeed, mapping-free SNP calling methods have been developed (44–46) and could be used in a BSA setup. However, without an assembly, more SNPs would have to be tested to map the locus. Moreover, having the sequences flanking the candidate variants improves the quality of the markers. Additionally, the *de novo* assemblies allow the identification of candidate genes. For instance, (13) using NLR-parser (47) could search for NLR genes in *de novo* assemblies. However, in the case of NLR genes, it has been shown that a *de novo* assembly based on short reads might result in chimeric scaffolds (48). In the same publication, by using PacBio long reads, the authors successfully assembled a complex resistance gene cluster and cloned the causal resistance gene.

In this paper, we mapped *Sen3* to a region containing the complex *R* gene cluster C76 which harbours 8 TNLs (7). The first locus involved in potato wart disease resistance identified, *Sen1*, was also mapped to chromosome 11 near the C76 cluster (16) (Figure 3). *Sen3* could be a different paralog from the same cluster as *Sen1*, but it could also be an allelic variant of *Sen1*. Only fine-mapping and the cloning of these genes will answer this question. The C76 cluster is in fact even more complex as the assembly of the DM genome is incomplete in this region (Figure 3). By adding contig DMB734, we added 3 TNLs and still at least one additional gap remains. This complexity was also reflected in the *de novo* assembly we generated from the resistant haplotype specific reads which was very scattered. Thus, we were unsuccessful in assembling the *R* gene cluster using the short reads. However, our analyses showed clearly that our assembly of the *R* haplotype was more complete than DM. Only 11.9% of the resistance specific k -mers that mapped to the three longest scaffolds of the *de novo* assembly could be mapped back to the reference genome with a perfect match and only 38.2% could be mapped allowing one mismatch. This matched our observations that the chromosome 11 *R* gene cluster but also many other *R* gene regions in DM are incomplete.

Conclusions

Overall, the CoSSA workflows were successful in mapping the potato wart disease resistance locus segregating in a full sib population of tetraploid non-inbred parents, with or without the use of the potato reference genome. This method is both quick and cheap as the bioinformatics pipeline required approximately only half a day and the price of whole genome sequencing using short reads is still decreasing. CoSSA is highly efficient as well: in the reference-free experiment, we needed to screen only three KASP markers to identify the resistance locus and design two valuable flanking markers. Moreover, we showed that the CoSSA *de novo* assembly of the R haplotype was more complete than the potato reference genome. The CoSSA workflows offer many possibilities to the user such as the quick mapping of major and minor effect loci, the identification of haplotype specific variants, the *de novo* assembly of haplotype specific reads and the *in silico* inspection of genotypes sharing the same trait. Finally, the different filters settings along the pipeline allow the user to adapt CoSSA to diverse crops and traits.

Declarations

Authors' contributions

CP, HJvE and JHV conceived and designed the experiment. CP performed experiments and all the analyses. DE and TB provided bioinformatics support. CP analysed all the results with the help of HJvE and JHV. CP wrote the manuscript. JHV, RGFV and HJvE helped to write, improve and review the latest version of this manuscript. All authors read and approved the final manuscript.

Acknowledgments

We thank all the partners involved in this public-private partnership (Averis seeds BV, Böhm-Nordkartoffel Agrarproduktion GmbH & Co. OHG, Danespo, HLB BV, HZPC Holland BV, C Meijer BV, SaKa Pflanzenzucht GmbH & Co. KG and Teagasc) for providing invaluable potato material, phenotypic data, KASP marker analysis, and scientific discussions. Wilbert van Workum and Limes Innovations are acknowledged for facilitating WGS at Hartwig Medical. We acknowledge the very kind help from Dr. Danuta Sekrecka, Laboratory of Potato Gene Resources and Tissue Culture, IHAR, Department of Potato Protection and Seed Science at Bonin, Poland, for providing varieties. A warm thank you goes also to Jarosław Przetakiewicz and IHAR who phenotyped the KxL population. Klaus Dehmer from IPK Gross Lüsewitz Potato Collection is acknowledged for providing material of Capella and BRA9089. We finally

acknowledge Bernd Truberg and Martin Sloksnat for inspiring discussions and pedigree information.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The datasets used and/or analysed during the current study are available in the published manuscript or available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Funding

This work was performed in a project called “An integrated genomics and effectormics impulse for potato wart resistance management and breeding” which was financially supported by the Dutch Topsector Horticulture & Starting Materials. Within the Topsector, private industry, knowledge institutes and the government are working together on innovations for sustainable production of safe and healthy food and the development of a healthy green environment. CP was supported by a Walsh fellowship provided by Teagasc.

Endnotes

^a To validate that the clone BRA9089 that we sequenced was indeed not the resistance donor, we decided to re-phenotype it. The phenotyping results showed that our sample of BRA9089 was indeed susceptible to pathotypes 1, 2, 6 and 18 (mean scores of 4.4, 4.2, 5 and 4.8 respectively). We also genotyped BRA9089 with chr11_1259552 and chr11_1772869 which was negative to the tested markers (Table 2).

References

1. Giovannoni JJ, Wing RA, Ganai MW, Tanksley SD. Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations. *Nucleic Acids Res.* 1991;19(23):6553–6568.
2. Michelmore RW, Paran I, Kesseli RV. Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proc Natl Acad Sci.* 1991;88(21):9828–9832.
3. Quarrie SA, Lazić-Jančić V, Kovačević D, Steed A, Pekić S. Bulk segregant analysis with molecular markers and its use for improving drought resistance in maize. *J Exp Bot.* 1999;50(337):1299–1306.
4. Becker A, Chao D-Y, Zhang X, Salt DE, Baxter I. Bulk Segregant Analysis Using Single Nucleotide Polymorphism Microarrays. El-Sherny H, editor. *PLoS ONE.* 2011 Jan 27;6(1):e15993.
5. Schneeberger K, Ossowski S, Lanz C, Juul T, Petersen AH, Nielsen KL, *et al.* SHOREmap: simultaneous mapping and mutation identification by deep sequencing. *Nat Methods.* 2009 Aug 1;6:550.
6. Liu S, Yeh C-T, Tang HM, Nettleton D, Schnable PS. Gene mapping via bulked segregant RNA-Seq (BSR-Seq). *PloS One.* 2012;7(5):e36406.
7. Jupe F, Witek K, Verweij W, Śliwka J, Pritchard L, Etherington GJ, *et al.* Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. *Plant J.* 2013 Nov;76(3):530–44.
8. Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, *et al.* Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. Fay JC, editor. *PLoS ONE.* 2008 Oct 13;3(10):e3376.
9. Hartwig B, James GV, Konrad K, Schneeberger K, Turck F. Fast Isogenic Mapping-by-Sequencing of Ethyl Methanesulfonate-Induced Mutant Bulks. *PLANT Physiol.* 2012 Oct 1;160(2):591–600.
10. Montenegro JD, Golicz AA, Bayer PE, Hurgobin B, Lee H, Chan C-KK, *et al.* The pangenome of hexaploid bread wheat. *Plant J.* 2017 Jun;90(5):1007–13.
11. Cantalapiedra CP, Contreras-Moreira B, Silvar C, Perovic D, Ordon F, Gracia MP, *et al.* A Cluster of Nucleotide-Binding Site–Leucine-Rich Repeat Genes Resides in a Barley Powdery Mildew Resistance Quantitative Trait Loci on 7HL. *Plant Genome.* 2016;9(2):0.
12. Rahman A, Hallgrímsdóttir I, Eisen M, Pachter L. Association mapping from sequencing reads using k-mers. *eLife.* 2018;7:e32920.
13. Arora S, Steuernagel B, Gaurav K, Chandramohan S, Long Y, Matny O, *et al.* Resistance gene cloning from a wild crop relative by sequence capture and association genetics. *Nat Biotechnol.* 2019 Feb;37(2):139–43.
14. Przetakiewicz J. The Viability of Winter Sporangia of *Synchytrium endobioticum* (Schilb.) Perc. from Poland. *Am J Potato Res.* 2015 Dec;92(6):704–8.
15. PM 9/5 (2) *Synchytrium endobioticum*. *EPPO Bull.* 2017 Dec;47(3):511–2.
16. Hehl R, Faurie E, Hesselbach J, Salamini F, Whitham S, Baker B, *et al.* TMV resistance gene *N* homologues are linked to *Synchytrium endobioticum* resistance in potato. *Theor Appl Genet.* 1999;98(3):379–386.
17. Brugmans B, Hutten RGB, Rookmaker ANO, Visser RGF, van Eck HJ. Exploitation of a marker dense linkage map of potato for positional cloning of a wart disease resistance gene. *Theor Appl Genet.* 2006 Jan;112(2):269–77.
18. Plich J, Przetakiewicz J, Śliwka J, Flis B, Wasilewicz-Flis I, Zimnoch-Guzowska E. Novel gene *Sen2* conferring broad-spectrum resistance to *Synchytrium endobioticum* mapped to potato

- chromosome XI. Theor Appl Genet [Internet]. 2018 Aug 9 [cited 2018 Aug 15]; Available from: <http://link.springer.com/10.1007/s00122-018-3154-y>
19. Ballvora A, Flath K, Lübeck J, Strahwald J, Tacke E, Hofferbert H-R, *et al.* Multiple alleles for resistance and susceptibility modulate the defense response in the interaction of tetraploid potato (*Solanum tuberosum*) with *Synchytrium endobioticum* pathotypes 1, 2, 6 and 18. Theor Appl Genet. 2011 Dec;123(8):1281–92.
20. Groth J, Song Y, Kellermann A, Schwarzfischer A. Molecular characterisation of resistance against potato wart races 1, 2, 6 and 18 in a tetraploid population of potato (*Solanum tuberosum* subsp. *tuberosum*). J Appl Genet. 2013 May;54(2):169–78.
21. Obidiegwu JE, Sanetomo R, Flath K, Tacke E, Hofferbert H-R, Hofmann A, *et al.* Genomic architecture of potato resistance to *Synchytrium endobioticum* disentangled using SSR markers and the 8.3k SolCAP SNP genotyping array. BMC Genet [Internet]. 2015 Dec [cited 2017 Aug 3];16(1). Available from: <http://www.biomedcentral.com/1471-2156/16/38>
22. Bartkiewicz A, Chilla F, Terefe-Ayana D, Lübeck J, Strahwald J, Tacke E, *et al.* Improved genetic resolution for linkage mapping of resistance to potato wart in monoparental dihaploids with potential diagnostic value in tetraploid potato varieties. Theor Appl Genet [Internet]. 2018 Aug 29 [cited 2018 Sep 11]; Available from: <http://link.springer.com/10.1007/s00122-018-3172-9>
23. JKI. [online] Available from: https://pflanzengesundheit.julius-kuehn.de/dokumente/upload/29991_bm-2018resistente_kartoffelsorten.pdf [Accessed 15 Oct. 2018]. 2018;
24. Glynne MD. Infection experiments with wart disease of potatoes. *Synchytrium endobioticum* (Schilb.) Perc. Ann Appl Biol. 1925;12(1):34–60.
25. Lemmerz J. A new simplified method of inoculation of potato cultivars to test for wart resistance. Züchter. 1930;2:288–297.
26. Fulton TM, Chunwongse J, Tanksley SD. Microprep protocol for extraction of DNA from tomato and other herbaceous plants. Plant Mol Biol Report. 1995;13(3):207–209.
27. Murray MG, Thompson WF. Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res. 1980;8(19):4321–5.
28. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Aug 1;30(15):2114–20.
29. Kaplinski L, Lepamets M, Remm M. GenomeTester4: a toolkit for performing basic set operations - union, intersection and complement on *k*-mer lists. GigaScience [Internet]. 2015 Dec [cited 2017 Jul 25];4(1). Available from: <https://academic.oup.com/gigascience/article-lookup/doi/10.1186/s13742-015-0097-y>
30. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009 Jul 15;25(14):1754–60.
31. Potato Genome Sequencing Consortium, Pan S, Cheng S, Zhang B, Mu D, Ni P, *et al.* Genome sequence and analysis of the tuber crop potato. Nature. 2011 Jul 10;475(7355):189–95.
32. Sharma SK, Bolser D, de Boer J, Søndergaard M, Amorós W, Carboni MF, *et al.* Construction of reference chromosome-scale pseudomolecules for potato: integrating the potato genome with genetic and physical maps. G3 Genes Genomes Genet. 2013;3(11):2031–2047.
33. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010 Mar 15;26(6):841–2.
34. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, *et al.* SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol. 2012 May;19(5):455–77.
35. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, *et al.* The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078–9.

36. Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics*. 2015 Jun 15;31(12):2032–4.
37. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp 365–386.
38. Bakker E, Borm T, Prins P, van der Vossen E, Uenk G, Arens M, *et al*. A genome-wide genetic map of NB-LRR disease resistance loci in potato. *Theor Appl Genet*. 2011 Aug;123(3):493–508.
39. Vossen JH, Dezhsetan S, Esselink D, Arens M, Sanz MJ, Verweij W, *et al*. Novel applications of motif-directed profiling to identify disease resistance genes in plants. *Plant Methods*. 2013;9(1):37.
40. Bukasov SM, Kameraz AJA. Bases of potato breeding. Gosudarstvennoe Izdatel'stvo Sel'skhozjajstvennoj Literatury, Moskva, Leningrad. 1959;528.
41. Ross H. Potato breeding - Problems and perspectives. Suppl. to *J. Plant Breeding*, Paul Parey, Berlin and Hamburg. 1986.
42. Murphy AM, De Jong H, Proudfoot KG, Proudfoot KG. A multiple disease resistant potato clone developed with classical breeding methodology. *Can J Plant Pathol*. 1999 Sep;21(3):207–12.
43. Nordström KJV, Albani MC, James GV, Gutjahr C, Hartwig B, Turck F, *et al*. Mutation identification by direct comparison of whole-genome sequencing data from mutant and wild-type individuals using *k*-mers. *Nat Biotechnol*. 2013 Apr;31(4):325–30.
44. Audano PA, Ravishankar S, Vannberg FO. Mapping-free variant calling using haplotype reconstruction from *k*-mer frequencies. Berger B, editor. *Bioinformatics*. 2018 May 15;34(10):1659–65.
45. Gardner SN, Hall BG. When Whole-Genome Alignments Just Won't Work: kSNP v2 Software for Alignment-Free SNP Discovery and Phylogenetics of Hundreds of Microbial Genomes. Tang H, editor. *PLoS ONE*. 2013 Dec 9;8(12):e81760.
46. Pajuste F-D, Kaplinski L, Möls M, Puurand T, Lepamets M, Remm M. FastGT: an alignment-free method for calling common SNVs directly from raw sequencing reads. *Sci Rep* [Internet]. 2017 Dec [cited 2018 Aug 2];7(1). Available from: <http://www.nature.com/articles/s41598-017-02487-5>
47. Steuernagel B, Jupe F, Witek K, Jones JDG, Wulff BBH. NLR-parser: rapid annotation of plant NLR complements. *Bioinformatics*. 2015 May 15;31(10):1665–7.
48. Witek K, Jupe F, Witek AI, Baker D, Clark MD, Jones JDG. Accelerated cloning of a potato late blight-resistance gene using RenSeq and SMRT sequencing. *Nat Biotechnol*. 2016 Apr 25;34(6):656–60.
49. van Berloo R, Hutten RCB, van Eck HJ, Visser RGF. An Online Potato Pedigree Database Resource. *Potato Res*. 2007 Jan;50(1):45–57.

Supplementary figures

Supplementary figure 1. Identification of stacks of 31 k -mers to design KASP markers

To identify a good candidate SNP to design diagnostic markers, we mapped the resistance specific k -mers minus S varieties to the reference genome and counted how many k -mers map to each chromosome bin. Each resistant haplotype specific SNP can have maximum k k -mers (in our case, 31 31-mers) mapped to itself. We designed most KASP markers for SNPs which had 31 k -mers mapped under the main k -mers peaks. Fluorescent dyes: F = FAM; H = HEX.

Supplementary figure 2. k -mers depth distribution

The total number of k -mers (unique k -mers x depth) for each sequencing depth from 2 to 100 are represented in this graph for the 4 different samples (red: Kuba, blue: Ludmilla, green: R-bulk, yellow: S-bulk). For the two varieties samples (Kuba and Ludmilla), the peaks of k -mers from simplex, duplex, triplex and quadruplex regions are visible. There is a shift in the simplex peak of the bulks as there are 8 possible haplotypes instead of 4 as in the parental genotypes.

Supplementary figure 3. Pedigree of Kuba

The pedigree tree of Kuba (according to (49)) and the known resistance pattern of its ancestors. BRA9089 is present in the pedigree of several varieties resistant to pathotypes 2, 6 and 18 and is thought to be the ancestral donor of their resistance. BRA9089 is present in the resistant grand-parent pedigree (Bzura) and in the susceptible grand-parent (Karlana) pedigree as well. under the main k-mers peaks. Fluorescent dyes: F = FAM; H = HEX.

Supplementary figure 4. CoSSA-based pedigree analysis

Number of unique *k*-mers in common between the resistance specific *k*-mers without the S varieties *k*-mers and Bzura (the resistant parent of Kuba), BRA9089 (putative resistant donor) and Kuba, (A) mapped to the reference genome (chromosome 11, 0 - 7Mb, DMB734 *k*-mers were added to the 1-2Mb bin) (B) mapped to the 10 longest scaffolds of the *de novo* assembly. Orange: Kuba, blue: Bzura, black: BRA9089. The bars numbers represent the percentages of the intersection *k*-mers.

Supplementary figure 5. Assessment of the depth cut-offs on the CoSSA results

Assessment of the effect of the lower and upper depth cut-off on the CoSSA output. (A) CoSSA output without the lower depth cut-off (R-bulk specific *k*-mers with a depth from 2 to 22x), (B) CoSSA output with the lower and upper cut-offs (R-bulk specific *k*-mers with a depth from 10 to 22x), (C) CoSSA output without the upper depth cut-off (R-bulk specific *k*-mers with a depth from 10 to ∞). Red: *k*-mers inherited from Kuba (resistance specific *k*-mers), blue: *k*-mers inherited from Ludmilla, green: *k*-mers inherited from both parents, grey: *k*-mers inherited from none of the parents. (D) Signal to noise ratio (SNR) for *Sen3* and *Ludmilla_chr9*. Black: SNR for CoSSA results when no lower cut-off is applied (2 to 22x),

stripped: SNR for CoSSA results when an upper and a lower cut-offs are applied (10 to 22x),
grey: SNR for CoSSA results when no upper cut-off is applied (22 to ∞).

Supplementary figure 6. CoSSA applied with Kuba and the S-bulk only

Kuba specific *k*-mers (depth 8x to 18x) obtained by the difference between Kuba and the S-bulk mapped to the reference genome.

Supplementary tables

Supplementary table 1. Statistics of the *de novo* assembly of the resistant haplotype

Sequencing depth per haploid genome	Number of scaffolds	Longest scaffold (bp)	Mean scaffold size (bp)	N50
~16x	21,306	46,690	1,037	1,115

Supplementary table 2. Comparison of the CoSSA workflows with/without reference

Scaffold	Length	# of mapped <i>k</i> -mers	# of <i>k</i> -mers mapped to DM 0 mismatches	# of <i>k</i> -mers mapped to DM 1 mismatch	# of <i>k</i> -mers mapped to DM 2 mismatches
NODE_1	46,690	11,511	340 (2,95%)	2,947 (25.60%)	5,199 (45,17%)
NODE_2	44,756	8,420	49 (0.58%)	2,788 (33.11%)	5,118 (60.78%)
NODE_3	38,969	6,629	2,774 (41.85%)	4,417 (66.63%)	5,426 (81,85%)

Supplementary table 3. Datasets used in the CoSSA simulations with lower sequencing depths

	Sample name	# read pairs	Produced total depth	Depth / haploid genome	<i>k</i> -mers cut-off depth
16x	R-bulk	172,190,144	61.9	15.5	[10; 22]
	S-bulk	228,585,110	82.2	20.5	
	Kuba	200,579,289	72.1	18.0	
	Ludmilla	195,099,847	70.1	17.5	
10x	R-bulk	111,258,278	40	10	[6; 14]
	S-bulk				
	Kuba				
	Ludmilla				
5x	R-bulk	55,629,139	20	5	[3; 7]
	S-bulk				
	Kuba				
	Ludmilla				

Supplementary files list

Supplementary files can be accessed at <https://doi.org/10.17026/dans-2zf-pnd3>

> Supplementary file 1 (Chapter3_S_file_1.xlsx)

The Excel sheet comprises: the genotypic data for all the tested markers, the number of tubers tested per genotype, the average score of the Glynne-Lemmerzahl assays performed in 2016 and 2017 for the pathotypes 2, 6 and 18, the individuals included in the subset ($n = 83$) used for the validation of the CoSSA results and the individuals selected for the R-bulk and the S-bulk.

> Supplementary file 2 (Chapter3_S_file_2.pdf)

Computational requirements for the biggest datasets we used in the CoSSA workflows (sequencing depth of 16x/haploid genome). We showed that a depth of 10x/haploid genome is sufficient when the parents are not sequenced and 5x when parents are sequenced which reduces these computational requirements. The different jobs were executed on the WUR Plant Breeding server. Only one node of 10.5GB memory was used. When the jobs are parallelized for the different CoSSA samples, the reference genome dependent workflow can be run in half a day. For the reference genome independent CoSSA workflow, we advise to allocate more nodes to the reads extraction and to the *de novo* assembly steps as they require more time. The most time consuming step at the moment which could be improved is the script used to extract the read pairs containing the k -mers.

> Supplementary file 3 (Chapter3_S_file_3.xlsx)

For each KASP marker used in the study is given: the marker name (chromosome_position), the experiment in which it was used, the population in which it was used, the two forward primer sequences and the common primer sequence.

> Supplementary file 4 (Chapter3_S_file_4.pdf)

Resistance distribution in the full-sib population ($n = 328$) for pathotypes 2 (A), 6 (B) and 18 (C). The phenotyping was performed with the Glynne-Lemmerzahl method in 2016. For each resistance category, the proportion of individuals with the R allele of *Sen3* (black), without the R allele (grey) and the proportion of recombinants (black stripes) is given. In (D) are given the Pearson correlations between the resistance mean scores of P2, P6 and P18. Chi-square tests were conducted to validate if the resistance to the three pathotypes segregated in a 1:1 ratio in the population (E).

> Supplementary file 5 (Chapter3_S_file_5.xlsx)

For all the samples sequenced, the number of reads and the sequencing depth is given. The number of unique k -mers (N_u) and the number of total k -mers (N_t) is given for all the sequenced samples and several subsets obtained after basic set operations for different depth cut-offs.

> Supplementary file 6 (Chapter3_S_file_6.pdf)

Number of R-bulk specific k -mers (depth 10 to 22) mapping to each 1Mb bin of (A) chromosome 0, (B) chromosome 3, (C) chromosome 4, (D) chromosome 5, (E) chromosome 9 and (F) chromosome 11 of the potato reference genome DM. Red: k -mers inherited from Kuba (resistance specific k -mers), yellow: k -mers inherited from Kuba minus the k -mers

present in the susceptible varieties, blue: *k*-mers inherited from Ludmilla, green: *k*-mers inherited from both parents.

> Supplementary file 7 (Chapter3_S_file_7.pdf)

Boxplots of the resistance scores for P2, P6 and P18 in function of the allele (0: S allele, 1: R allele) of the KASP markers chr11_1519485 (A) and chr09_55113777 (B) in the KxL population subset (n = 83).

Chapter 4

Sen1 fine-mapping and candidate gene analysis

Charlotte Prodhomme, Gert van Arkel, Marjan Bergervoet,
Danny Esselink, Johan Rijk, Moffat Makechemu,
Xizheng Chen, Michiel Schrijen, Joop Hillen,
Richard G. F. Visser, Herman J. van Eck, Jack H. Vossen

Abstract

Key message

Sen1, the potato wart disease *R* gene that confers recognition of *AvrSen1*, is a TNL encoding gene that resides in a cluster, syntenic to C76, on the north arm of chromosome 11.

Abstract

Synchytrium endobioticum is a soil borne obligate biotrophic fungus responsible for wart disease in potato (*Solanum tuberosum*). Potato wart disease has a worldwide quarantine status as winter spores of *S. endobioticum* can remain viable in the soil for more than 40 years, yield losses can reach 50-100% and no fungicides are efficient against the pathogen. Therefore, it is essential to breed for resistant varieties to manage the disease and to identify the *R* genes involved. So far, several loci involved in qualitative and quantitative resistance to the potato wart disease have been mapped, but nothing is known about the underlying genes. *Sen1* was the first wart resistance locus mapped and brings resistance to the pathotype 1 (D1) of the wart disease. Two bi-parental populations in which *Sen1* was segregating have been used to identify haplotype specific markers flanking *Sen1*. Recombinants were screened for in a population of 1,104 seedlings to fine-map *Sen1*. The fine-map position of *Sen1* encompassed a TNL cluster on the north arm of chromosome 11 of the reference genome DM. An RNAi construct silencing TNL genes from chromosome 11 confirmed that *Sen1* is an *R* gene from the TNL group, however the exact TNL gene is still unknown. A phased assembly of a tetraploid variety holding *Sen1* was used to continue the fine-mapping of *Sen1* and to identify TNL candidate genes. Complementation assays were performed in *Nicotiana benthamiana* and in potato to identify which candidate TNL gene caused recognition of *AvrSen1*.

Keywords

Potato wart disease, fine-mapping, Comparative Subsequence Sets Analysis (CoSSA), *R* gene cluster, TNL, tetraploid, *de novo* phased assembly.

Introduction

Potato wart disease, caused by the obligate biotrophic fungus *Synchytrium endobioticum*, is a quarantine disease causing serious losses in potato (*Solanum tuberosum*) cultivation. *S. endobioticum* is a soil born fungus, member of the phylum Chytridiomycota. The formation of galls, a characteristic symptom of the wart disease,

is due to a hypertrophic growth of the infected meristematic tissues. Winter spores, that can remain viable in the soil for more than 40 years (1), are formed in the infected tissues. Because of this long infectiousness combined with the lack of efficient fungicides (2) and the yield losses ranging from 50-100% (3), potato wart disease is categorized as a quarantine pest world-wide (4) and its management depends on the cultivation of resistant varieties. To date, more than 39 pathotypes of *S. endobioticum*, which are defined according to their virulence patterns on a differential set of potato varieties (5), have been identified (6). Several genetic studies were successful to map potato resistance (*R*) genes and QTLs for pathotype specific or broad-spectrum resistance. However, up to now, no gene involved in potato wart disease resistance has been cloned and the type of genes responsible for resistance is still unknown.

Plants rely on a complex innate immunity system to recognize and fight pathogen attacks. The majority of the genes involved in resistance that are cloned so far encode intracellular Nucleotide-binding domain, Leucine Rich containing (NLR) receptor proteins that recognize effector proteins secreted by the pathogen in the cytoplasm. The NLR proteins consist of a central nucleotide-binding domain (NB), a C-terminal leucine-rich repeat domain (LRR) and an N-terminal domain consisting either of a coiled-coil domain (CC) for the CNL group or of a TIR (Toll, interleukin-1 receptor) domain for the TNL group (7). In the potato reference genome DM1-3 516 R44 (DM) (8), 755 NLR genes have been identified among which 584 belong to the CNL group and 157 to the TNL group (9). So far, only three functional TNL genes have been cloned in potato. Y-1 is a TNL giving an hypersensitive response (HR) upon infection by the potato virus Y (PVY) and is located on the north arm of chromosome 11 (10). *Gro1-4*, a TNL involved in resistance to the nematode *Globodera rostochiensis* (11), is located on chromosome 7. Recently, the *Rysto* gene from *S. stoloniferum*, conferring extreme resistance to Potato Virus Y (PVY) and located on the south arm of chromosome 12, has been cloned as well (12). In other species, TNLs are involved in resistance to several pathogens such as bacteria (13), fungi (14,15), nematodes (16), oomycetes (17) and viruses (18).

The first gene involved in potato wart disease resistance identified was *Sen1* which brings resistance to the pathotype 1 (D1) isolates. *Sen1* was mapped for the first time in a diploid population to the north arm of chromosome 11 (19). This region is known to contain *R* gene clusters (9) in the potato reference genome and several *R* genes have already been successfully introgressed in this region from different sources such as *S. andigena* (20), *S. avilesii* (21), *S. capsicobaccatum*, *S. circaefolium* ssp. *quimense*

(22) and *S. stoloniferum* (23). Since its first identification, *Sen1* has been mapped in several populations with different genetic backgrounds (24–27) but no candidate genes were identified and examined. Recently, we performed a Genome Wide Association Study (GWAS) in a panel of 330 potato cultivars representative of the commercial potato gene pool (28). SNP markers significantly linked with *Sen1* in the GWAS panel were validated in two independent tetraploid populations. We mapped *Sen1* to the chromosome 11 region distal to PotVar0066530 (2.7 Mb) which is in agreement with (19) where *Sen1* mapped distal to the markers NI25 and NI27. In DM, this region contains the two *R* gene clusters C76 and C77 which contain only TNL genes (9). Therefore, we hypothesized that *Sen1* is an NLR gene from the TNL group. Interestingly, the *Sen3* gene was mapped to a syntenic locus and could also encode a TNL.

In this study, we used a Comparative Subsequence Sets Analysis (CoSSA) (29) in the two sibling populations containing *Sen1* from (28) to identify haplotype specific SNPs and for the fine-mapping of *Sen1* in a recombinant population. An RNAi construct silencing genes with homology to clusters C76 and C77 of chromosome 11 was designed to validate the hypothesis that *Sen1* is an NLR gene from the TNL group. For the further fine-mapping of *Sen1* and the identification of candidate genes, we used a phased *de novo* assembly of the tetraploid potato variety Colomba. The candidate genes remaining after fine mapping and expression analysis were cloned and complementation assays were performed in *N. benthamiana* and in *S. tuberosum*.

Materials and Methods

Plants materials

For the fine-mapping of *Sen1*, we used the two populations Aventura x Desiree (AxD) and Kuras x Aventura (KxA) from (28), segregating for *S. endobioticum* pathotype 1 (D1) resistance, hereafter referred as P1. A resistant bulk, named AxDK_Sen1_RB, was composed of 13 and 11 descendants from the AxD and the KxA populations respectively (Supplementary file 1). The descendants from the AxDK_Sen1_RB bulk were selected because of their resistance to P1 and because they hold the minor allele of PotVar0066530, linked to *Sen1*. A susceptible bulk, named AxDK_SB, composed of 5 and 5 progeny clones from the AxD and the KxA populations respectively, were selected based on their susceptibility to P1 and the absence of the minor allele of PotVar0066530.

To identify more recombinants, an additional 1,104 seedlings from Aventra x Desiree were sown *in vitro*. Seedlings were screened with two KASP (Kompetitive Allele Specific PCR) markers flanking *Sen1* to identify recombinants.

DNA extraction, Whole Genome Sequencing with short reads and Nanopore sequencing

Genomic DNA of the descendants selected for the bulks (AxDK_Sen1_RB bulk and AxDK_SB) and the three parents (Aventra, Desiree, Kuras) was extracted from freshly harvested leaves according to (30) and purified using the Qiagen DNeasy Plant Mini Kit. The DNA concentration was assessed using a Qubit Fluorometer (Invitrogen). For AxDK_Sen1_RB and AxDK_SB, 42 ng and 100 ng of DNA of each individual was pooled respectively. For all the WGS samples, 1 µg of (pooled) genomic DNA was used for the library preparation. Each sample was sequenced on an Illumina platform generating 150 bp paired end (PE) reads (Hartwig Medical foundation, Amsterdam, The Netherlands).

The DNA extraction of the AxD (n = 42) and the AxK populations (n = 35) was performed by (28) from freshly harvested freeze-dried tubers or fresh leaves according to (30). Genomic DNA of the recombinants screening population (Aventra x Desiree, n = 1,104) was extracted from freshly harvested leaves from *in vitro* material according to (30). The DNA quality was assessed on ethidium bromide containing agarose gels. For the markers assays, the DNA concentration was estimated using a Nanodrop ND-1000 spectrophotometer (Thermo Scientific) and adjusted to a concentration of 5 to 50 ng/µl.

DNA of the variety Colomba was extracted with the protocol described in (31) with some modifications. The DNA library was constructed using the 1D DNA ligation sequencing kit SQK-LSK109. 1 µg of DNA, measured with a Qubit Fluorometer (Invitrogen), was used. The DNA library was sequenced on MinION using R9.4.1 FLO-MIN106 flow cells (Oxford Nanopore Technologies).

RNA extraction and RNAseq

In vitro plants of Deodara, Desiree, Kuras and the Kuras RNAi plant AI124-03 were grown three weeks in the greenhouse to reach the developmental stage used for agroinfiltrations. Leaf samples were taken from fully grown leaves. RNA was extracted for the three genotypes with the RNeasy Mini Kit (Qiagen, The Netherlands). Concentration and quality of the RNA was assessed using a Nanodrop ND-1000 spectrophotometer (Thermo Scientific). For Desiree, 1.25 µg in 25 µL was sent for

sequencing on a NextSeq500 platform with 150 bp PE reads (MAD Dutch Genomics Service & Support Provider, Amsterdam, Netherlands). For Deodara, Kuras and AI124-03, 3 µg in 30 µL were sent for sequencing on a BGISEQ-500 platform (BGI-Shenzhen, Shenzhen, China) with 100 bp PE, 150 bp PE and 150 bp PE reads respectively.

CoSSA to map and fine-map *Sen1* using the reference genome

To identify SNPs linked in coupling phase with *Sen1*, we conducted a Comparative Subsequence Sets Analysis (CoSSA) as described in (29). In brief, *k*-mer tables (*k* = 31) were produced using the glistmaker function of GenomeTester4 (32) for AxDK_Sen1_RB, AxDK_SB and the three parents. The *k*-mers with a frequency (~ depth) of 1 were removed. To select for *k*-mers specific to the AxDK_Sen1_RB bulk, the difference between AxDK_Sen1_RB and AxDK_SB was made using the glistcompare function of GenomeTester4. Moreover, to select for *k*-mers coming from single copy and simplex sequences, only AxDK_Sen1_RB bulk specific *k*-mers with a frequency between 9x and 27x were kept. These *k*-mers were then divided in eight different sets: inherited from Desiree, from Kuras, from Avenra, from Desiree and Kuras (hereafter called *Sen1* specific *k*-mers), from Desiree and Avenra, from Kuras and Avenra, from the three parents and from none of the parents. This last set was discarded from further analysis as it contained sequencing errors and DNA contamination. The seven AxDK_Sen1_RB bulk specific *k*-mers sets were mapped to the potato reference genome *Solanum phureja* DM1-3 (DM) version 4.03 (8,33). The number of *k*-mers mapping to each chromosome 1Mb bin were counted using the bedtools suite (v2.25) (34) and plotted in Microsoft Excel.

Phenotyping using *AvrSen1*

The 56 recombinants were phenotyped with the effector *AvrSen1* which is specifically recognized by *Sen1*. *AvrSen1* was transiently expressed in the leaves of the recombinants and positive and negative controls as described in (35). For each recombinant, 3 leaves/plant and 3 plants/genotype were agroinfiltrated with a positive control, a negative control and *AvrSen1*. The optical density (OD) used for each construct was 0.3. When results were not clear, the experiments were repeated. The positive controls used were either the co-agroinfiltration of *R8/Avr8* (36) or the co-agroinfiltration of *R3b/Avr3b* (37). The negative controls used were *R8* or *R3b* alone. Negative and positive control plants known to contain or not *Sen1* were included in each experiment. Three days after the agroinfiltration, the Hypersensitive Responses (HR) were scored with a quantitative scale ranging from 0 (no HR), to 1 (50 % cell

death), to 2 (100 % cell death in the infiltrated zone) (35). In the end, each recombinant was given a qualitative score in function of its recognition (HR) or non-recognition (n) of *AvrSen1*.

The RNAi plants were also phenotyped for their HR responses to *AvrSen1*. As the silencing through the RNAi system is partial, we used different ODs for the agroinfiltration of *AvrSen1* to measure the reduction in *Avrsen1* response. ODs of 0.3, 0.1, 0.03, 0.01 were used in the first three experiments and ODs of 0.02, 0.01, 0.005 and 0.0025 were used in the last experiment. For each RNAi plant and the untransformed genotypes (Desiree and Kuras), 3 leaves/plant and 3 to 6 plants/genotype were agroinfiltrated. The phenotyping of RNAi plants which showed HR reduction was repeated. Two days after the agroinfiltration, a quantitative score was given for each agroinfiltration. Kruskal-Wallis test and post-hoc non-parametric multiple comparisons Dunn's test were performed to validate if the HR responses differed between the genotypes. All the statistical tests were performed on R v3.5.0 (38).

Selection of *Sen1* haplotype specific scaffolds

A haplotype-phased *de novo* assembly of the tetraploid variety Colomba which is resistant to P1 (39) was generated by NRGene (NRGene, Nes Ziona, Israel) with the v3.0 of DeNovoMagic™ (<https://www.nrgene.com/solutions/denovomagic/>). As the Colomba assembly was not ordered into chromosomes, we used the *Sen1* specific *k*-mers from CoSSA to select the Colomba scaffolds belonging to the *Sen1* haplotype. We mapped the *Sen1* specific *k*-mers to the full Colomba assembly with BWA aln allowing 0 mismatches (command line option: -n 0) (40). The number of *Sen1* specific *k*-mers mapped to each Colomba scaffold was determined using Qualimap v2.2 (41).

Annotation of candidate genes

Putative NLR genes were searched on the five *Sen1* haplotype scaffolds with NLR-Annotator (42) using the standard parameters. The genes present on these scaffolds were predicted using Augustus v3.3.1 (reference species = tomato) (43). To inquire about the completeness of NLR gene prediction by Augustus, we searched for the NLR motifs from (44) using the motif alignment and search tool (MAST) from the MEME suite (45). To inquire about the expression of the candidate genes, we trimmed the RNAseq data of Desiree, Kuras, AI124-03 (RNAi plant) and Deodara (negative control) using Trimmomatic version 0.32 (46) (settings LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:70) and mapped the trimmed reads to the fine-

mapped region using STAR with stringent settings (option -- outFilterMismatchNmax 0) (47). The alignment files and coverage plots were visualized in IGV v2.3.72 (48).

RNAseq based expression analysis of candidate genes

Reads unique for the different candidate genes were selected. To do so, we extracted with SAMtools view (49) the *Sen1* specific *k*-mers which mapped to each candidate gene with a mapping quality greater than 10 (option -q 10). The RNAseq reads from each sequenced sample containing the *Sen1* specific *k*-mers having a mapping quality > 10 were retrieved using the kfastqfilter script from (29). The number of extracted reads was counted and compared between the samples.

Gap filling using long reads

To fill the gap in the sequence of the candidate gene NLR_5, we used Nanopore reads from the variety Colomba. To identify the long-reads spanning the candidate gene, we selected the reads with a size > 10kb and mapped them to the Colomba phased assembly using minimap2 (50). The reads mapping to the fine-mapping interval with a mapping quality of 60 were retrieved using Samtools view. To maximize the chances that the selected long-reads were specific to the resistant haplotype, we mapped to them the *Sen1* specific *k*-mers allowing 0 mismatches (option -n 0 of bwa aln as described previously) and kept the reads with at least 31 *k*-mers mapped. The selected long-reads were assembled using Canu v1.8 (51).

KASP markers genotyping

For the mapping and fine-mapping of *Sen1*, KASP markers were designed using the reference genome and using the phased re-assembly of Colomba to define the alternative allele. The KASP markers were designed using the CoSSA output as described in (29). All the primers of the markers designed in this study are available in Supplementary file 2. The KASP assays were performed as described in (29).

RNAi plants construction

A general RNAi construct was designed to silence the TNLs from the chromosome 11 clusters. As a template to design the RNAi construct, we used the *Rpi-cap1* gene candidate CAP19H1gn16 which belongs to the same TNL cluster. Alignments were made with all the TNLs from clusters C49 and C51 (44) to check the presence of common sequences of at least 21bp in a row. The primers used to amplify the wratchr11en1 RNAi fragment were: wratchr11en1-1F1 caccCTGTGCTCGATGGAAGTAT

GATGTCT and wratchr11en1-1R1 TCTGCATCCCCTCAACATCATCC. The RNAi fragment of a size of 400 bp was amplified using Phusion polymerase (ThermoFisher Scientific F530L), cloned via pEntr/D-Topo kit (ThermoFisher Scientific K240020) to pHellsgate12 (51) using the Gateway™ LR Clonase™ II system and transformed to *Agrobacterium tumefaciens* strain AGL + VirG (52) with a standard electroporation procedure. The transformation to plants was performed in the variety Kuras as described in (53).

Candidate genes cloning and complementation assays

The predicted CDSs of the two best candidate genes were synthesized by GenScript (GenScript, Leiden, The Netherlands) with the standard Gene Synthesis procedure, with adding the Gateway att L1/L2 sites, with avoiding the restriction sites of *BpiI* and *BsaI* (without any amino acid changes) but without any codon optimization. The gene expression constructs were cloned into pK7WG2 using the Gateway™ LR Clonase™ II system and transformed to *A. tumefaciens* strain AGL + VirG (52) with a standard electroporation procedure. The correct sequence of the CDSs was confirmed using Sanger sequencing.

Complementation assays with *AvrSen1* were performed in *N. benthamiana* and in *S. tuberosum* as described previously. For each experiment, *Avr8/R8* was used as a positive control and *AvrSen1* and the candidate genes alone were used as negative controls. Each candidate gene was then co-infiltrated with *AvrSen1*, separately and also combined. For the complementation assays in potato, we selected three genotypes that did not contain *Sen1*: one genotype from the AxD population (12-352-19) and two genotypes from the KxA population (12-380-24 and 12-380-36). For each complementation experiment, two plants of each genotype, three leaves per plant and two different ODs were tested (0.6 and 0.3). For the complementation assays in *N. benthamiana*, three plants per experiment, three leaves per plant and two different ODs were tested (1.2 and 0.6). As a positive control for the *AvrSen1* construct, we agroinfiltrated the *Sen1* positive AxD descendant 12-352-30.

Results

CoSSA to identify *Sen1* flanking markers

As a first step towards fine mapping, we identified markers that flanked *Sen1*. To identify *Sen1* haplotype specific SNPs, we pursued a CoSSA strategy. Pools of genomic DNA were made to build the AxDK_Sen1_RB and AxDK_SB bulks composed of 24

individuals holding the minor allele of PotVar0066530 and being resistant to P1, and 10 individuals without the minor allele of PotVar0066530 and susceptible to P1, respectively (Supplementary file 1). The two bulks and the three parents were sequenced (Supplementary file 3). *K*-mer tables were produced for each sample (Supplementary Figure 1). The difference between AxDK_Sen1_RB and AxDK_SB *k*-mers was made resulting in 718,151,348 *k*-mers (Nt) among which 197,803,897 were unique (Nu). *K*-mers with a frequency between 9x and 27x corresponding to genomic regions present in one copy were kept (Nt = 137,353,036, Nu = 12,031,271) and divided in several subsets in function of their parental origin (Supplementary file 3) and mapped to the potato reference genome (Figure 1).

A high peak of the *Sen1* specific *k*-mer (= AxDK_Sen1_RB bulk specific *k*-mers inherited by Desiree and Kuras) was observed on the north arm of chromosome 11, where *Sen1* has been mapped before (Figure 1, b). This peak was located on the 16 first bins of chromosome 11 and we could observe that it was composed of a first high peak located between 300 kb and 4.6 Mb and a second flat peak between 6,8 Mb and 15.7 Mb (Supplementary Figure 2). Both peaks contained 711,873 *k*-mers, which is 5.9% of the total. A new marker at the beginning of the first peak of *Sen1* specific *k*-mers was designed at 384,068 bp (AxDK_ch11_384068) (Supplementary Figure 2).

On the chromosome 0 containing the unanchored scaffolds of the reference genome, a peak of 48,567 *Sen1* specific *k*-mers was observed between 20 Mb and 21 Mb. This bin contains the superscaffold DMB734 which was recently anchored to the north arm of chromosome 11 (29). Another broad peak composed of 1,302,280 *Sen1* specific *k*-mers was observed on chromosome 4 between 25 Mb and 46 Mb. Regarding the AxDK_Sen1_RB bulk specific *k*-mers from the other subsets (Figure 1, c to h), we observed an interesting peak located between 47 Mb and 52 Mb on the south arm of chromosome 5, composed of 93,831 *k*-mers inherited from Aventura.

Screening for recombinants, recombinants phenotyping and *Sen1* fine-mapping using DM

To screen for recombinants, we used DM_11_384068 and the flanking marker PotVar0106272 identified in (28). With these two KASP markers, we screened 1,104 progeny clones from the cross between Aventura and Desiree and identified 56 recombinants (Figure 2, Supplementary file 4).

From the 56 identified recombinants, 48 could be phenotyped using the *S. endobioticum* effector *AvrSen1* which triggers HR in *Sen1* plants (Figure 3). The eight

Figure 1. Comparative Subsequence Sets Analysis (CoSSA) to map *Sen1*

The *Sen1*-bulk specific *k*-mers from seven different sets were mapped to the 12 potato chromosomes of DM v4.03 and to the unanchored scaffolds (chromosome 0). The average number of AxDK_Sen1_RB bulk specific *k*-mers mapped to each chromosome has been calculated and for each 1Mb bin of each chromosome, the number of mapped *k*-mers from each *k*-mer set has been transformed to a percentage of the chromosome average ($y_{\max} = 549\%$). **a** Number of NLR genes from DM according to (9) per 1Mb bin. *Sen1* bulk specific *k*-mers inherited **b** from Desiree and Kuras = *Sen1* specific *k*-mers, **c** from Aventra, **d** from Desiree, **e** from Kuras, **f** from Desiree and Aventra, **g** from Kuras and Aventra and **h** from the three parents.

Figure 2. Mapping, fine-mapping and candidate genes identification for *Sen1*

a Genetic map of the *Sen1* region in the combined AxD (n = 42) and AxK (n = 35) populations. **b** Genetic map of the *Sen1* region in the recombinants screening population (Aventra x Desiree, n = 1,104). **c** Genetic map of the *Sen1* region in a subset of 20 recombinants. **d** Physical map of two Colomba scaffolds spanning the *Sen1* region. **e** Physical map of the fine-mapped *Sen1* region on the Colomba scaffold 1148. Candidate genes present in this fine-mapped region were represented by an arrow. For each map, the name of the designed KASP markers was written on the right side of the map. For each genetic map, the number of observed recombinants was written on the left side. Remaining recombinants did not show competence for agroinfiltration as no HR was triggered when the positive control was infiltrated. For the 48 “agro-competent” recombinants a final qualitative score was given to each genotype (Supplementary file 1).

For the fine-mapping of *Sen1*, we used two CoSSA *Sen1* specific SNPs which were located at 1,308,927 bp and 1,666,351 bp of the DM genome. The corresponding markers were tested in the recombinant population and appeared to flank the *AvrSen1* response leaving a physical window of 357,424 bp in the reference genome (Figure 2). The flanking markers were located at the edge of the TNL cluster C76 (9). We showed previously that this *R* gene cluster region was poorly assembled (29) and that the unanchored superscaffold DMB734 belongs to this region.

Chromosome 11 TNL RNAi plants

To prove that *Sen1* is an NLR protein encoding gene from the TNL group, we used the TNL sequences from clusters C49 and C51 from DM (44) to design an RNAi construct that targets twelve of these TNLS. The RNAi construct was transformed to Kuras and the transgenic plants were agroinfiltrated with the effector *AvrSen1* to assess if the HR was reduced compared to Kuras. In a first agroinfiltration experiment (Exp1_2018_01_14, Supplementary Figure 3), six out of nine and eight out of nine Kuras RNAi plants showed a reduced HR to *AvrSen1* compared to the wild type Kuras for ODs of 0.3 and 0.1 respectively. The RNAi plants AI124-03 and AI124-10 showed the highest reduction of the HR compared to Kuras. For an OD of 0.3, the mean HR was 1.22, 0.04 and 0.1 for Kuras, AI124-03 and AI124-10 respectively and, for an OD of 0.1, their means were 0.75, 0.022 and 0. In a second experiment, significantly reduced HRs were observed for *AvrSen1* ODs of 0.03 and 0.01 (Exp2_2018_04_12, Supplementary Figure 4, Figure 4). Again, AI124-03 was the RNAi plant which showed the most reduced HR (HR means of 0.56 and 0.17 for ODs of 0.03 and 0.01 respectively) compared to the wild type (HR means of 1.5 and 1.09 for ODs of 0.03 and 0.01 respectively). Two plants with the lowest response were retested in a third experiment (Exp3_2018_04_19, Supplementary Figure 5), and indeed showed a significantly reduced HR for ODs of 0.1, 0.03 and 0.01 (Figure 4). The HR means for these ODs were respectively 1.95, 1.92 and 1.67 for Kuras and 1.56, 0.22 and 0.26

for AI124-03. In a last experiment (Exp4_ 2018_05_03, Supplementary Figure 6), lower ODs were used and significantly reduced HRs were observed for ODs of 0.02, 0.01, 0.005 and 0.0025. Overall, we observed significantly reduced HRs of several *Kuras* RNAi plants compared to *Kuras* wild type which validated our hypothesis that *Sen1* is indeed a TNL.

Figure 3. Phenotyping of the *Sen1* recombinants using *AvrSen1*

The *Sen1* recombinants were phenotyped by the agro-infiltration of *AvrSen1*. For each recombinant, 3 plants and 3 leaves / plant were agroinfiltrated with a positive control (here R3b/*AvrR3b*), a negative control (here R3b) and *AvrSen1*. In this figure are shown a recombinant not showing any HR when infiltrated with *AvrSen1* (AxD_557) and a recombinant for which an HR is triggered by the infiltration of *AvrSen1* (AxD_794).

Figure 4. Phenotyping of the RNAi plants using *AvrSen1*

a Hypersensitive response to two different ODs (0.03 and 0.01) of *AvrSen1* in Kuras and five different Kuras RNAi plants silenced with the wratchr11en1 RNAi construct (experiment of 12.04.2018). Kruskal-Wallis test was used with Dunn's test as post-hoc non-parametric multiple comparisons. RNAi plants which showed no differences in their HR response to *AvrSen1* were marked with the same letter. **b** Leaves of Kuras and AI124-03, the Kuras RNAi plant showing the most reduced response to *AvrSen1*, agroinfiltrated with a positive control (Avr3b/R3b), a negative control (R3b) and four different ODs of *AvrSen1*.

***Sen1* fine-mapping and candidate genes identification using a tetraploid phased assembly**

To continue the fine-mapping of *Sen1* and to confirm the right TNL candidate gene, we used a phased *de novo* assembly of the tetraploid variety Colomba (NRGene assembly). To identify scaffolds belonging to the *Sen1* haplotype, the *Sen1* specific *k*-mers identified with CoSSA were mapped to the full Colomba phased assembly allowing no mismatches. From the 5,532,061 *Sen1* specific *k*-mers, 1,971,891 (35,64%) were mapped. Among the scaffolds of a length of at least 100,000 nucleotides, we selected

the five scaffolds with the highest ratio mapped k -mers / length to which were mapped 42,1% (830,624 k -mers) of the mapping k -mers (Table 1).

Table 1. NLR genes identification on the *Sen1* haplotype scaffolds

		scaffold 2054	scaffold 1148	scaffold 4877-3	scaffold 2055	scaffold 1347
Length scaffold		3,505,740	1,339,855	372,875	325,214	276,239
Mapped k-mers		582,591	134,858	22,306	59,716	53,459
Mapped k-mers / length		0.17	0.10	0.06	0.18	0.19
TNL	Complete	9	7		3	3
	Complete (pseudogene)	1	3			1
	Partial	12	4		3	4
	Partial (pseudogene)	2	2			1
CNL	Complete	1	2	1		
Unknown	Complete	1				
	Partial	2				
Total		28	18	1	6	9

Candidate NLR genes were searched on the selected *Sen1* haplotype scaffolds using NLR-Annotator (42). In total 28, 18, 1, 6 and 9 complete or partial NLRs were identified on scaffolds 2054, 1148, 4877-3, 2055 and 1347, respectively (Table 1). Scaffolds 2054 and 1148 were the longest selected scaffolds with the highest number of TNLs. To anchor these scaffolds to the *Sen1* genetic map, we performed a BLAST of the previously described *Sen1* markers to the scaffolds (Figure 2). Scaffold 2054 was spanning the region just downstream *Sen1* where the previously identified flanking markers DM_11_1666351, PotVar0066530 and PotVar0106272 were located. Previously identified flanking markers DM_11_384068 and DM_11_1308927 were found on scaffold1148 (Figure 2). New markers were designed at the extremities of scaffold1148 to test if this scaffold was spanning the *Sen1* region. The newly designed markers were screened in the recombinants and anchored to the genetic map. The genetic order of the markers was co-linear with the sequence of the Colomba assembly. As markers Col_1148_1249152 and Col_1148_118477 were flanking the *Avrsen1* response, we concluded that scaffold 1148 was spanning the *Sen1* region and must therefore contain the *Sen1* gene. The NLR cluster on scaffold 1148 is composed of 16 TNL sequences. New markers were designed on that scaffold to pinpoint candidate *Sen1* genes. The *Sen1* genetic window was reduced to a 97,438 bp region between the markers Col_1148_338212 and Col_1148_240774. Both markers were one

recombination event away from *Sen1* (Figure 2; Supplementary file 1). This interval contained five putative TNL genes, two of which were annotated by NLR-Annotator as being complete, one partial and two complete but pseudogenes (Supplementary file 4).

Candidate TNL genes for *Sen1*

The NLR genes models were predicted using Augustus. The five predicted TNL genes ranged in size from 5,112 bp for NLR_7 to 13,097 bp for NLR_3 and they consisted of 4, 5, 5, 6 and 21 exons for NLR_4, NLR_5, NLR_6, NLR_7 and NLR_3 respectively (Supplementary file 4). Sequence alignments of the five complete amino acids sequences and subsequent phylogenetic analysis showed that they indeed grouped with other TNL genes (Supplementary Figure 7).

Candidate gene NLR_5 contained a gap in its third intron (Supplementary Figure 8). This gap could be due to the difficulty to assemble paralogous sequences or due to low complexity of the sequence. The size of the gap estimated by the assembler was 494 bp. To validate the size of the gap and to verify that it is not bigger and does not contain another exon or (an)other complete hidden gene(s), we used WGS Nanopore reads from Colomba. The Nanopore reads were mapped to the full Colomba phased assembly and 232 reads mapping to the *Sen1* interval (scaffold1148 240,774 bp to 338,212 bp) were extracted. To further confirm that the mapping Nanopore reads were indeed derived from the *Sen1* haplotype, we mapped the *Sen1* specific *k*-mers allowing 0 mismatches to the 232 fished reads and selected reads to which at least 31 *k*-mers mapped. Next, the 152 remaining reads were subjected to a *de novo* assembly. According to this assembly of the Nanopore reads (Canu assembly), the gap of NLR_5 was 898 bp long. We decided to fill the gap of the NRGene assembly with the Canu assembly and to re-do the gene prediction. The gene prediction was exactly the same as previously (Supplementary Figure 8), leaving the coding sequence unaltered after filling the gap. The gap was composed of very low complexity sequences which could not be assembled using short reads.

To validate the completeness of the five predicted NLR proteins, we searched for the NLR motifs (44) in the Augustus predicted protein sequences (Supplementary Figure 9). The candidate genes NLR_3, NLR_4 and NLR_5 contained all the motifs usually found in the TIR (motifs 18, 15 and 13), NB-ARC (motifs 1, 4, 5, 10, 3, 12) and LRR (motifs 9 and 11) domains of TNLs genes. In the predicted protein of NLR_6, the entire NB-ARC domain was missing. The motifs of the NB-ARC domain were however found

in the DNA sequence (Supplementary file 4) so we could not conclude if this gene is truly incomplete or if its incompleteness is due to a sequence error erroneously causing a frame shift. The NLR_7 protein missed the motif 18 of the TIR domain. This motif was present in the DNA sequence as well so again, we cannot conclude if this gene is incomplete or wrongly assembled.

Expression of the candidate genes

To determine if the five candidate genes were expressed in *Sen1* plants, we mapped RNAseq data from Desiree, Kuras and Deodara with a high stringency to the *Sen1* interval and counted the number of reads mapping to each candidate gene (Table 2). The read counts suggested that the five genes were expressed in Desiree and Kuras. Taking the gene size into account, the gene with the highest number of cDNA paired reads in the three genotypes is NLR_5. As expected, the number of cDNA reads from the P1 susceptible Deodara mapping to the candidate genes was very low.

Table 2. Expression analysis of the five candidate TNL genes

NLR	Size (bp)	Unique k-mers ^a	Desiree	Kuras		AI124-03		Deodara
			Total mapping reads ^b	Total mapping reads ^b	Gene specific reads ^c	Total mapping reads ^b	Gene specific reads ^c	Total mapping reads ^b
3	13,097	3,217	270	4067		4099		10
4	5,883	1,668	151	1301	37 (0.20)	1328	31 (0.16)	21
5	7,339	92	266	3386	98 (0.52)	3534	98 (0.52)	5
6	5,959	115	130	1015		1172		12
7	5,112	978	79	1042		1119		10

^a number of *Sen1* specific *k*-mers coming from unique sequences (mapping with a quality > 10) mapping to each of the candidate genes. ^b number of cDNA paired reads mapping to each candidate gene. ^c number of cDNA paired reads containing at least one *Sen1* specific *k*-mer from unique sequences (number of extracted reads * 10⁶ / total number of reads).

Coverage plots and alignments of the RNAseq reads were inspected manually to assess the evenness of the read coverage over the candidate NLR gene sequences. The mapping of reads to NLR_3, NLR_6 and NLR_7 contradicted the predicted gene model. Moreover, the depth is uneven. This showed that the reads were not derived from these candidates but were due to the mapping of paralogous sequences. The coverage of NLR_4 and NLR_5 confirmed the gene model prediction and the depth was more

even, except for some regions in the LRR domain (Figure 5, Supplementary Figure 10). So, only candidates NLR_4 and NLR_5 could be validated as expressed genes. We attempted to further prioritize *Sen1* candidate genes using the RNAi plant that had reduced expression of *Avrsen1* response. We tested if there was a difference in the RNAseq read count of the expressed candidate genes NLR_4 and 5 between Kuras and RNAi plant AI124-03. It was important to focus on sequences unique to each candidate gene and to get rid of reads from conserved regions in TNLs from paralogs or TNLs from other haplotypes. Therefore, we selected *Sen1* specific *k*-mers mapping uniquely to each of the five candidates and extracted the cDNA reads from each genotype containing at least one of these *k*-mers.

Figure 5. Coverage, gene model and NLR motifs of candidates NLR_4 and NLR_5

NLR_4 coverage plot of **a** Kuras RNAseq data (max depth = 355) and **b** Desiree RNAseq data (max depth = 25). **c** Gene model of NLR_4 predicted by Augustus. **d** NLR motifs found in NLR_4. NLR_5 coverage plot of **e** Kuras RNAseq data (max depth = 634) and **f** Desiree RNAseq data (max depth = 62). **g** Gene model of NLR_5 predicted by Augustus. **h** NLR motifs found in NLR_5. **i** Motif legend according to (44).

The extracted reads were counted and their relative abundance was calculated (ratio extracted reads * 10^6 / total number of reads) for each genotype (Table 2). For NLR_5, the ratio was exactly the same between the two genotypes (0.52). For the candidate gene NLR_4, the ratio of the unique cDNA reads was slightly smaller in AI124-03 (0.16) than in Kuras (0.20). As expected, none of the Deodara cDNA reads contained unique *Sen1* specific *k*-mers.

Candidate genes cloning and complementation assays

Because the RNAseq data did not confirm the gene model prediction of NLR_3, NLR_6 and NLR_7, and the RNAseq analysis of the RNAi plant did not conclusively point out either gene as a preferred candidate, we decided to synthesize the predicted CDSs of NLR_4 and NLR_5 (Supplementary file 5). The complementation assays were performed in *N. benthamiana* and in potato genotypes which did not contain *Sen1*. NLR_4 and NLR_5 were co-infiltrated individually with *AvrSen1* and we also co-infiltrated the two candidate genes together with *AvrSen1*. Surprisingly, neither of the *Sen1* candidate constructs provided complementation of *AvrSen1* recognition in any of the experiments performed in *N. benthamiana* and in potato (Supplementary Figure 11).

Discussion

Fine-mapping of *Sen1* using WGS approaches

R gene cloning can be a very long and labor-intensive process. Until recently, marker development for fine-mapping was tedious due to the limited availability of sequence polymorphisms. Subsequently, it involved the creation of bacterial artificial chromosome (BAC) libraries combined with several rounds of BAC landing and BAC sequencing. In this study, combined WGS approaches were applied, which could bypass these labor-intensive steps. First, CoSSA resulted in the identification of several peaks of AxDK_Sen1_RB bulk specific *k*-mers: two peaks of *k*-mers inherited from Desiree and Kuras on chromosome 4 and 11 and a peak of *k*-mers inherited from Aventura on chromosome 5. The peak from chromosome 4 is probably a false positive peak due to the small size of the susceptible bulk as the bi-modal distribution of P1 resistance in the two populations suggests the segregation of only one gene (28). The peak from chromosome 5 might be involved in resistance to other pathotypes, Aventura being resistant to pathotypes 2, 6 and 18. The CoSSA peak on chromosome 11 allowed us to identify haplotype specific polymorphisms that could be swiftly converted into markers. Using the reference genome, we mapped *Sen1* to the sub-telomeric TNL

cluster C76 between 1,308,927 bp and 1,666,351 bp on the long arm of chromosome 11. The potato reference genome, being distantly related to the cultivated potato *S. tuberosum* and being susceptible to the wart disease, could not be used to further fine-map *Sen1* inside the TNL cluster C76 and to identify candidate genes. Instead of creating a BAC clone library of a *Sen1* variety such as Desiree or Kuras, we decided to use a phased *de novo* assembly of the tetraploid potato variety Colomba which is resistant to pathotype 1 (D1) and holds *Sen1*. Using the *Sen1* specific *k*-mers from CoSSA, we could select the scaffolds which belonged to the *Sen1* haplotype. After the scaffolds selection, we aligned the primers of the previously designed *Sen1* specific KASP markers to identify scaffolds spanning the *Sen1* fine-mapped region. These two *in silico* scaffold identification and anchoring steps are the equivalent of a BAC landing and walking approach but were considerably faster. The design of only four KASP markers was necessary to show the positions of scaffolds 1148 and 2054 relative to *Sen1*. Using the CoSSA *Sen1* specific *k*-mers, we were able to design more KASP markers and the fine-mapping located *Sen1* in a 97 kb region containing five TNL genes.

Sen1* genotypes screening using *AvrSen1

The fine-mapping was sped up by the availability of *AvrSen1* (35). Without its availability, the recombinants phenotyping should have been performed using one of the potato wart disease assays. To get reliable phenotypes, the Spieckermann (54) and Glynne Lemmerzähl (55,56) assays require the availability of a sufficient number of tubers which can take up to two years of tubers multiplication. An alternative to these tuber assays could be the inoculation of aboveground parts of the plant (57) which does not require the multiplication of tubers but requires a high number of repeats because of the possibility of escapes. Another advantage of using *AvrSen1* is the use of the unambiguous direct gene-for-gene recognition of a specific effector protein by a specific host resistance protein. Results are not impacted by the presence of other species than the host and pathogen of the studied pathosystem. Moreover, results can differ in function of the *S. endobioticum* isolate used, the isolates classification in pathotypes being based only on their virulence pattern and not on their effectors content. The use of *Agrobacterium*-mediated transient transformation (agroinfiltration), unambiguously allowed us to assess if *Sen1* was present in the recombinants in a very short amount of time and a low number of repeats. A disadvantage however is that we have now cloned the *Avrsen1* response candidate

genes, which are most likely *Sen1*, but it remains to be shown if either of our candidates can indeed provide wart resistance in potato plants.

Fine-mapping of *Sen1* to a 97 kb interval

The first haplotype specific markers used in this study to build the resistant and susceptible bulks in the AxD and KxA populations were derived from a GWAS study performed in a panel of 330 potato cultivars (28). In this GWAS, *Sen1* was mapped 10 cM upstream PotVar0066530 (coordinate on DM v4.03: chr11 2,706,497 bp) which is in agreement with our mapping results. No marker flanking *Sen1* on the telomeric side could be identified due to the use of the SolSTW 20K SNP array (58). With the new markers screened in the same populations in our study, we observed that there was an overestimation of the number of recombinants between PotVar0066530 and *Sen1*. We reduced the number of recombinants from seven to three. One of the four individuals wrongly considered as a recombinant had resistance scores just at the edge of the threshold between resistant and susceptible clones. The three remaining individuals turned out to be false negatives as they do not hold *Sen1* but were scored as being resistant with the Glynne-Lemmerzahn assays (Supplementary File 1). The first mapping of *Sen1* (19) located the gene distal to the markers NI25 and NI27 which is also in agreement with our mapping results. In the diploid F1 population in which *Sen2* was identified, however, *Sen1* was mapped to a 2.05 Mb interval between 1,636,806 bp and 3,686,135 bp. These mapping results are in disagreement with our fine-mapping (27). Indeed, our DM-based flanking marker DM_11_1308927 was only one recombination event away from *Sen1* whereas DM_11_1666351 was nine recombination events away, suggesting *Sen1* was closer to 1.3 Mb than 1.67 Mb. Moreover, the BLAST of the DM-based flanking markers to the Colomba *Sen1* scaffolds showed that DM_11_1666351 was located on scaffold 2054 and DM_11_1308927 on scaffold 1148 at 344,270 bp so 6,058 bp from Col_1148_338212 and 103,496 bp from Col_1148_240774, the two closest flanking markers to *Sen1*. These discrepancies between the two mapping studies can be due to the different genetic background of the F1 population used by (27). It would be interesting to test the resistant parental clones of this population with *AvrSen1* to verify that it is the same gene segregating in both populations.

***Sen1* is a TNL gene from the C76 cluster**

Agroinfiltrations in RNAi plants transformed with an RNAi construct specific to the chromosome 11 TNL clusters C76 and C77 caused a reduction of the HR reactions triggered by *AvrSen1* as compared to the untransformed plant. These experiments

confirmed our hypothesis that *Sen1* is an NLR gene from the TNL group and the fine-mapping located *Sen1* in the cluster C76. The *Sen1* haplotype scaffold 1148 from Colomba was spanning this cluster and contained 16 potential TNL genes (Table 1, Supplementary file 4) whereas the C76 cluster from DM v4.03 contains only seven potential TNL genes (9). The number of *Sen1* candidate genes was reduced by the design of new markers, leaving five candidate TNLs. From those five TNLs, three were discarded due to their gene model and their RNAseq data coverage. Twenty-one exons were predicted for NLR_3 which makes it more likely a pseudogene as the three potato TNLs cloned so far possess four (*Gro1.4* and *Rysto*) to six (*Y-1*) exons (10–12). The RNAseq data of NLR_6 and NLR_7 did not support the predicted gene model which might be indicative for misassemblies or for being pseudogenes, the mapped reads coming from paralogous sequences.

Complementation assays with the two remaining candidates, however, did identify neither of both candidates as being *Sen1*. The reason could be the cloning system that we used. 35S driven expression of the CDSs of the candidates did not result in a functional AvrSen1 receptor, implying an essential role for the non-coding sequences of the gene. Non coding sequences like promoters and introns may affect mRNA levels and (alternative) splicing. Alternative splicing has previously been shown to be important for the function of several TNL genes such as for instance *N* (18), *L6* (15), *Y-1* (10), *BS4* (59) or *RPS4* (60). We did not find evidence for alternative splicing of the two candidate *Sen1* genes in our RNAseq data. This might be due to insufficient sequencing depth or due to the activation of alternative splicing upon infection/effector recognition. It has been shown for instance that the *Arabidopsis thaliana* TNL gene *RPS4* underwent expression changes and alternative splicing in response to the *Pseudomonas syringae* effector *avrRps4* (60). To identify the putative alternative transcripts of NLR_4 and NLR_5, we could extract and sequence RNA from leaves before and after agroinfiltration by *AvrSen1*. Another reason for the lack of complementation could be an incorrect prediction of the gene models. For instance, the visual inspection of the RNAseq data coverage plot of NLR_4 might suggest the existence of a fifth exon at the end of the gene. If alternative splicing is necessary for *Sen1* function or if the predicted gene models are incorrect, the cloning of each of the complete candidate gene regions will be required for positive complementation assays.

So far, only three potato TNL genes have been cloned: the *Y-1* gene involved in PVY resistance, located in the same cluster as *Sen1* (10), *Gro1-4*, involved in *G. rostochiensis* resistance, located on chromosome 7 (11) and *Rysto*, conferring extreme

resistance to PVY and located on chromosome 12 (12). The *Sen1* candidate genes showed a smaller phylogenetic distance with the *BS4* gene located on chromosome 5 of tomato (59) than with *Y-1*, *Gro1-4* or *Rysto* (Supplementary Figure 7). The *Sen3* gene was mapped to the same TNL cluster between 1,259,552 bp and 1,519,485 bp and could therefore be an NLR gene from the TNL group as well (29). The cloning of these two genes would bring valuable knowledge as well for the identification of other potato wart disease *R* genes but also for biotechnical applications such as the gene editing of their susceptible alleles to improve susceptible varieties.

The fine-mapping of *Sen1* and *Sen3* suggests that TNL genes play an important role in resistance to *Synchytrium endobioticum*. Up to now, no clear correlation has been observed between the NLR genes group and the recognition specificity of a certain type of pathogen. For instance, TNL genes have been shown to be involved in resistance to a wide range of pathogens such as bacteria, fungi, oomycetes, nematodes and viruses. This observation is coherent with the common observation that it is the LRR domain which is implicated in the specificity of effector recognition and not the TIR and CC N-terminal domains, which are involved in the activation of the defense signaling pathways (7). In flax, it has been shown however that the TIR domain of the flax rust resistance gene *L* can also play a role in specificity determination (61,62). Several alleles from the *L* locus having different resistance specificities have been cloned and sequenced. It was observed that natural or artificial *L* alleles with the same LRR region but different TIR and NBS regions could give different resistance specificities. Interestingly, up to now, all the cloned NLR genes giving resistance to flax rust are from the TNL group (63). Flax rust is caused by *Melampsora lini*, an obligate biotrophic fungus which is not culturable (64). This similarity with *S. endobioticum* gives rise to the hypothesis that TNL genes might be more efficient to give resistance to obligate biotrophs. TNL and CNL genes differ by their signaling pathways which involve different genes such as EDS1 (65) and NRG1 (66–69) for TNLs and NDR1 for CNLs (65). The efficiency of TNLs for obligate biotrophs might be explained by a quicker recruitment of defense responses with the TNL signaling pathway.

De novo assembly of *R* gene clusters

The potato reference genome genotype DM1-3 516 R44 (8) is a doubled monoploid from the *S. tuberosum* group Phureja which is quite distant from the commercial potato varieties *S. tuberosum* group Tuberosum. It has been shown that monoploid and doubled monoploid genomes of *S. tuberosum* group Phureja encompass an extensive copy number variation (CNV) impacting apparently a large part of the genome (70).

Almost a third of the potato genes showed at least partial deletions or duplications. More specifically, four of the ten genomic regions the most highly enriched for CNVs were *R* gene clusters which are known to evolve quickly through duplication or recombination (71). It can be expected that these structural changes observed in monoploids are even more abundant in tetraploid material. A study of the diversity of the wheat pangenome showed that only 64% of the genes were shared between the 19 wheat cultivars, the rest showing presence-absence variation (72). It was observed that the variable genome was enriched with genes involved in defense response. The same was observed in rice: the dispensable genome was enriched as well with genes involved in the defense to biotic stresses, including NLR genes (73). Because of this extreme variation of *R* genes between genotypes, relying on only one or a few reference genomes for genetic studies is insufficient to pursue the fine-mapping and the identification of candidate genes within *R* gene clusters. *De novo* assemblies of the *R* gene clusters of resistant varieties is therefore needed.

R gene clusters are however very difficult to assemble due to the high homology between paralogues and the presence of repeats. The use of short reads alone to assemble *R* gene rich regions is vain as they can lead to very scattered assemblies (29), chimeric genes (74), and will be inefficient in solving repeats rich regions. An alternative to the use of short reads is the use of long reads to assemble the targeted *R* gene clusters. Combining PacBio reads with RenSeq (SMRT RenSeq), (74) were capable of correcting the chimeric genes assembly of the MiSeq data and their final assembly allowed them to clone a new late blight resistance gene. Recently, the same SMRT RenSeq approach was successfully applied for the cloning of the TNL *Ry_{sto}* gene conferring extreme resistance to PVY (12). The phased *de novo* assembly we used in our study did not involve the long reads technology but combined the use of DNA libraries of various insert sizes and the use of the 10X technology for the phasing. The quality of the Colomba phased assembly in the *Sen1* region was sufficient to efficiently design haplotype specific markers inside this complex TNL cluster. The quality of the TNL genes assembly was however uneven. Indeed, for the five candidate TNLs inside the *Sen1* interval, we observed that the gene model was supported by the RNAseq data only for the two candidates NLR_4 and NLR_5 which could be a sign of misassemblies in the three other candidates. Moreover, we observed that the use of short reads combined with 10X data was inefficient in assembling low complexity regions and led to a gap in the candidate gene NLR_5. In this case, the use of Nanopore reads from Colomba could solve the gap composed of low complexity sequences.

Despite these errors, the phased *de novo* assembly was of a sufficient quality to synthesize the CDSs of the two *Sen1* candidate genes NLR_4 and NLR_5. One should not only consider the quality of a *de novo* assembly of *R* gene clusters but also the costs. In our case, the phased *de novo* assembly can be used for many different traits and purposes as the full genome has been assembled. However, the costs and time investment are substantial. If one's aim is the fine-mapping and cloning of *R* gene(s), a complexity reduction approach such as RenSeq combined with the long-read technology could be preferred.

Acknowledgements

We acknowledge all the partners involved in this public-private partnership (Averis seeds BV, Böhm-Nordkartoffel Agrarproduktion GmbH & Co. OHG, Danespo, HLB BV, HZPC Holland BV, C Meijer BV, SaKa Pflanzenzucht GmbH & Co. KG and Teagasc) for providing invaluable potato material, phenotypic data, KASP marker analysis, and scientific discussions. Richard Finkers and Linda Kodde are thanked for making the Colomba assembly and Nanopore data available.

Funding

This work was performed in a project called "An integrated genomics and effectormics impulse for potato wart resistance management and breeding" which was financially supported by the Dutch Topsector Horticulture & Starting Materials. Within the Topsector, private industry, knowledge institutes and the government are working together on innovations for sustainable production of safe and healthy food and the development of a healthy green environment. CP was supported by a Walsh fellowship provided by Teagasc.

Authors contribution

CP, HJvE and JHV conceived and designed the experiments. CP, GVA, MB, JR and MM performed the experiments and analyses. CP analyzed all the results with the help of JR, MM and JHV. CP wrote the manuscript. JHV and RGFV helped to write, improve and review the latest version of this manuscript. All authors read and approved the final manuscript.

References

1. Przetakiewicz J. The Viability of Winter Sporangia of *Synchytrium endobioticum* (Schilb.) Perc. from Poland. *Am J Potato Res.* december 2015;92(6):704–8.
2. Hampson MC. Control of potato wart disease through the application of chemical soil treatments: a historical review of early studies (1909-1928). *EPPO Bull.* march 1988;18(1):153–61.
3. Hampson M. History, biology, and control of potato wart disease in Canada. *Can J Plant Pathol.* december 1993;15(4):223–44.
4. EPPO. *Synchytrium endobioticum*: Diagnostic protocols. *EPPO Bull.* august 2004;34(2):213–8.
5. EPPO. PM 7/28 (2) *Synchytrium endobioticum*. *EPPO Bull.* december 2017;47(3):420–40.
6. Przetakiewicz J. First Report of New Pathotype 39(P1) of *Synchytrium endobioticum* Causing Potato Wart Disease in Poland. *Plant Dis.* 2015;99:285.
7. Dodds PN, Rathjen JP. Plant immunity: towards an integrated view of plant–pathogen interactions. *Nat Rev Genet.* august 2010;11(8):539–48.
8. Potato Genome Sequencing Consortium, Pan S, Cheng S, Zhang B, Mu D, Ni P, e.a. Genome sequence and analysis of the tuber crop potato. *Nature.* 10 july 2011;475(7355):189–95.
9. Jupe F, Witek K, Verweij W, Śliwka J, Pritchard L, Etherington GJ, e.a. Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. *Plant J.* november 2013;76(3):530–44.
10. Vidal S, Cabrera H, Andersson RA, Fredriksson A, Valkonen JPT. Potato Gene *Y-1* is an *N* Gene Homolog That Confers Cell Death Upon Infection with *potato virus Y*. *Mol Plant Microbe Interact.* july 2002;15(7):717–27.
11. Paal J, Henselewski H, Muth J, Meksem K, Menéndez CM, Salamini F, e.a. Molecular cloning of the potato *Gro1-4* gene conferring resistance to pathotype Ro1 of the root cyst nematode *Globodera rostochiensis*, based on a candidate gene approach. *Plant J.* april 2004;38(2):285–97.
12. Grech-Baran M, Witek K, Szajko K, Witek AI, Morgiewicz K, Wasilewicz-Flis I, e.a. Extreme resistance to *Potato virus Y* in potato carrying the *Ry_{sto}* gene is mediated by a TIR - NLR immune receptor. *Plant Biotechnol J* [Internet]. 4 september 2019 [cited 2 october 2019]; Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/pbi.13230>
13. Hinsch M, Staskawicz B. Identification of a new Arabidopsis disease resistance locus, *RPS4*, and cloning of the corresponding avirulence gene, *avrRps4*, from *Pseudomonas syringae* pv pisi. *Mol Plant Microbe Interact.* 1996;9(1):55–61.
14. Menz I, Straube J, Linde M, Debener T. The TNL gene *Rdr1* confers broad-spectrum resistance to *Diplocarpon rosae*: *Rdr1* -mediated disease resistance in roses. *Mol Plant Pathol.* may 2018;19(5):1104–13.
15. Lawrence G, Finnegan E, Ayliffe M, Ellis J. The *L6* Gene for Flax Rust Resistance is Related to the Arabidopsis Bacterial Resistance Gene *RPSP* and the Tobacco Viral Resistance Gene *N*. *The Plant Cell.* 1995;7:1195–206.
16. Claverie M, Dirlwanger E, Bosselut N, Van Ghelder C, Voisin R, Kleinhentz M, e.a. The *Ma* Gene for Complete-Spectrum Resistance to *Meloidogyne* Species in *Prunus* Is a TNL with a Huge Repeated C-Terminal Post-LRR Region. *Plant Physiol.* june 2011;156(2):779–92.
17. van der Biezen EA, Freddie CT, Kahn K, ParkerP JE, Jones JDG. Arabidopsis *RPP4* is a member of the RPP5 multigene family of TIR-NB-LRR genes and confers downy mildew resistance through multiple signalling components. *Plant J.* february 2002;29(4):439–51.
18. Dinesh-Kumar SP, Tham W-H, Baker BJ. Structure-function analysis of the tobacco mosaic virus resistance gene *N*. *Proc Natl Acad Sci.* 19 december 2000;97(26):14789–94.

19. Hehl R, Faurie E, Hesselbach J, Salamini F, Whitham S, Baker B, e.a. TMV resistance gene *N* homologues are linked to *Synchytrium endobioticum* resistance in potato. *Theor Appl Genet.* 1999;98(3):379–386.
20. Hämäläinen JH, Watanabe KN, Valkonen JPT, Arihara A, Plaisted RL, Pehu E, e.a. Mapping and marker-assisted selection for a gene for extreme resistance to potato virus Y: *Theor Appl Genet.* february 1997;94(2):192–7.
21. Verzaux E, Budding D, de Vetten N, Niks RE, Vleeshouwers VGAA, van der Vossen EAG, e.a. High Resolution Mapping of a Novel Late Blight Resistance Gene *Rpi-avl1*, from the Wild Bolivian Species *Solanum avilesii*. *Am J Potato Res.* december 2011;88(6):511–9.
22. Verzaux E, van Arkel G, Vleeshouwers VGAA, van der Vossen EAG, Niks RE, Jacobsen E, e.a. High-Resolution Mapping of Two Broad-Spectrum Late Blight Resistance Genes from Two Wild Species of the *Solanum circaefolium* Group. *Potato Res.* june 2012;55(2):109–23.
23. van Eck HJ, Vos PG, Valkonen JPT, Uitdewilligen JGAML, Lensing H, de Vetten N, e.a. Graphical genotyping as a method to map *Ny (o,n)sto* and *Gpa5* using a reference panel of tetraploid potato cultivars. *Theor Appl Genet.* march 2017;130(3):515–28.
24. Ballvora A, Flath K, Lübeck J, Strahwald J, Tacke E, Hofferbert H-R, e.a. Multiple alleles for resistance and susceptibility modulate the defense response in the interaction of tetraploid potato (*Solanum tuberosum*) with *Synchytrium endobioticum* pathotypes 1, 2, 6 and 18. *Theor Appl Genet.* december 2011;123(8):1281–92.
25. Groth J, Song Y, Kellermann A, Schwarzfischer A. Molecular characterisation of resistance against potato wart races 1, 2, 6 and 18 in a tetraploid population of potato (*Solanum tuberosum* subsp. *tuberosum*). *J Appl Genet.* may 2013;54(2):169–78.
26. Obidiegwu JE, Sanetomo R, Flath K, Tacke E, Hofferbert H-R, Hofmann A, e.a. Genomic architecture of potato resistance to *Synchytrium endobioticum* disentangled using SSR markers and the 8.3k SolCAP SNP genotyping array. *BMC Genet* [Internet]. december 2015 [cited 3 augustus 2017];16(1). Available from: <http://www.biomedcentral.com/1471-2156/16/38>
27. Plich J, Przetakiewicz J, Śliwka J, Flis B, Wasilewicz-Flis I, Zimnoch-Guzowska E. Novel gene *Sen2* conferring broad-spectrum resistance to *Synchytrium endobioticum* mapped to potato chromosome XI. *Theor Appl Genet.* november 2018;131(11):2321–31.
28. Prodhomme C, Vos PG, Paulo MJ, Tammes JE, Visser RGF, Vossen JH, e.a. Distribution of P1(D1) wart disease resistance in potato germplasm and GWAS identification of haplotype specific SNP markers. *Theor Appl Genet.* 2020 <https://doi.org/10.1007/s00122-020-03559-3>;
29. Prodhomme C, Esselink D, Borm T, Visser RGF, van Eck HJ, Vossen JH. 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* [Internet]. december 2019 [cited 12 june 2019];15(1). Available from: <https://plantmethods.biomedcentral.com/articles/10.1186/s13007-019-0445-5>
30. Fulton TM, Chunwongse J, Tanksley SD. Microprep protocol for extraction of DNA from tomato and other herbaceous plants. *Plant Mol Biol Report.* 1995;13(3):207–209.
31. Bernatzky R, Tanksley SD. Genetics of actin-related sequences in tomato. *Theor Appl Genet.* juni 1986;72(3):314–21.
32. Kaplinski L, Lepamets M, Remm M. GenomeTester4: a toolkit for performing basic set operations - union, intersection and complement on *k*-mer lists. *GigaScience* [Internet]. december 2015 [cited 25 july 2017];4(1). Available from: <https://academic.oup.com/gigascience/article-lookup/doi/10.1186/s13742-015-0097-y>
33. Sharma SK, Bolser D, de Boer J, Sønderkær M, Amorós W, Carboni MF, e.a. Construction of reference chromosome-scale pseudomolecules for potato: integrating the potato genome with genetic and physical maps. *G3 Genes Genomes Genet.* 2013;3(11):2031–2047.
34. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 15 march 2010;26(6):841–2.

35. van de Vossen BTLH, Prodhomme C, van Arkel G, van Gent-Pelzer MPE, Bergervoet M, Brankovics B, e.a. The *Synchytrium endobioticum* AvrSen1 Triggers a Hypersensitive Response in *Sen1* Potatoes While Natural Variants Evade Detection. *Mol Plant Microbe Interact.* 17 september 2019;MPMI-05-19-0138.
36. Vossen JH, van Arkel G, Bergervoet M, Jo K-R, Jacobsen E, Visser RGF. The *Solanum demissum* R8 late blight resistance gene is an *Sw-5* homologue that has been deployed worldwide in late blight resistant varieties. *Theor Appl Genet.* september 2016;129(9):1785–96.
37. Li G, Huang S, Guo X, Li Y, Yang Y, Guo Z, e.a. Cloning and Characterization of *R3b* ; Members of the *R3* Superfamily of Late Blight Resistance Genes Show Sequence and Functional Divergence. *Mol Plant Microbe Interact.* october 2011;24(10):1132–42.
38. R core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria [Internet]. 2018. Available from: <https://www.R-project.org/>
39. NVWA. Resistente aardappelrassen [Internet]. 2018 [cited 5 august 2019]. Available from: <https://www.nvwa.nl/onderwerpen/plantenziekten-en-plagen/wratziekte/resistente-aardappelrassen>
40. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 15 july 2009;25(14):1754–60.
41. Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics.* 1 october 2015;btv566.
42. Steuernagel B, Witek K, Krattinger SG, Ramirez-Gonzalez RH, Schoonbeek H, Yu G, e.a. Physical and transcriptional organisation of the bread wheat intracellular immune receptor repertoire. *bioRxiv* [Internet]. 5 june 2018 [cited 23 july 2019]; Available from: <http://biorxiv.org/lookup/doi/10.1101/339424>
43. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics.* 1 march 2008;24(5):637–44.
44. Jupe F, Pritchard L, Etherington GJ, MacKenzie K, Cock PJ, Wright F, e.a. Identification and localisation of the NB-LRR gene family within the potato genome. *BMC Genomics.* 2012;13(1):75.
45. Bailey TL, Gribskov M. Combining evidence using *p*-values: application to sequence homology searches. *Bioinformatics.* 1 february 1998;14(1):48–54.
46. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 1 august 2014;30(15):2114–20.
47. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, e.a. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* january 2013;29(1):15–21.
48. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander E, Getz G, e.a. Integrative genomics viewer. *Nature.* 2011;29(1):3.
49. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, e.a. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 15 august 2009;25(16):2078–9.
50. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Birol I, redacteur. Bioinformatics.* 15 september 2018;34(18):3094–100.
51. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive *k* -mer weighting and repeat separation. *Genome Res.* may 2017;27(5):722–36.
52. van der Fits L, Deakin EA, Hoge HC, Memelink J. The ternary transformation system: constitutive virG on a compatible plasmid dramatically increases Agrobacterium-mediated plant transformation. *Plant Molecular Biology.* 2000;43:495–502.

53. Zhu S, Vossen JH, Bergervoet M, Nijenhuis M, Kodde L, Kessel GJT, e.a. An updated conventional- and a novel GM potato late blight *R* gene differential set for virulence monitoring of *Phytophthora infestans*. Euphytica. march 2015;202(2):219–34.
54. Spieckermann A, Kothoff P. Testing potatoes for wart resistance. Deutsche Landwirtschaftliche Presse 51:114–115. 1924;
55. Glynne MD. Infection experiments with wart disease of potatoes. *Synchytrium endobioticum* (Schilb.) Perc. Ann Appl Biol. 1925;12(1):34–60.
56. Lemmerz J. A new simplified infection procedure for testing potato cultivars for wart resistance. Zücher. 1930;2(288):97.
57. van de Vossen BTLH, van Gent-Pelzer MPE, Boerma M, van der Gouw LP, van der Lee TAJ, Vossen JH. An Alternative Bioassay for *Synchytrium endobioticum* Demonstrates the Expression of Potato Wart Resistance in Aboveground Plant Parts. Phytopathology. juni 2019;109(6):1043–52.
58. Vos PG, Uitdewilligen JGAML, Voorrips RE, Visser RGF, van Eck HJ. Development and analysis of a 20K SNP array for potato (*Solanum tuberosum*): an insight into the breeding history. Theor Appl Genet. december 2015;128(12):2387–401.
59. Schornack S, Ballvora A, Gürlebeck D, Peart J, Ganai M, Baker B, e.a. The tomato resistance protein Bs4 is a predicted non-nuclear TIR-NB-LRR protein that mediates defense responses to severely truncated derivatives of AvrBs4 and overexpressed AvrBs3. Plant J. januari 2004;37(1):46–60.
60. Zhang X-C, Gassmann W. Alternative Splicing and mRNA Levels of the Disease Resistance Gene *RPS4* Are Induced during Defense Responses. Plant Physiol. december 2007;145(4):1577–87.
61. Ellis JG, Lawrence GJ, Luck JE, Dodds PN. Identification of Regions in Alleles of the Flax Rust Resistance Gene *L* That Determine Differences in Gene-for-Gene Specificity. Plant Cell. 1999;11:495–506.
62. Luck JE, Lawrence GJ, Dodds PN, Shepherd KW, Ellis JG. Regions outside of the Leucine-Rich Repeats of Flax Rust Resistance Proteins Play a Role in Specificity Determination. Plant Cell. 2000;12:1367–77.
63. Cullis CA. Disease Resistance Genes in Flax. In: Cullis CA, redacteur. Genetics and Genomics of Linum [Internet]. Cham: Springer International Publishing; 2019 [cited 8 september 2019]. p. 215–25. Available from: http://link.springer.com/10.1007/978-3-030-23964-0_13
64. Lawrence GJ, Dodds PN, Ellis JG. TECHNICAL ADVANCE: Transformation of the flax rust fungus, *Melampsora lini*: selection via silencing of an avirulence gene: Transformation of flax rust. Plant J. 13 november 2009;61(2):364–9.
65. Aarts N, Metz M, Holub E, Staskawicz BJ, Daniels MJ, Parker JE. Different requirements for *EDS1* and *NDR1* by disease resistance genes define at least two *R* gene-mediated signaling pathways in Arabidopsis. Proc Natl Acad Sci. 18 august 1998;95(17):10306–11.
66. Castel B, Ngou P, Cevik V, Redkar A, Kim D, Yang Y, e.a. Diverse NLR immune receptors activate defence via the *RPW 8*-*NLR NRG 1*. New Phytol. april 2019;222(2):966–80.
67. Peart JR, Mestre P, Lu R, Malcuit I, Baulcombe DC. NRG1, a CC-NB-LRR Protein, together with N, a TIR-NB-LRR Protein, Mediates Resistance against Tobacco Mosaic Virus. Curr Biol. may 2005;15(10):968–73.
68. Qi T, Seong K, Thomazella DPT, Kim JR, Pham J, Seo E, e.a. *NRG1* functions downstream of *EDS1* to regulate TIR-NLR-mediated plant immunity in *Nicotiana benthamiana*. Proc Natl Acad Sci. 13 november 2018;115(46):E10979–87.
69. Wu Z, Li M, Dong OX, Xia S, Liang W, Bao Y, e.a. Differential regulation of TNL-mediated immune signaling by redundant helper CNLs. New Phytol. april 2019;222(2):938–53.

70. Hardigan MA, Crisovan E, Hamilton JP, Kim J, Laimbeer P, Leisner CP, e.a. Genome Reduction Uncovers a Large Dispensable Genome and Adaptive Role for Copy Number Variation in Asexually Propagated *Solanum tuberosum*. *Plant Cell*. february 2016;28(2):388–405.
71. Bergelson J. Evolutionary Dynamics of Plant R-Genes. *Science*. 22 juni 2001;292(5525):2281–5.
72. Montenegro JD, Golicz AA, Bayer PE, Hurgobin B, Lee H, Chan C-KK, e.a. The pangenome of hexaploid bread wheat. *Plant J*. june 2017;90(5):1007–13.
73. Yao W, Li G, Zhao H, Wang G, Lian X, Xie W. Exploring the rice dispensable genome using a metagenome-like assembly strategy. *Genome Biol*. december 2015;16(1):187.
74. Witek K, Jupe F, Witek AI, Baker D, Clark MD, Jones JDG. Accelerated cloning of a potato late blight-resistance gene using RenSeq and SMRT sequencing. *Nat Biotechnol*. 25 april 2016;34(6):656–60.

Supplementary figures

Supplementary figure 1. *K*-mer frequency and distribution in parents and bulks

The volume of *k*-mers (unique *k*-mers x depth) for each *k*-mer frequency from 2 to 100 for the five different sequenced samples (green: *Sen1*-bulk, red: *S*-bulk, blue: *Desiree*, yellow: *Kuras*, black: *Aventra*). For the five samples, the peak of *k*-mers from simplex regions is visible.

AxDK_Sen1_RB bulk specific k-mers inherited from Desiree and Kuras (= Sen1 specific k-mers) mapping to the 17.8 first Mb of the north arm of chromosome 11. The position of the markers, designed using DM as a reference, are indicated.

Supplementary figure 3. AvrSen1 agroinfiltrations in RNAi plants. Experiment performed on January 14th 2018

Boxplots of the HR scores of each RNAi plant tested are presented for each construct (positive control, negative control and *AvrSen1*) and OD tested. Kruskal-Wallis tests were used with Dunn's tests as post-hoc non-parametric multiple comparisons. RNAi plants which showed no differences in their HR response to *AvrSen1* were marked with the same letter. If no letters are shown, the genotype effect was not significant.

Supplementary figure 5. AvrSen1 agroinfiltrations in RNAi plants. Experiment performed on April 19th 2018

Boxplots of the HR scores of each RNAi plant tested are presented for each construct (positive control, negative control and *AvrSen1*) and OD tested. Kruskal-Wallis tests were used with Dunn's tests as post-hoc non-parametric multiple comparisons. RNAi plants which showed no differences in their HR response to *AvrSen1* were marked with the same letter. If no letters are shown, the genotype effect was not significant.

Supplementary figure 7. Maximum likelihood tree of the Sen1 candidate proteins and various R proteins from Solanaceae

The tree is based on the proteins full amino acid sequences. CED4 from *Caenorhabditis elegans* was used as outgroup. The nodes number represent their bootstrap support (% out of 100 bootstraps). The clades represented on the right side of the tree are according to (9). Hero: CAD29729, Rpi-blb2: AAZ95005, Sw5-b: AAG31014, R8: ANJ02805, R1: Q8W1E0, Prf: AAC49408, Rpi-amr3: AMY98955, BS2: AAF09256, Gpa2: AAF04603, Rx2: CAB56299, Rx: CAB50786, Rpi-abpt: ACU65455, R2: ACU65456, Tm2²: AAQ10736, Rpi-vnt1.1: ACJ66594, R3b: AEC47890, R3a: AAW48299, Rpi-bt1: ACI16480, Rpi-blb1: Q7XBQ9, Rpi-sto1: ACI25288, Gro1-4: AAP44390, N: Q40392, Y-1: (10), BS4: XP_010320695, CED4: C35D10.9, Rysto: QEL52751.1.

Supplementary figure 8. Coverage, gene model and gap filling of candidate NLR_5

RNA-seq coverage of the candidate gene NLR_5 in **a** Kuras (max depth = 634), **b** Desiree (max depth = 62) and **c** Deodara (max depth = 30). **d** Gene model of NLR_5 predicted by Augustus. The 494 bp gap in intron 3 is pinpointed with dashed lines. **e** NLR domains and motifs from (44) identified with MAST. **f** Nanopore reads from Colomba, which belonged to the fine-mapped resistant haplotype, mapped to the **g** Canu assembly of the gap of NLR_5 which turned-out to be 898 bp long. When comparing the NRGene and the Canu assemblies, we observed the presence of 48 small deletions and 11 SNPs in the Canu assembly which were probably due to the high error rate of the Nanopore reads. **h** Augustus gene model of NLR_5 with the gap filled with the Canu assembly. **i** Motif legend according to (44).

Supplementary figure 9. Motifs from (44) found by MAST in the protein sequences of the five Sen1 candidates

Motifs 13, 15 and 18 are specific to the TIR domain of TNLS. Motifs 1, 2, 3, 4, 5, 6, 10, 12 are specific to the NB-ARC domain. Motifs 9, 11 and 19 are specific to the LRR domain. Motifs 7 and 8 are linkers. Motifs 16 and 17 are pre-NB motifs specific to CNLS. Motifs 14 and 20 are found in monocotyledons.

Supplementary figure 10. RNAseq data coverage of the five *Sen1* candidates

For each candidate gene, the coverage for A1124-03 (A), Kuras (K), Desiree (Ds) and Deodara (Do) is given. For each track, the depth range is also given on the right hand side of the Y axis. Reads mapped to the candidates NLR_3, NLR_6 and NLR_7 although we suspect these genes to be not expressed. The depth along these candidates is uneven with peaks of high depth which suggest the mapping of conserved paralogous sequences. The mapping of reads to these candidates contracted the predicted gene model.

Supplementary figure 11. Complementation assays in *N. benthamiana* and in potato

Complementation assays performed in *N. benthamiana* (OD = 0.6) and in the potato genotype 12-380-36 (OD = 0.3). Only the positive control showed a Hypersensitive Response, no complementation was observed for any of the candidate with AvrSen1.

Supplementary files list

Supplementary files can be accessed at <https://doi.org/10.17026/dans-2zf-pnd3>

> Supplementary file 1 (Chapter4_S_file_1.xlsx)

Genotypic and phenotypic data of the AxDK populations (sheet 1) and of the AxD recombinants (sheet 2). Sheet 1: Progeny clones selected for the *Sen1*-bulk and the S-bulk are indicated in two extra columns as well as recombinants and false negatives. Sheet 2: Genotypic data of all the tested markers and AvrSen1 response data in the recombinants population. For each marker, the reference used for the design is given. For each recombinant, the result of the agroinfiltration of *AvrSen1* is given (HR = HR reaction triggered by *AvrSen1*, n = no HR).

> Supplementary file 2 (Chapter4_S_file_2.xlsx)

Primer sequences of all the KASP markers used in this study. For each marker is given the reference genome used to design the marker (DM v4.03 or the Colomba phased re-assembly), the chromosome or the scaffold, the coordinate, the forward primer of the resistant and the susceptible alleles and the reverse primer. The markers are ordered according to Figure 2.

> Supplementary file 3 (Chapter4_S_file_3.xlsx)

CoSSA results in the AxDK population. For each sequenced sample, the number of sequenced reads is given, the sequencing depth is calculated, the number of total *k*-mers (Nt) and the number of unique *k*-mers (Nu) is given for the sequenced samples and several subsets obtained after basic set operations for different depth cut-offs.

> Supplementary file 4 (Chapter4_S_file_4.xlsx)

Potential NLR genes identified by NLR-Annotator on scaffold 1148 from the Colomba phased *de novo* assembly. For each candidate NLR gene, the completeness, the start and the end position, the strain and the motifs from (44) are given by NLR-Annotator. According to the motifs identified, we assigned each candidate to the CNL or the TNL group. For each candidate, the corresponding gene name, start and end position, number of exons and gene size predicted by Augustus are given as well. The KASP markers Col_1148_338212, Col_1148_240774 and Col_1148_118477 are represented by grey rows.

> Supplementary file 5 (Chapter4_S_file_5.pdf)

Genomic sequences, predicted cDNA sequences and protein sequences of the two synthesized *Sen1* candidates NLR_4 and NLR_5.

Chapter 5

A Hitchhiker's Guide to the Potato Wart Disease Resistance Galaxy

Charlotte Prodhomme, Gert van Arkel, Jarosław Plich,
Jasper E. Tammes, Johan Rijk, Herman J. van Eck,
Richard G. F. Visser, Jack H. Vossen

Abstract

Key message

Two novel major effect loci (*Sen4* and *Sen5*) and several minor effect QTLs for potato wart disease resistance have been mapped. The importance of minor effect loci to bring full resistance to wart disease was investigated. Using the newly identified and known wart disease resistances, a panel of potato breeding germplasm and *Solanum* wild species was screened. This provided a state-of-the-art inventory of complementary wart disease resistance sources.

Abstract

Potato wart disease, caused by the obligate biotrophic soil-born fungus *S. endobioticum*, is the most important quarantine disease of potato. Because of its huge impact on yield, the lack of chemical control and the formation of resting spores with long viability, breeding for resistant varieties combined with strict quarantine measures are the only way to efficiently and durably manage the disease. In this study, we set out to make an inventory of the different resistance sources. Using a Genome Wide Association Study (GWAS) in the potato breeding genepool, we identified *Sen4*, associated with pathotypes 2, 6 and 18 resistance. Associated SNPs mapped to the south arm of chromosome 12 and were validated to be linked to resistance in one full-sib population. Also, a Bulk Segregant Analysis (BSA) combined with a Comparative Subsequence Sets Analysis (CoSSA) resulted in the identification of *Sen5*, associated with pathotypes 2, 6 and 18 resistance, on the south arm of chromosome 5. In addition to these two major effect loci, the GWAS and CoSSA allowed the identification of several Quantitative Traits Loci (QTLs) necessary to bring full resistance to certain pathotypes. Panels of varieties and *Solanum* accessions were screened for the presence of *Sen1*, *Sen2*, *Sen3*, *Sen4* and *Sen5*. Combined with pedigree analysis, we could trace back some of these genes to the ancestral resistance donors. This analysis revealed complementary resistance sources and allows to eliminate redundancy in wart resistance breeding programs.

Keywords

Synchytrium endobioticum, haplotype specific SNPs, pedigree analysis, Comparative Subsequence Sets Analysis (CoSSA), Bulk Segregant Analysis (BSA), Genome-Wide Association Study (GWAS).

Introduction

Every year, food crop production suffers huge yield losses due to pests and pathogen attacks. For potato (*Solanum tuberosum*), the yield losses have been estimated between 8% to 21% depending on the region of production (1). The potato wart disease, caused by *Synchytrium endobioticum*, can cause yield losses up to 100%. *S. endobioticum* is an obligate biotrophic soil-born fungus from the Chytridiomycota phylum which causes the formation of galls on potato tubers. After the sexual phase of its life cycle, *S. endobioticum* produces winter spores (2) which can remain viable in the soil for decades (3). Because of this longevity and the lack of efficient chemical control (4), *S. endobioticum* has a quarantine status worldwide. The fungus originates from the Andean region of South America where it co-evolved with potato species and was introduced in North America and Europe at the end of the nineteenth century (5). At that time, only one variant (pathotype 1 (D1)) of the pathogen existed and breeders were successful in breeding for resistant varieties at the beginning of the twentieth century. However, new pathotypes emerged in the 1940's (6) and resistance to the pathotype 1 was not effective against them. Today, more than 40 pathotypes of *S. endobioticum* have been reported and the most frequent occurring in Europe are pathotypes 1(D1), 2(G1), 6(O1) and 18(T1). There is a strong need of identifying loci bringing resistance against these higher pathotypes as the only efficient and durable way to control the potato wart disease is to apply strict quarantine measures and cultivate resistant potato varieties.

Several potato wart disease resistance genes have already been identified. The first gene to be identified was *Sen1* (7), a TIR-NBS-LRR (TNL) gene from the chromosome 11 cluster C76 (8), which brings resistance to pathotype 1 through the recognition of the pathogen effector AvrSen1 (9). *Sen1* has been intensively used in breeding (10) and its resistance is not effective against higher pathotypes. The *Sen2* gene, bringing resistance to a wide range of pathotypes, has been identified recently in a complex diploid species hybrid. Its presence in commercial varieties is unknown as the markers flanking *Sen2* have not yet been screened in a panel of resistant and susceptible varieties. *Sen3*, bringing resistance to pathotypes 2, 6 and 18, has been mapped to the same TNL cluster as *Sen1* (11). The flanking markers of *Sen3* have been screened in a wider panel of varieties (11,12) and *Sen3* was found to be the main cause of resistance in Polish and German varieties. These three *Sen* genes are dominant genes giving a qualitative type of resistance. Several Quantitative Trait Loci (QTLs) have also been mapped in tetraploid populations. In the BNA1 and SaKa1 tetraploid populations,

Sen1 was identified as well as a QTL on chromosome 1 giving resistance to pathotypes 2, 6 and 18 and a QTL on chromosome 9 giving resistance to pathotype 18 (13). In the full-sib population obtained from the cross between Saturna and Panda, *Sen1* was identified, as well as several QTLs on chromosomes 1, 2, 6, 7, 8, 10 and 11 (14). In the BNA2 population, several QTLs were also identified on chromosomes 1, 3, 4, 6, 10 and 12 (15).

Wild *Solanum* species are a reservoir of resistance genes for many diseases, which is a valuable tool for breeders. Indeed, numerous resistance genes were discovered in wild *Solanum* species, such as *S. demissum* and *S. bulbocastanum* for late blight resistance or *S. tuberosum* ssp. *andigena* for viruses resistance (16), and were introgressed in the potato breeding genepool. *Sen1* was probably present very early in the ancestors of cultivated potato as its frequency is high in potato breeding germplasm (10). Indeed, (17) screened 52 landrace genotypes from *S. phureja*, *S. stenotomum*, *S. tuberosum* ssp. *andigenum* and *S. tuberosum* ssp. *tuberosum* with the *Sen1* linked marker NI25 (7) but they did not observe any correlation between wart resistance to pathotype 1 and taxonomy, ploidy level or geographic origin. This confirms our hypothesis that *Sen1* was already present in the ancestors of cultivated potato. Resistance to the higher pathotypes must, however, come from later introgressions, maybe as linkage drag during the introgression of *R* genes for other diseases such as nematodes or viruses. Several wild *Solanum* species have been reported to be potential sources for the higher pathotypes resistance. This is the case of *S. tuberosum* ssp. *andigena* (6,18,19), *S. acaule* (6,19) or *S. demissum* (6,18) which were reported several times as resistance sources. Recently, the *Sen3* gene could be traced back to the variety Ora, the pedigree of which contains *S. edinense* origins and a cultivar from Chiloe (11). The *Sen2* gene was identified in a complex hybrid with *S. acaule*, *S. chacoense*, *S. demissum*, *S. gourlayi*, *S. microdontum*, *S. phureja*, *S. tuberosum*, *S. verrucosum* and *S. yungasense* in its pedigree (20). Knowing which species have been used to introgress resistance to wart disease in breeding material would be useful information for the elimination of redundancy in breeding germplasm and the identification of new resistances.

Therefore, we set out to identify potential additional *Sen* genes in potato breeding germplasm. First, we performed a Genome-Wide Association Study for resistance to pathotypes 2, 6 and 18 in the potato breeding genepool. The significantly associated markers were screened in three independent full-sib populations, resulting in the identification of *Sen4* on the south arm of chromosome 12 and *Rse-XIc-VTN62.33.3*, a

QTL on the north arm of chromosome 11, both bringing resistance to pathotypes 2, 6 and 18. Bulk Segregant Analyses (BSA) (21,22) combined with Comparative Subsequence Sets Analyses (CoSSA) (11) were used to find additional markers for *Sen4*. Also, CoSSA was pursued to map the resistance of Aventura, *Sen5*, on the south arm of chromosome 5, involved in pathotypes 2, 6 and 18 resistance. Furthermore, CoSSA was used to design haplotype specific markers for the resistances segregating in the SaKa1 population (13). Finally, we identified a minor effect locus necessary to bring full resistance to pathotype 18 in combination with the *Sen3* gene (11). Markers flanking *Sen1*, *Sen2*, *Sen3*, *Sen4* and *Sen5* were screened in a panel of resistant and susceptible varieties and in a panel of wild *Solanum* accessions. The varieties and accessions sequenced in the study of (23) were screened using CoSSA to identify the presence of *Sen1*, *Sen3*, *Sen4* and *Sen5*. The distribution over the breeding germplasm and the origin of the different *Sen* genes are inquired and discussed.

Materials and Methods

Plant materials

For the GWAS, we used the genotypic dataset of 569 varieties produced with the 20K SolSTW SNP array and kept genotypes with a SNP call rate greater than 75%. We gathered phenotypic data for pathotypes 2, 6 and 18 (hereafter referred as P2, P6 and P18) from public and private sources (Supplementary file 1). The GWAS panels for P2, P6 and P18 were composed of 117, 138 and 53 genotypes respectively (Supplementary file 1).

We used the five segregating populations AxV, AxD, KxA, SaKa1 and KxL to validate the GWAS results and to map novel and previously identified loci involved in potato wart disease resistance (Supplementary file 2). The AxV population was composed of 100 descendants of a cross between Axion (resistant to P1, P2, P6 and P18) and VR808 (resistant to P1). The AxD population resulted from a cross between Aventura (susceptible to P1, resistant to P2, P6 and P18) and Desiree (resistant to P1), and was composed of 42 descendants (10). The KxA population resulted from a cross between Kuras (resistant to P1) and Aventura, and was composed of 35 descendants (10). The SaKa1 population (13) was composed of 124 descendants from the cross between Andante (resistant to P1, P2, P6 and P18) and Alegria (resistant to P1). The KxL population consisted of 328 descendants from a cross between Kuba (resistant to P1, P2, P6 and P18) and Ludmilla (resistant to P1) (11).

To make an inventory of the potato wart disease resistance present in the potato breeding germplasm and in wild *Solanum* species, we collected a panel of 118 potato breeding clones and varieties, called hereafter the variety panel. The variety panel was composed of resistant and susceptible varieties (mainly tetraploid), old and recent, bred in at least 11 different countries. Also, a panel of 118 *Solanum* accessions, some from which wart resistance phenotypes were available through CGN (Centrum voor Genetische Bronnen Nederland), was compiled and called hereafter the Solanum panel. The Solanum panel was composed of diploid and polyploid accessions from the Solanum section petota and contained 38 different species. They originated from various regions ranging from Central and South America (Supplementary file 3).

Phenotyping

The wart resistance phenotype data used in the GWAS were gathered from diverse sources such as National Lists, various websites, booklets from breeding companies and scientific papers (Supplementary file 1). The resistance scales used in the different sources were different and for the purpose of GWAS, the quantitative scores were all transformed to a 1 (highly susceptible) to 10 (highly resistant) scale. Qualitative scores (i.e. resistant or susceptible) were transformed into quantitative scores, as indicated in Supplementary file 1. For each genotype, we calculated a final resistance score corrected for the source (origin of the phenotypic data) effect using restricted maximum likelihood (REML) as follows: $Resistance = source + genotype$, the source being included as a random effect and the genotype as a fixed effect.

The phenotypic assays performed on the mapping populations in this study are summarized in Supplementary table 1. For the Spieckermann assays (24), each assessed tuber was given a quantitative score ranging from 10 (highly resistant, corresponding to the type 1 in Germany and the type - in the Netherlands) to 1 (highly susceptible, corresponding to the type 5 in Germany and to the type X in the Netherlands). A mean score was calculated for each genotype (Supplementary file 2). For the Glynne-Lemmerzähl assays (25,26), disease symptoms were rated from 1 (highly resistant, early defence necrosis) to 5 (highly susceptible) and mean scores were calculated between replicates. Pathotype 1 resistance scores for the populations AxD and KxA were previously described by (10). Phenotypic data for the KxL population was described by (11). Phenotypic data for the SaKa1 population was obtained by (13).

For some *Solanum* accessions from the Solanum panel, phenotypic data for pathotypes 1, 2, 6 and/or 8 could be retrieved from the CGN (Centrum voor Genetische Bronnen

Nederland) database ((27)). Five to ten tubers per accession were phenotyped with the Glynne-Lemmerzähl method between 1980 and 1994. A qualitative score was given to the phenotyped accessions: resistant (R), intermediate (I) or susceptible (S) (Supplementary file 3).

Genome Wide Association Study of pathotypes 2, 6 and 18 resistance

The association analysis was carried out using two different models. The first model was a naive model which did not include a correction for the panel structure. For the second model, we calculated the van Raden kinship (28) between the genotypes using a random subset of 1,000 markers and performed a Principal Coordinate Analysis on the kinship. The 30 first principal coordinates (PCOs) were included as a fixed effect in a mixed linear regression model. As potato resistance to the higher pathotypes is rare in potato germplasm, we reduced the minor allele frequency (MAF) threshold: for each panel, we kept markers present in at least three genotypes. Markers with more than 20% of missing data were removed from the dataset. In total, 12,279, 12,486 and 11,392 markers were used for the P2, P6 and P18 GWAS studies. The two GWAS models were fitted in GenStat version 18 (VSN International, 2015). For each dataset (P2, P6, P18), the genome wide significance threshold was calculated by the procedure QTHRESHOLD, using the method developed by Li and Ji, 2005, similarly as described in (10).

Bulked Segregant Analysis (BSA) in the mapping populations

To design haplotype specific markers flanking *Sen4* segregating in AxV, a resistant bulk (AxV_RB) and a susceptible bulk (AxV_SB) were built containing 18 and 15 descendants, respectively (Supplementary file 2). To identify the resistance locus present in Aventura (resistant to P2, P6 and P18), we used the bulks previously compiled and sequenced by (8). In the AxDK_Sen1_RB bulk, 15 out of 24 descendants showed resistance to other pathotypes than P1 (Supplementary file 2). The susceptible bulk AxDK_SB was composed of descendants susceptible to all pathotypes. In the SaKa1 population, we selected 16 descendants resistant to P1, P2, P6 and P18 to build the SaKa1_RB bulk and 16 descendants resistant to P1 but susceptible to P2, P6 and P18 to build the SaKa1_P1RB bulk (Supplementary file 2). To identify minor effect loci providing full resistance to P18 in the KxL population from (11), we compiled one new bulk in this population. The KxL_P2P6RB bulk was composed of 17 descendants which were fully resistant to P2 and P6 but weakly susceptible to P18 with the Glynne-Lemmerzähl phenotyping method, and contained the *Sen3* flanking markers.

Comparative Subsequence Sets Analysis (CoSSA) workflows

Several CoSSA workflows have been conducted to identify the resistance loci segregating in the different populations. For each population, the workflow used was the CoSSA with reference genome described in (11). To summarize: R-bulk specific k -mers were selected by performing the difference between the R-bulk and the S-bulk. The R-bulk specific k -mers were filtered based on their frequency in function of the R-bulk sequencing depth: the k -mers selected had a frequency of $\frac{R-bulk\ depth}{4} \pm 0.5 * \frac{R-bulk\ depth}{4}$. Because of the relatively low sequencing depth of the R-bulks from the AxV and Saka1 populations, we adapted the depth cut-off filter by increasing the lower and upper thresholds to improve the signal (Supplementary file 4). Next, the R-bulk specific k -mers were divided in function of their inheritance (from the R parent(s), referred to as resistance specific k -mers; from the S parent(s); from all parents; from none of the parents). Where required, the difference was made between the resistance specific k -mers and the k -mers present in susceptible varieties (called hereafter “minus S varieties”) to increase the haplotype specificity of the remaining k -mers. The resistance specific k -mers (minus S varieties) were then mapped to the potato reference genome DM v4.03 (30). The number of mapped k -mers per 1Mb bin for each chromosome were counted and plotted in Microsoft Excel. The details of each CoSSA workflow applied for each population are described in the Supplementary file 4.

CoSSA was also used to test the presence of *Sen1*, *Sen3*, *Sen4* and *Sen5* in the 67 varieties and accessions from (23). The intersection between each of the above-mentioned resistance gene specific k -mers (without the S varieties when applicable) was made with the k -mers from each of the 67 samples. The resulting k -mers were then mapped to the potato reference genome DM v4.03, k -mers per 1Mb bin were counted and the number of k -mers mapping to the *Sen1*, *Sen3*, *Sen4* and *Sen5* (fine)-mapped regions was compared with positive controls (varieties which hold the gene of interest).

DNA extraction and sequencing

Genomic DNA of the parents of the populations (except SaKa1), of the bulked progeny clones and of susceptible varieties was extracted from freshly harvested leaves according to (31) and purified using the Qiagen DNeasy Plant Mini Kit. The samples DNA concentration was measured using a Qubit Fluorometer (Invitrogen). For the different bulks, an equal amount of DNA from each progeny clone was pooled to obtain a final DNA quantity of 1 μ g. For all the samples, 1 μ g of (pooled) genomic DNA was

used for the library preparation and sequenced on an Illumina platform. Paired-end (PE) reads of 151 bp were produced (Hartwig Medical foundation, Amsterdam, The Netherlands). A summary of the sequencing depth obtained for all sequenced samples is given in Supplementary file 4.

Lyophilised leaves from the SaKa1 population were obtained from SaKa Pflanzenzucht GmbH & Co. KG. The DNA of the genotypes in the bulks was extracted as described previously. The DNA concentration and quality were verified with a Qubit Fluorometer (Invitrogen) and on ethidium bromide containing agarose gels. The DNA was partially degraded, probably due to storage of the lyophilised material at room temperature. Consequently, we compiled four sub-bulks in function of the level of degradation of the DNA as assessed on gel. The two resistant sub-bulks were named SaKa1_RB_small (containing samples with smaller DNA fragments), SaKa1_RB_big, and contained each eight individuals that were resistant to P1, P2, P6 and P18. SaKa1_P1RB_small and SaKa1_P1RB_big sub-bulks contained both eight individuals that were resistant to P1 and were positive for the *Sen1* markers, but that were susceptible to P2, P6 and P18. During sequencing library preparation, the two sub-bulks with smaller DNA fragments were less sheared than the two with big DNA fragments. The four sub-bulks were sequenced as described above. Appropriate sequencing quality and yield of the four sub-bulks was confirmed using FastQC (32) and the small and big fragments sub-bulks were merged *in silico* to form SaKa1_RB and SaKa1_P1RB bulks for subsequent CoSSA.

Genomic DNA of the different siblings, parents and the variety panel was extracted from freshly harvested leaves or from tubers according to (31). The DNA quality was assessed on ethidium bromide containing agarose gels. Genomic DNA from the *Solanum* panel was extracted by the Dr Van Haeringen Laboratorium (Wageningen, The Netherlands). The DNA concentration was measured using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific).

KASP and PCR markers genotyping

Kompetitive Allele Specific PCR (KASP) markers were designed as described in (11). Adjusted concentrations of 5-50 ng/μl of genomic DNA were used for the KASP assays. The KASP assays were performed in house or by C. Meijer BV (Rilland, The Netherlands) as described in (11).

A Cleaved Amplified Polymorphic Sequence (CAPS) marker designed by (20) was used to validate the presence of *Sen2* in the variety and *Solanum* panels. Adjusted concentrations of 5-10 ng/μl of genomic DNA were used for the *Sen2* marker assays.

PCR amplified a 1046 bp fragment which was visualised on 2% agarose gels. The 1046 bp fragment is then digested with a restriction enzyme. DG 97-264 from (20) was used as a positive control and its 1046 bp fragment is digested into 8 fragments.

We used χ^2 test to verify the goodness of fit of the segregation ratio of the tested KASP markers in the five populations with the expected ratio (defined in function of the dosage observed in the parents). We used Kruskal-Wallis test to validate the association of the tested markers with resistance. All the statistical tests were performed on R v3.2.3.

Phylogenetic analysis in the Solanum panel

Genotypic data from 222 AFLP markers were available for 108 of the 118 accessions from the Solanum panel (33). We used Mesquite v3.6 (34) to format the data and MrBayes v3.2.7 (35) to infer a Bayesian rooted tree. MrBayes was run with four chains, 10,000,000 generations, a sampling frequency of 10,000 and a temperature setting for the heated chains of 0.25. Six *S. etuberosum* accessions were added as an outgroup to the analysis in order to root the phylogenetic tree.

Results

GWAS of pathotypes 2, 6 and 18 resistance identifies a minor and a major effect QTL

Two GWAS models were used to find markers associated with P2, P6 and P18 resistance in panels of 117, 138 and 53 potato genotypes respectively: one naive model and a model corrected with a PCoA performed on 1,000 random SNPs. With the naive model, 32, 292 and 9 markers were significantly associated with P2, P6 and P18 respectively (Figure 1). The correlation between the 30 first PCo of the PCoA and P2, P6 and P18 resistance was of 0.72, 0.70 and 0.79, respectively. Remarkably, with the PCoA corrected model, none of the markers reached the significance thresholds of 4.567, 4.608 and 4.295 for P2, P6 and P18 respectively, suggesting that resistance is present in related genotypes. Pursuing with the results of the naive model, we found that 23 SNPs were significantly associated with both P2 and P6, ten of which were located on the north arm of chromosome 11 between 0.81 Mb and 4.35 Mb. Seven SNPs were associated with both P6 and P18 resistance, six of which were located on the south arm of chromosome 12 between 43.3 Mb and 50.4 Mb. From the 303 associated SNPs, we selected 42 SNPs having a positive effect on resistance and having the minor alleles present in simplex or duplex in the resistant parents and absent from the susceptible parents of the validation populations (AxV, AxD, KxA,). These 42 selected SNPs were

converted to KASP markers in order to validate their linkage to resistance in the three sibling populations. The distribution of resistance in these populations can be found in the Supplementary figures 1 to 3.

Figure 1. GWAS performed on pathotypes 2, 6 and 18 resistance

Manhattan plots of the GWAS performed on **a** P2, **b** P6 and **c** P18 resistance. The x axis represents the 12 potato chromosomes. Chromosome u represents the markers from unanchored scaffolds, chloroplastic and mitochondrial markers. The red line is the threshold of significance as calculated by the method of (29). Significant markers are highlighted in green. The markers from the *Sen4* haplotype and from *Rse-XIc-VTN62.33.3* associated with resistance in the AxV, KxA and AxV populations are indicated.

The three markers from chromosome 12 located between 43.3 Mb and 50.4 Mb (PotVar0031912, PotVar0036325 and PotVar0037666) co-segregated and were strongly associated with P2, P6 and P18 resistance in the AxV population (Supplementary table 2). This chromosome 12 haplotype showed a strong effect on resistance (Supplementary figure 4). Therefore, we decided to give it the name of *Sen4*, following the naming system of the dominant, major effect *S. endobioticum* resistance genes. The oldest variety in which we found the *Sen4* markers in the GWAS panels is Alcmaria (Supplementary file 1). Twelve co-segregating markers from the north arm of chromosome 11 were also significantly associated with resistance. We called this haplotype *Rse-XIc-VTN62.33.3*. We followed the naming system from (36) and added the name of the oldest clone in which the haplotype markers were found in order to distinguish different haplotypes with overlapping genomic locations. The minor alleles of the first four markers on the telomeric side (~0.81 Mb) of *Rse-XIc-VTN62.33.3* (PotVar0064519, PotVar0064625, PotVar0064771, PotVar0064787) were only associated with P2 resistance while the first seven markers from the centromeric side (between 2.04 Mb and 4.35 Mb; PotVar0066219, PotVar0066337, PotVar0067017, PotVar0067459, PotVar0110432, PotVar0106057 and PotVar0106019) were associated with P2 as well as P6 resistance. The last marker from *Rse-XIc-VTN62.33.3*, PotVar0059284 (coordinate 8,816,275), was present in duplex in Axion and was indeed associated with P2 but also with P18 resistance in the AxV population. The effect of *Sen4* was stronger than the effect of *Rse-XIc-VTN62.33.3* for P6 (mean scores of 9.5 and 8.5 for *Sen4* and *Rse-XIc-VTN62.33.3* respectively) and especially for P18 resistance (mean scores of 9.1 and 6.0; Supplementary figure 4). The highest level of resistance was achieved when both loci were present.

CoSSA to fine-map *Sen4*

To identify markers flanking *Sen4*, we built and sequenced a resistant bulk (AxV_RB) from 18 selected AxV descendants which were strongly resistant to P6 and P18, and a susceptible bulk (AxV_SB) composed of 15 descendants that were highly susceptible to P6 and P18 (Supplementary file 2). We performed CoSSA and identified 91,868,050 unique (Nu) AxV_RB specific *k*-mers from a total number (Nt) of 492,018,414 *k*-mers

that occurred at a frequency from 4 to 20 (Supplementary file 4). From these *k*-mers, 42% (Nu = 38,573,330; Nt = 195,954,466) were inherited from Axion. From these Axion resistance specific *k*-mers, we removed *k*-mers from the susceptible varieties Alegria, Desiree, Kuras and Ludmilla to select for haplotype specific *k*-mers. This last set was composed of 17,440,189 unique and 88,818,285 total *k*-mers. Firstly, a peak composed of 163,965 Axion resistance specific *k*-mers was observed on the north arm of chromosome 11 between 1 Mb and 3 Mb. The chromosome 11 peak included four of the 12 SNPs that were associated with P2 and P6 in the GWAS study (PotVar0066337, PotVar0067017, PotVar0106057 and PotVar0106019). The re-identification of *Rse-XIc-VTN62.33.3* confirmed the contribution of this QTL to resistance. Secondly, from these Axion resistance specific *k*-mers, 10% (Nu = 1,752,553) mapped to a broad peak ranging from 10 Mb till 54 Mb on chromosome 12 (Figure 2). Within the chromosome 12 peak, *k*-mers mapped to the three markers identified with the GWAS (PotVar0031912, PotVar0036325 and PotVar0037666). We sought additional SNPs in the chromosome 12 peak to design markers flanking *Sen4*. Three SNPs to which *Sen4* specific *k*-mers were matching and for which PotVar markers had already been designed ((37); PotVar0036489, PotVar0037687 and PotVar0037404) and two other SNPs (chr12_48501410 and chr12_51499015), were selected to design novel KASP markers. Among 100 AxV descendants, only three recombinants could be identified between the outermost selected SNPs. Recombinant AxV_13_33 was susceptible to P6 and P18 in the Spieckermann assays of 2016 (scores of 7.5 and 6.4 respectively). Recombinant AxV_13_84 was scored as resistant to P6 and P18 in the Spieckermann assays of 2015 and 2016 and did not hold *Rse-XIc-VTN62.33.3* (Supplementary file 2). Recombinant AxV_13_38 was resistant to P6 and P18 in the Spieckermann assays of 2015 and 2016 but it holds *Rse-XIc-VTN62.33.3* as well and therefore was not informative for *Sen4* mapping. Consequently, we could map *Sen4* in a 3 Mb region between 48.5 Mb and 51.5 Mb (between chr12_48501410 and chr12_51499015) (Figure 3).

CoSSA to identify *Sen5*

Also in the AxD population, *Rse-XIc-VTN62.33.3* was significantly associated with P6 and P18 as assessed by the Glynne-Lemmerzähl method and P18 as assessed by the Spieckermann method (Supplementary table 3). In the KxA population, *Rse-XIc-VTN62.33.3* was also significantly associated with P2 and P6 in the Glynne-Lemmerzähl method and P18 in the Spieckermann method. None of the *Rse-XIc-VTN62.33.3*

markers were associated with P1 resistance, ruling out the possibility that these markers are derived from the *Sen1* haplotype that resides in the same genomic region.

Several descendants in the AxD and KxA populations which did not hold *Rse-Xlc-VTN62.33.3* showed a strong resistance to P2, P6 and P18, suggesting that at least one other resistance locus was segregating. To identify the additional resistance(s) segregating in the AxD and KxA populations, we re-used the bulks that were used to fine-map *Sen1* (AxDK_Sen1_SB and AxDK_Sen1_RB) (8) as 15 out of 24 individuals within AxDK_Sen1_RB showed resistance to P2, P6 and/or P18 while all 10 individuals in AxDK_Sen1_SB were highly susceptible to all the tested pathotypes. To identify *k*-mers linked to P2, P6 and P18 resistance, we lowered the frequency threshold. As AxDK_Sen1_RB had a sequencing depth of $\sim 72\times$ and 15 from the 24 descendants composing the bulks were resistant to the higher pathotypes, and assuming that a

Figure 2. Comparative Subsequence Sets Analysis to fine-map *Sen4* and identify *Sen5*

a Number of NLR genes per bin of 1 Mb from the potato reference genome DM v4.03 according to (38). **b** CoSSA performed in the AxV population. The AxV_RB specific *k*-mers inherited from Axion minus the S varieties *k*-mers were mapped to the reference genome ($y_{\max} = 67,500$ *k*-mers). **c** CoSSA performed in the AxDK population. The AxDK_RB specific *k*-mers inherited from Aventura minus the S varieties *k*-mers were mapped to the reference genome ($y_{\max} = 58,000$ *k*-mers).

single simplex locus was causal for resistance, the resistant haplotype(s) from Aventura were expected to have a sequencing depth between 6x and 17x. CoSSA allowed the identification of 32,183,695 unique *k*-mers ($N_t = 267,071,227$) specific to the AxDK_Sen1_RB bulk (Supplementary file 4). From these *k*-mers, 18% ($N_u = 5,934,080$, $N_t = 48,978,404$) were inherited from the resistant parent Aventura. From these Aventura resistance specific *k*-mers, commonly occurring *k*-mers were removed by subtracting *k*-mers from the susceptible varieties Alegria, Ludmilla and VR808, resulting in a subset of 3,533,028 unique *k*-mers ($N_t = 28,794,943$). From these *k*-mers, 207,335 mapped to the south arm of chromosome 5 and produced a peak between 47 Mb and 52 Mb (Figure 2). To validate the association of the chromosome 5 locus with resistance, we identified eight PotVar SNPs from (37) to which Aventura resistance specific *k*-mers without S varieties *k*-mers mapped. We also designed two new markers flanking the *k*-mers peak. These ten markers co-segregated and were strongly associated with P2, P6 and P18 in AxD and in KxA (Supplementary tables 3 and 4). The effect of the chromosome 5 haplotype on resistance was stronger than the effect of *Rse-XIc-VTN62.33.3* in AxD for P2 (mean scores of 1.9 and 2.8 for the chromosome 5 locus and *Rse-XIc-VTN62.33.3*, respectively) with Glynne-Lemmerzähl and for P18 with Spieckermann (mean scores of 8.0 and 5.6, respectively; Supplementary figure 5). In KxA, the chromosome 5 haplotype provided a stronger effect on resistance for P6 (mean scores of 3.5 and 4.0 for the chromosome 5 locus and *Rse-XIc-VTN62.33.3*, respectively) and P18 (mean scores of 3.7 and 4.1, respectively) with Glynne-Lemmerzähl and P18 (mean scores of 8.1 and 4.9, respectively) with Spieckermann (Supplementary figure 6). The highest resistance was always achieved when both loci were present. Because Aventura chromosome 5 resistance is a strong effect locus, we decided to name it *Sen5*. Two recombinants in AxD and two recombinants in KxA could be identified in the *Sen5* haplotype. The two recombinants from AxD are not informative as they hold *Rse-XIc-VTN62.33.3*. The two KxA recombinants do not hold *Rse-XIc-VTN62.33.3* and were clearly susceptible to P6 and P18 in the Spieckermann assay. Therefore, we could map *Sen5* to a 1.96 Mb region

of the south arm of chromosome 5 between 49.2 Mb and 51.16 Mb (between PotVar0123123 and PotVar0034831) (Figure 3).

CoSSA in the SaKa1 population

In the SaKa1 population, resistance to P1 was skewed towards resistance whereas resistance to P2, P6 and P18 was skewed towards susceptibility (13). The segregation of P1 resistance was bimodal whereas resistance to the other pathotypes was quantitative. A PCA performed on the phenotypes showed that the descendants could be divided in three different groups: group 1 contained 37 descendants which were resistant to the four pathotypes, group 2 contained 47 descendants which were only resistant to P1 and group 3 contained 39 descendants susceptible to the four pathotypes (Supplementary figure 7). To identify the loci providing resistance to P2, P6 and P18, we built the SaKa1_RB bulk composed of 16 descendants resistant to the four pathotypes and the SaKa1_P1RB bulk composed of 16 descendants only resistant to P1 (Supplementary file 2). SaKa1_RB specific *k*-mers were identified (Nt = 68,252,446; Nu = 6,255,989; frequency between 9 to 40) and divided in subgroups in function of their parental origin (Supplementary file 4). The 2,626,263 (Nu) SaKa1_RB specific *k*-mers inherited from Andante and 2,222,171 (Nu) *k*-mers from Alegria were mapped to the potato reference genome (Supplementary figure 8A). A peak of 90,013 *k*-mers inherited from Andante was found on the south arm of chromosome 1 between 72 Mb and 80 Mb. A second peak of 62,861 *k*-mers inherited from Andante was found on the south arm of chromosome 9 between 55 Mb and 61.5 Mb. A third broad peak on chromosome 11 between 0 and 31 Mb was composed of 152,829 *k*-mers. A fourth peak of 54,300 *k*-mers inherited from Alegria was observed on the south arm of chromosome 12 between 57 Mb and 60 Mb (this latter interval was not overlapping with the *Sen4* locus). KASP markers were designed to validate the association of these peaks with resistance. The *Sen1* flanking markers (8) were also included in this screening.

The *Sen1* flanking markers were significantly associated with P1 resistance in the SaKa1 population (Supplementary table 5). The effect of *Sen1* on P1 resistance was very strong, with a mean score of 2.19 when *Sen1* was present and 3.56 when *Sen1* was absent (Supplementary figure 9). The markers on the south arm of chromosome 1 between 72.77 Mb and 79.69 Mb were also associated with P1 resistance albeit with a lower effect than the *Sen1* markers (mean score of 2.4 for *Sen1* alone, mean score of 3.45 for chr01_76425362 alone). The resistance to P1 was improved when both loci were present (mean score of 2.13; Supplementary figure 9). The chromosome 1 locus,

Figure 3. Genetic and physical maps of *Sen4* and *Sen5*

Genetic maps of *Sen4* (a) and *Sen5* (c) in the AxV and AxDK populations respectively. The number of recombinants found between the screened markers is given on the left side of the maps. Physical maps of *Sen4* (b) and *Sen5* (d) according to the potato reference genome DM v4.03. On the right side of the physical maps are shown the screened markers, the RLK genes (grey) according to (39), the NLR genes (grey) and clusters (black blocks) according to (38).

which we called *Rse-Ib-Andante*, was also associated with P2, P6 and P18 resistance (Supplementary table 5). The marker with the strongest effect on resistance was chr01_76425362 which segregated in a 1:4:1 (nulliplex:simplex:duplex) ratio. The resistance was stronger when the marker was present in duplex (Supplementary figure 9) which can be due either to a dosage effect or to the fact that chr01_76425362 is present on two different haplotypes, and one of the two haplotypes brings a stronger resistance than the other. The markers chr01_72774086, chr01_74148509, chr01_74162620, chr01_77750280, chr01_77801278, chr01_79026840 and chr01_79694600 were located on the same haplotype of chromosome 1 because they co-segregated (referred hereafter as *Rse-Ib-Andante-a*). Markers located under the 0-31 Mb peak of chromosome 11 were co-segregating and were also significantly associated with P2, P6 and P18 resistance. The effect of this locus, that we called *Rse-XId-Andante*, on resistance was weaker than *Rse-Ib-Andante* but the resistance was improved when both loci were present (Supplementary figure 9). The markers designed under the chromosome 9 peak were co-segregating and significantly associated with P18 resistance only. The effect of this locus, that we called *Rse-IXa-Andante*, on P18 resistance was weaker than *Rse-Ib-Andante* (duplex marker) but stronger than *Rse-XId-Andante*. The concomitant presence of *Rse-Ib-Andante*, *Rse-IXa-Andante* and *Rse-XId-Andante* brought a strong resistance to P18 (mean score of 2.37; Supplementary figure 9).

The *Rse-Ib-Andante-a* haplotype was associated with P1, P2, P6 and P18 resistance but not all the descendants which hold this haplotype were resistant to P2, P6 and P18. This observation suggests that this haplotype is dependent on other QTLs in the background.

In an attempt to identify the other locus/loci responsible for *Rse-Ib-Andante-a* haplotype to bring full resistance, we built two new bulks, SaKa1_chr01_RB and SaKa1_chr01_SB, composed respectively of 17 descendants holding the *Rse-Ib-Andante-a* haplotype and fully resistant to the four pathotypes and of 17 descendants holding the *Rse-Ib-Andante-a* haplotype but with a weak resistance to the four

pathotypes (Supplementary file 2). In a new CoSSA, the difference between the *k*-mers from SaKa1_chr01_RB and SaKa1_chr01_SB was made (Supplementary file 4). A high peak composed of 113,781 SaKa1_chr01_RB bulk specific *k*-mers inherited from Andante was observed on chromosome 1 between 67 Mb and 74 Mb, overlapping with the position of *Rse-Ib-Andante-a* (Supplementary figure 8B). Two markers were designed under this peak (chr01_70066624 and chr01_73527005). Both markers were present on the same haplotype (called *Rse-Ib-Andante-c*) which did not co-segregate with *Rse-Ib-Andante-a*, and chr01_73527005 was significantly associated with P2, P6 and P18 resistance (Supplementary table 5). This was in agreement with our hypothesis that marker chr01_76425362 might be present on two different haplotypes from *Rse-Ib-Andante*. When comparing the presence/absence of *Rse-Ib-Andante-a* and *Rse-Ib-Andante-c* markers, it seems that marker chr01_76425362 was indeed present on both haplotypes. In another CoSSA workflow, the difference between SaKa1_chr01_SB and SaKa1_chr01_RB was made to identify haplotypes which potentially contributed to susceptibility (Supplementary file 4). A high peak of 215,605 SaKa1_chr01_SB bulk specific *k*-mers inherited from Andante was observed again on the south arm of chromosome 1 between 74 Mb and 84 Mb. Two markers were designed under this peak (chr01_77329972 and chr01_79751572). Both were located on the same haplotype (called *Rse-Ib-Andante-b*) and showed a significant association with P1, P2 and P6 susceptibility. This is most likely due to a repulsion effect of *Rse-Ib-Andante-b* with the haplotypes *Rse-Ib-Andante-a* and *Rse-Ib-Andante-c*. These results reveal a complex architecture of wart disease resistance in the SaKa1 population which involves several haplotypes of the *Rse-Ib-Andante* locus. *Rse-Ib-Andante* has a major effect on resistance but because of its co-dominance, we decided not to include it in the *Sen* naming system for dominant major effect wart resistance genes.

CoSSA to identify minor effect loci in the KxL population

In the KxL population ($n = 328$) from (11), we performed a Principal Component Analysis (PCA) on the P2, P6 and P18 resistance scores assessed with the Glynne-Lemmerzähl method (Supplementary figure 10). The two first principal components (PC), explaining 97.8% of the variance, showed that the descendants of the population could be separated in three groups. Group 1 was composed of 67 descendants resistant to P2, P6 and P18. Group 2 contained 92 descendants resistant to P2 and P6 but slightly susceptible to P18. Group 3 contained 169 descendants susceptible to P2, P6 and P18. The proportion of group 1+2 and 3 reflected a 1:1 segregation ratio and *Sen3* co-

segregated with group 1 and 2 phenotypes (except for 5 false positives out of 328). The groups 1 and 2 represent each 25% of the KxL population suggesting that there is another locus segregating in a 1:1 ratio which is required by *Sen3* to bring full resistance to P18. To identify this locus, we built the KxL_P2P6RB which was composed of 17 individuals belonging to phenotypic group 2. We re-used the KxL_RB from (11) which consisted of 17 individuals of group 1. CoSSA identified 2,851,467 (Nu) KxL_RB specific *k*-mers with a frequency between 8 and 60 (Nt = 26,856,174) (Supplementary file 4). Four peaks of KxL_RB bulk specific *k*-mers were observed that were inherited from Kuba, and six peaks that were inherited from the susceptible parent Ludmilla (Supplementary figure 11). We designed KASP markers to validate eight of these peaks and screened the entire offspring with them (Supplementary table 6). Only the two markers located on the south arm of chromosome 8 were significantly associated with P18 resistance obtained with the Glynne-Lemmerzähl method. This chromosome 8 locus, that we called *Rse-VIIIb-Kuba*, significantly improved P18 resistance in combination with *Sen3*. Although the *Rse-VIIIb-Kuba* effect alone was very small compared to *Sen3* alone (Supplementary figure 12).

Inventory of the wart disease resistances present in the potato breeding gene pool

A panel of 118 potato varieties and clones was screened with the flanking markers of *Sen1* (chr11_1666351), *Sen2* (*Sen2_CAPS*), *Sen3* (chr11_1259552 and chr11_1772869), *Sen4* (chr12_48501410 and PotVar0037404) and *Sen5* (PotVar0034831, PotVar0035016). Also 23 varieties whose genomes were sequenced by (23) were screened for the *Sen1*, *Sen3*, *Sen4* and *Sen5* genes using CoSSA (Table 1; Supplementary file 5). *Sen1* was present in 59.6% (n = 84) of the tested varieties, 78 of which are resistant to P1, two have an intermediate level of resistance and for four the resistance is unknown. *Sen1* is not present in any of the varieties susceptible to P1. *Sen2* marker was absent from all the varieties except for the variety Bonza which showed a banding pattern which suggests it might be positive for *Sen2*. *Sen3* was found in 22 varieties, 17 of which are known to be resistant to P1 and at least one of the higher pathotypes. Thirteen of the *Sen3* varieties hold BRA9089 in their pedigree, for the rest, the pedigree is incomplete or unknown. *Sen4* was found in 13 varieties, ten of which are resistant to P2, twelve to P6, and nine to P18. Only Berber is susceptible to P2, P6 and P18. The oldest clone in which we identified *Sen4* is AM78-3704 and nine of the *Sen4* varieties hold AM78-3704 in their pedigree. In the GWAS genotypic data, the *Sen4* markers were found in Alcmaria, the grand-parent of AM78-3704 (Supplementary file 1). Among the twelve resistant varieties holding *Sen4* in our

variety panel, ten hold Alcmaria in their pedigree, making it a very likely source of resistance.

Table 1. Screening of the potato wart disease *Sen* genes associated markers in the variety panel

Variety	Registration		<i>Sen</i> genes ^a					Resistance to ^b					Clone in pedigree ^c			
	Year	Country	<i>Sen1</i>	<i>Sen2</i>	<i>Sen3</i>	<i>Sen4</i>	<i>Sen5</i>	P1	P2	P6	P8	P18	BRA9089	Alcmaria	VE71-105	Jubel
Bzura	1983	POL	1	0	1	0	0	R	R	R	R	R	y			
Cekin	2004	POL	1	0	1	0	0	R	R	R	R	R	?			
Defender		USA	1	0	1	0	0	R					y			
Eurotonda		BRD	1	0	1	0	0	R	R		R	R	?			
Gandawa	2004	POL	1	0	1	0	0	R	R	R	R	R	?			
Gawin	2010	POL	0	0	1	0	0	R	R	R		R	?			
Goldrush	1992	USA	0	0	1	0	0						y			
Ibis	1987	POL	0	0	1	0	0	R					y			
Ikar	1996	POL	0	0	1	0	0	R	R	R	R	R	y			
Karolin	1990	BRD	0	0	1	0	0	R	R	R	R	R	y			
Kuba	1999	POL	0	0	1	0	0	R	R	R	R	R	y			
Lady Sara	2013	HOL	1	0	1	0	0	R					y			
Megusta	2010	BRD	0	0	1	0	0	R	R	R		R	?			
Sonda	2002	POL	0	0	1	0	0	R	R				?			
Transit	2009	BRD	1	0	1	0	0	R	R	R		R	y			
Ulme	1991	BRD	0	0	1	0	0	R	R	R		R	y			
Zagloba	2007	POL	1	0	1	0	0	R	R	R		R	y			
Ora	1952	DDR	0	0	1	r	0	R	R	R	R	R	y	n		n
Temp	1965	SU	1	0	1	r	0						y	n		y
Adam	2005	POL	0	0	1	r	0	R	R	R	R	R	?	?		?
Constantina		BRD	1	0	1	0	0	R	R	R	R	R	?			
Rudawa	2002	POL	1	0	1	0	0	R		R			?			
Bonza		GER	1	1?	r	0	0	R	R	R			?			
Jennifer	2015	HOL	1	0	r	0	0	I	S	S		S	n			
MPI 44.1016_10		GER	0	0	r	0	0	R	S	S		S	n			
Amanda	1986	SWE	1	0	r	0	0	R	R	R		S	n			
Inca Dawn			0	0	r	0	0						?			
Inca Rosa			0	0	r	0	0						?			
Actaro	2013	HOL	1	0	0	1	0	R	R	R		R		y		y
AM78-3704	> 1969	HOL	1	0	0	1	0	R	R	R		R		y		y
Avano	2005	HOL	1	0	0	1	0	R	R	R		R		y		y
Avaya	2005	HOL	1	0	0	1	0	R	R	R		R		y		y
Averia	2004	HOL	1	0	0	1	0	R	R	R				y		y
Axion	2008	HOL	1	0	0	1	0	R	R	R		R		y		y
Berber	1984	HOL	1	0	0	1	0	R	S	S		S		y		y
Celandine	2013	HOL	0	0	0	1	0	S	S	R		I		y		y
Katinka	1998	HOL	1	0	0	1	0	R	R	R		R		y		y
Panda	1986	BRD	0	0	0	1	0	R	R	R	R	R		n		y
Plasstarke	2007	HOL	1	0	0	1	0	I	R	R		R		y		y
Seresta	1994	HOL	1	0	0	1	0	R	R	R		R		y		y
Spectra	2016	HOL	NA	0	0	1	r	R		R				?	?	?
Andante	2004	BRD	1	0	0	r	0	R	R	R	R	R		n		y

Aventra	2005	HOL	0	0	0	0	1	S	R	R		R				?	
Kanjer	1996	HOL	0	0	0	0	1	I	R	R						y	
VE71-105		HOL	1	0	0	0	1	R	I	R		S				y	
Achilles	2008	HOL	1	0	0	0	0	R									
Adretta	1975	DDR	1	0	0	0	0	R									
Agria	1985	BRD	1	0	0	0	0	R	S	S							
Alegria	2003	BRD	1	0	0	0	0	R		S		S					
Annabelle	2001	HOL	1	0	0	0	0	R	S	S		S					
Antares	1961	DDR	1	0	0	0	0	R	R								
Apollo	1956	DDR	1	0	0	0	0	R	R			R					
Aquila	1942	GER	0	0	0	0	0	R									
Astarte	1976	HOL	1	0	0	0	0	R	I	I		R					
Asterix	1991	HOL	1	0	0	0	0	R	S								
Atlantic	1976	USA	0	0	0	0	0	S									
Aveka	2001	HOL	1	0	0	0	0	R		I		S					
Avenance	2005	HOL	1	0	0	0	0	R	S	R		I					
Aviala	2004	HOL	1	0	0	0	0	R	R	R		S					
Belita	1986	HOL	1	0	0	0	0	R	R	R		R					
Bintje	1910	HOL	0	0	0	0	0	S		S							
Capella	1943	GER	1	0	0	0	0	R									
Carrera	1999	HOL	1	0	0	0	0	R	S	R		S					
Caruso	2004	BRD	0	0	0	0	0	R	R	R		S					
Colomba	2011	HOL	1	0	0	0	0	R	S	S		S					
Delcora	1988	HOL	1	0	0	0	0	R	R	R		S					
Deodara	1913	GER	0	0	0	0	0	S	S	S	S	S					
Desiree	1962	HOL	1	0	0	0	0	R	S	S	S	S					
DM 1-3 516 (R44)			0	0	0	0	0										
Elkana	1978	HOL	1	0	0	0	0	R	S	S		S					
Eurobona		BRD	1	0	0	0	0	R		R							
Euroviva		BRD	1	0	0	0	0	R	R	R	R	R					
Festien	2000	HOL	0	0	0	0	0	S	R	R		I					
Fidelia		GER	1	0	0	0	0	R									
Folva	1989	DEN	1	0	0	0	0	R	S	S		S					
Giewont	1955	POL	1	0	0	0	0	R	S	S	S	S					
Hansa	1957	BRD	0	0	0	0	0	S	R	S							
HB93-7108-08			0	0	0	0	0										
Inca Sun			0	0	0	0	0										
Innovator	1999	HOL	1	0	0	0	0	R	S	S		S					
Kantara	1997	HOL	1	0	0	0	0	R	R	R		S					
Kardal	1988	HOL	1	0	0	0	0	R		S							
Karlana	1988	DDR	1	0	0	0	0	R									
Kuras	1996	HOL	1	0	0	0	0	R	S	S							
Labadia	2001	HOL	1	0	0	0	0	R									
Lorch	1931	SU	0	0	0	0	0										
Ludmilla	2008	BRD	1	0	0	0	0	R	S	S		S					
Markies	1997	HOL	0	0	0	0	0	S	S	S							
Mayan Gold			0	0	0	0	0										
Mayan Twilight			0	0	0	0	0										
Miriam	1988	BRD	1	0	0	0	0	R	R	R		S					
Monalisa	1982	HOL	1	0	0	0	0	R	R								
MPI 44.1016_24		GER	0	0	0	0	0	R	R	R		R					
Nomade	1995	HOL	1	0	0	0	0	R	R	I		S					
Opal	1961	POL	1	0	0	0	0	R	S	I		S					
Otolia		BRD	1	0	0	0	0	R		R		R					
Panamera	2013	HOL	1	0	0	0	0	R	S	S		S					
Parella		HOL	1	0	0	0	0	R	R	R		S					

Paroli		GER	1	0	0	0	0	R											
Plasettie	2009	HOL	1	0	0	0	0	R		R			R						
Polessky Rosovy	1978	SU	0	0	0	0	0												
Princess	1998	BRD	0	0	0	0	0	S	S	S			R						
Producent	1984	HOL	1	0	0	0	0	R	S	S			S						
Provento	1994	HOL	1	0	0	0	0	R	R	R									
RH90-37-012			0	0	0	0	0												
Royal	2008	DEN	1	0	0	0	0	R	R	R			S						
Saphir	1960	BRD	1	0	0	0	0	R	S	R	R		R						
Sarpo Mira	<2003	HUN	0	0	0	0	0	S		R									
Simply Red	2001	HOL	1	0	0	0	0	R	S				R						
Smart	2008	HOL	1	0	0	0	0	R	S	S			R						
Starga	2000	HOL	1	0	0	0	0	R	R	I									
Talent	2006	BRD	1	0	0	0	0	R	I	I			S						
Tivoli	2002	DEN	1	0	0	0	0	R	R	R			R?						
Tomensa	1989	BRD	0	0	0	0	0	S	S	S	S		S						
Ultra	1999	HOL	0	0	0	0	0	S		R									
Van Gogh	1989	HOL	0	0	0	0	0	R											
VR-808	2009	HOL	1	0	0	0	0	R	S	S			S						
VTN62.33.3		HOL	1	0	0	0	0	R	S	S			S						
Atlantic*	1976	USA	0		0	0	0	S											
Burbank*	1876	USA	1		0	0	0	R											
Dakota Diamond*	2006	USA	0		0	0	0												
Early Rose*	1867	USA	0		0	0	0	S											
Garnet Chili*	1857	USA	0		0	0	0												
Irish Cobbler*	1876	USA	1		0	0	0	R											
Kalkaska*	2007	USA	0		0	0	0												
Katahdin*	1932	USA	0		0	0	0	S											
Kennebec*	1948	USA	0		0	0	0	S											
Missaukee*	2009	USA	0		0	0	0												
Mountain Rose*		USA	0		0	0	0												
Norland*	1957	USA	1		0	0	0												
Premier Russet*	2006	USA	1		0	0	0												
Purple Majesty*		USA	0		0	0	0												
Rio Grande Russet*		USA	0		0	0	0												
Russet Burbank*	1908	USA	1		0	0	0	R											
Russet Norkotah*	1987	USA	1		0	0	0												
Shepody*	1980	CAN	0		0	0	0	S											
Sierra Gold*	2003	USA	0		0	0	0												
Snowden*	1990	USA	0		0	0	0												
Spunta*	1968	HOL	1		0	0	0	R											
Superior*	1961	USA	0		0	0	0												
Yukon Gold*	1980	CAN	0		0	0	0	I											

^a The presence/absence of the *Sen* genes is coded as 0: absent, 1: present, r: recombinant and is based on results from Supplementary file 3. ^b Resistance scores, R: resistant, I: intermediate, S: susceptible. ^c Putative resistance sources in the varieties pedigree, y: yes, n: no, ?: unknown. * These varieties were screened for *Sen1*, *Sen3*, *Sen4* and *Sen5* using CoSSA, see Supplementary file 5

In the variety panel, only Aventura, Kanjer and the breeding clone VE71-105 hold *Sen5*. Aventura and VE71-105 are resistant to P2, P6 and slightly resistant to P18, Kanjer is resistant to P2 and P6 (it is slightly susceptible to P1 and its resistance to P18 is unknown). Aventura's pedigree is unknown and Kanjer holds VE71-105 in its pedigree (40).

Sen resistance gene recognition spectrum and representation in the EPPO panel

Among the 54 varieties from our panel which hold *Sen1* but do not hold any of the identified *Sen* genes, 51 are resistant to P1 and 19 do not show resistance to any of the higher pathotypes (Table 1). The resistance pattern of *Sen1* is specific to the pathotype 1 of *S. endobioticum*, as observed previously in several studies (13,14,20,41). *Sen2* was shown to bring resistance to P1, P2, P3, P6, P8, P18 and P38 (20). In our variety panel, the *Sen2* marker was observed in Bonza which is resistant to pathotypes 1, 2 and 6. In our variety panel, there are 11 varieties which hold *Sen3* without *Sen1*, ten of which are known to be resistant to P1 (Tables 1 and 2). Moreover, five of these varieties are also resistant to pathotype 8. These observations suggest that *Sen3* is involved in resistance to pathotypes 1, 2, 6, 8 and 18. *Sen4* was shown to bring resistance to the tested pathotypes 2, 6 and 18 in the AxV population. In the variety panel, only two varieties hold *Sen4* without *Sen1*, one being susceptible to P1 (Celandine) and one resistant (Panda) (Tables 1 and 2). Therefore, no conclusion can be made about an extended resistance spectrum of *Sen4*. *Sen5* is present in two varieties that do not hold *Sen1* and are susceptible (Aventra) or slightly susceptible (Kanjor) to P1 (Tables 1 and 2). This observation and the study we made in the AxD and KxA populations show that the resistance spectrum of *Sen5* is specific to pathotypes 2, 6 and 18.

Among the varieties from the EPPO set (42,43), *Sen1* was found in six of the ten varieties resistant to P1 (Table 2). *Sen3* was present in the four remaining varieties resistant to P1 which are also resistant to P2, P6 and P18. Neither *Sen2*, *Sen4* or *Sen5* was found in the EPPO set varieties. The resistance present in Talent, Delcora, Saphir, Miriam and Belita is not explained by any of the *Sen* genes identified so far.

Resistance in *Solanum* species

A panel of 118 accessions from 38 wild *Solanum* species was screened with markers flanking the *Sen* genes (Figure 4, Supplementary file 3) and an additional set of 44 wild species and landraces (23) was analysed using CoSSA. The *Sen1* linked marker chr11_1666351 was found in seven accessions of the Tuberosa 3 series (44) and one diploid *S. bulbocastanum* accession. Three of the accessions holding the *Sen1* flanking marker are known to be resistant to P1 (Supplementary file 3). With CoSSA, we observed that *Sen1* was found in only one accession (landrace PI 258885), belonging to the *S. tuberosum* species from the Chilotanum subgroup, which shared 63.3% of

the *Sen1* specific *k*-mers mapping between 1.2 Mb and 1.7 Mb on chromosome 11 (Supplementary file 5).

Table 2. *Sen* genes resistance spectrum

		P1	P2	P6	P18	Sen1	Sen2	Sen3	Sen4	Sen5
EPO set	Tomensa	S	S	S	S	-	-	-	-	-
	Deodara	S	S	S	S	-	-	-	-	-
	Producent	R	S	S	S	+	-	-	-	-
	Talent	R	I	I	S	+	-	-	-	-
	Delcora	R	R	R	S	+	-	-	-	-
	Saphir	R	S	R	R	+	-	-	-	-
	Miriam	R	R	R	S	+	-	-	-	-
	Ikar	R	R	R	R	-	-	+	-	-
	Gawin	R	R	R	R	-	-	+	-	-
	Karolin	R	R	R	R	-	-	+	-	-
	Ulme	R	R	R	R	-	-	+	-	-
	Belita	R	R	R	R	+	-	-	-	-
	Celandine	S	S	R	I	-	-	-	+	-
	Panda	R	R	R	R	-	-	-	+	-
	Aventra	S	R	R	R	-	-	-	-	+
	Kanjer	I	R	R	?	-	-	-	-	+

With the *Sen2* CAPS marker, we identified positive signals in *Solanum acaule* accessions from the Acaulia series. Three accessions from the Demissum series were also positive for the *Sen2* marker (one *S. demissum* accession and two *S. guerreroense*). Finally, three accessions from the Tuberosa series (one *S. spegazzini* and one *S. verrucosum*) were positive for the *Sen2* marker (Figure 4, Supplementary file 3, Supplementary figure 13).

For *Sen3*, *Sen4* and *Sen5*, we observed that none of the screened accessions held both flanking markers. The *Sen3* flanking marker chr11_1259552 was present in 24 accessions whereas the other flanking marker chr11_1772869 was present in only three accessions (Supplementary file 3). This second marker mapped closer to *Sen3*

in the KxL population (11) was present in two *S. guerreroense* accessions (Demissa series) which have an intermediate level of resistance to P8 and one *S. tuberosum* ssp. *andigena* accession (Tuberosa 3 series). With CoSSA, *Sen3* was also found in a *S. tuberosum* landrace (PI 245847) from the Chilotanum subgroup. It shared 93.2% of the *Sen3* specific *k*-mers mapping between 1.2 Mb and 1.8 Mb on chromosome 11 (Supplementary file 5).

The *Sen4* flanking markers chr12_48501410 and PotVar0037404 were found each in one accession (Supplementary file 3). In our AxV population, PotVar0037404 co-segregated with *Sen4*. It was found in a *S. circaelifolium* accession from the Circaelifolia series, known to have an intermediate level of resistance to P2 and P6. For *Sen4* with CoSSA, it was more difficult to determine if an accession held the gene or not as a smaller proportion of *k*-mers was shared in the wide genomic region identified. To determine if an accession held the *Sen4* gene, we compared the distribution pattern of *Sen4* specific *k*-mers mapped to each chromosome 12 bin between Axion and the tested accessions (Supplementary file 5). The *S. brevicaulis* accessions PI 498112 and PI 545987 shared 42.6% and 43% respectively of the *Sen4* specific *k*-mers between 48.5 Mb and 50.7 Mb, and had a distribution pattern of mapped *k*-mers similar to Axion. It was also the case of four *S. tuberosum* group Andigena accessions (PI 245935, PI 245940, PI 546023 and PI 607886) and one *S. tuberosum* group Chilotanum (PI 245847) accession in which was found *Sen3* as well.

The *Sen5* flanking markers PotVar0035016, which was co-segregating with *Sen5* in AxD and KxA, and PotVar0034831, which was two recombination events from *Sen5*, were found in zero and 40 accessions, respectively (Supplementary file 3). PotVar0034831 was found in all the accessions from the Demissa and Acaulia series, nine accessions from the Tuberosa 3 series (including five *S. vernei* accessions), seven accessions from the North and central American diploid series and two accessions from the Circaeifolia series (Figure 4). Fourteen of these accessions are known to show an intermediate or strong level of resistance to at least one of the higher pathotypes (Supplementary file 3). *Sen5* was found with CoSSA in the *S. vernei* accession PI 473305 which contained 56.4% of the *Sen5* specific *k*-mers between 49.2 Mb and 51.2Mb (Supplementary file 5).

Figure 4. Classification of *Solanum* accessions screened for the *Sen* genes

Bayesian rooted tree of 108 *Solanum* accessions screened for the *Sen* genes, based on the genotypic data of 222 AFLPs. Branch lengths are proportional to the number of changes/site, posterior probabilities are shown at each node. The classification is adapted from Hawkes' series (44). The *S. etuberosum* accessions were included in the phylogenetic analysis to form the outgroup but they were not screened for the *Sen* genes. The presence of *Sen1*, *Sen2*, *Sen3*, *Sen4* and *Sen5* is based on markers chr11_1666351, Sen2_CAPS, chr11_1772869, PotVar0037404 and PotVar0034831 respectively

Discussion

Identification of *Sen4*, important source of resistance in the Dutch breeding material

In this study, we performed a GWAS in the potato breeding genepool to identify markers linked with pathotypes 2, 6 and 18 resistance. The GWAS and the validation of the significant markers in a full-sib population allowed to identify the major effect locus *Sen4*, located on the south arm of chromosome 12, linked with P2, P6 and P18 resistance. CoSSA was used to identify new markers flanking *Sen4* which could be mapped to a 3 Mb interval between 48.5 Mb and 51.5 Mb. In this region of the reference genome, no large NLR or RLP or RLK clusters are found (38,39). Only a single CNL gene (DMG401007575; (38)) and a single RLK (StRLK66) were found as candidate genes. We found that *Sen4* was present in the variety Panda which was used as the resistant parent of the mapping population in the study of (14). A QTL matching the *Sen4* locus was, however, not identified in their study which is probably due to the use of a limited number of SSR markers. In the BNA2 population from (15), the locus *RSe-XIIa* from the central region of chromosome 12 was associated with P2, P6 and P18. The alternative allele of Solcap_c2_33630 was however present in many susceptible varieties and was present in duplex in the resistant parent (Karolin) and in the susceptible parent (unknown). For these reasons, it is unlikely that *RSe-XIIa* is the same locus as *Sen4*. Moreover, we tested Karolin which was negative for *Sen4*.

We observed that *Sen4* is Identical-By-Descent (IBD) in the Dutch varieties and could be traced back with markers to the breeding clone AM78-3704 and even further to its grand-parent Alcmaria. Alcmaria comes from a cross between Sirtema and a progeny from Saskia x (CPC 1673-20 x Furore) (40). We tested genome sequencing data from Sirtema for the chromosome 12 associated SNPs but could not find them. So, we concluded that Sirtema does not contain *Sen4* (data not shown). Therefore, *Sen4* must come from Furore or CPC 1673-20. Interestingly, wart resistance had already been observed in CPC1673 which was used to introgress nematode resistance in the breeding genepool (19). The only variety which holds *Sen4* not IBD from Alcmaria is Panda. Panda's *Sen4* allele probably comes from a different source than the Dutch material. In the genotypic data from (37), we observed that the *Sen4* associated markers were present in the variety Hindenburg which is resistant to the pathotype 1 (45) but whose resistance to the higher pathotypes is unknown. Hindenburg is present in the pedigree of Panda through the varieties Aquila and Mittelfruhe. We tested Aquila

which was negative for the *Sen4* markers. Testing the Mittelfruhe pedigree branch with the *Sen4* markers may reveal if Hindenburg can be the German source of *Sen4*.

In the Solanum panel, the *Sen4* marker the most closely linked with resistance was present in one *S. circaeifolium* accession (Circaeifolia series). *S. circaeifolium* is known to contain resistance to other pathogens such as *P. infestans*, *G. pallida* and *E. carotovora* (46). This accession was phenotyped and found to be moderately resistant to P2 and P6 but susceptible to P8. In the panel from (23), *Sen4* was found in two *S. brevicaulis* accessions, a landrace from the group Chilotanum and several landraces from the *S. tuberosum* group Andigena. This last observation strengthens our hypothesis that the Dutch material source of *Sen4* might be the clone CPC1673 which possesses Andigenum origins.

Identification of *Sen5*, underrepresented source of resistance in the breeding gene pool

The two bulks from (8) were re-used in an adapted CoSSA workflow to characterise the resistance from Aventura. We identified and mapped *Sen5* on the south arm of chromosome 5 between 49.2 Mb and 51.16 Mb. In the reference genome, the C40 and C41 NLR clusters are located between 49.48 Mb and 49.69 Mb (38) and there are four gaps in this region in the reference genome (Figure 3). In the variety panel, *Sen5* was very rare as it was found in only two varieties (Aventura and Kanjer) and one breeding clone (VE71-105). A closer look to the pedigree showed that Kanjer is a descendant of VE71-105. The pedigree of Aventura is unknown, but we hypothesize that these three resistant genotypes are IBD for *Sen5*, sharing VE71-105 as the donor. The screening of the two flanking markers in the Solanum panel gave contrasting results as one marker was not found in any accession and the other was found in 40 accessions including all the accessions from the Demissa and Acaulia series, accessions from the Tuberosa series (including several *S. vernei* accessions) and several accessions from North and Central America and the Circaeifolia series. As it is unlikely that *Sen5* is present in so many accessions from so many different phylogenetic series, we hypothesize that marker PotVar0034831 detects a more common SNP adjacent to the introgression segment of the *Sen5* donor haplotype. CoSSA gave less ambiguous results than the marker analysis, and *Sen5* was found in one *S. vernei* accession. This is in agreement with the observation that VE71-105 has multiple *S. vernei* sources in its pedigree (40). *S. vernei* has previously been reported as a potential source of wart disease resistance (6).

Diagnostic power of the *Sen1* flanking markers and origin of *Sen1*

In the SaKa1 population, (13) identified the *Sen1-XI* locus on the north arm of chromosome 11 which brings a quantitative resistance to P1. The authors suspected this locus to be an allelic variant of *Sen1* or a different *R* gene from the same cluster. However, the screening of the *Sen1* flanking markers from (8) showed that *Sen1* is linked to P1 resistance in the SaKa1 population. In the study of (13), the GP259 markers from the *Sen1* region were most associated with P1 resistance. The BLAST of GP259 to DM v4.03 showed this fragment is located between 2,805,301 bp and 2,805,774 bp which is more than 1 Mb from *Sen1*. This may explain why the *Sen1-XI* locus did not bring a clear qualitative resistance to P1, while in our studies the two flanking *Sen1* markers (chr11_1308927 and chr11_1666351) did detect qualitative P1 resistance.

Sen1 was present in 65% of the varieties screened and as described previously (10), it is not possible to trace it back to a common ancestor donor. In the accessions panels, *Sen1* was found mainly in the diploid Tuberosa 3 series (44) accessions originating from Bolivia and Argentina. More specifically, it was found in the *S. tuberosum* group Chilotanum, in *S. microdontum*, *S. vernei* and one accession from *S. bulbocastanum*. Members of the group Chilotanum, adapted to long days, are ancestors of commercial cultivars (*S. tuberosum* ssp. *tuberosum*) and contributed greatly to its genetic background. A high number of introgressions from *S. microdontum* have been found in long-day adapted cultivars (23). The intensive use of Chilotanum group and *S. microdontum* to adapt potato to long-days could have led to multiple introgressions of *Sen1* early in the group Tuberousum which would explain the lack of identity-by-descent (IBD) of *Sen1*. In (17), the *Sen1* linked marker NI25 (7) was identified in two accessions from *S. tuberosum* ssp. *andigenum* and three accessions from *S. tuberosum* ssp. *tuberosum*. In this study, and in the broad panel of accessions screened for P1 resistance by (47), no correlation was found between the resistance to P1 and the taxonomy, the ploidy level or the geographic origin, which might be explained by the presence of *Sen1* very early in the ancestors of cultivated potato.

Sporadic presence of *Sen2* in the breeding material

Sen2 was identified for the first time in the complex diploid hybrid DG 97-264 (20). In our variety panel, none of the varieties showed a similar banding pattern compared to DG 97-264 except for Bonza which is known to be resistant to P1, P2 and P6. However, Bonza's resistance might come from *Sen3* as it holds one of the two *Sen3* flanking markers. The pedigree of Bonza is unknown so no hypothesis can be made about if

Bonza hold a recombination in the *Sen3* haplotype. Most likely, *Sen2* has not yet been included in the breeding programs and is therefore a new source of wart disease resistance.

In the accessions panel, four accessions from the Demissum series (*S. demissum* and *S. guerreroense*), four *S. acaule* accessions, two *S. verrucosum* accessions and one *S. spegazzini* accession showed a banding pattern similar to DG 97-264. *S. acaule*, *S. demissum* and *S. verrucosum* have been used in the pedigree of DG 97-264 and either of them might be the resistance donor. *S. acaule* has been reported as being a potential source of wart disease resistance (6) and more specifically as being the source of the P1, P6 and P18 resistance of Saphir through the donor MPI44.1016/24 (19). We could, however, reject the hypothesis that Saphir or MPI 44.1016/24 were positive to the *Sen2* CAPS marker.

The scoring of the *Sen2*_CAPS marker was not unambiguous, however. In order to validate if variety Bonza and the accessions that we scored positive for *Sen2* indeed hold the same allele as the clone DG 97-264, we would recommend to sequence the *Sen2* marker PCR products. Another unambiguous way to verify the presence of *Sen2* in other sequenced varieties or accessions would be to identify *k*-mers specific to *Sen2*. This could be done by sequencing two bulks of resistant and susceptible descendants from the SEN 12-01 population (20) or by sequencing the PCR fragments which are linked with resistance in DG 97-264 and comparing their *k*-mers with other genotypes using CoSSA.

***Sen3* is an important source of resistance in the German and Polish material**

It has been shown previously that *Sen3* was IBD in all the *Sen3* positive varieties and could be traced back to the variety Ora, a descendant of Capella and BRA9089 (11). In old literature, resistance to the higher pathotypes was traced back to the clone BRA9089 which was found at that time to be resistant to P1 and to the higher pathotypes. In (11), BRA9089 was found to be susceptible to all pathotypes and negative for the *Sen3* flanking markers, which is most likely due to the fact that the old and recent BRA9089 clones are different. In our variety panel, *Sen3* was mainly found in German and Polish varieties. All these *Sen3* varieties for which the pedigree is complete hold BRA9089 in their pedigree. Surprisingly, the *Sen3* markers were also found in the American varieties Defender and Goldrush and in the Dutch variety Lady Sara (descendant of Defender). We did not find any information about the resistance of these varieties to the higher pathotypes, nor did we pursue phenotyping of these

clones. These three varieties also hold BRA9089 in their pedigree, which makes it the putative unique donor of *Sen3* in the breeding genepool. BRA9089 originates from crosses involving a cultivar from Chiloe (*S. tuberosum* ssp. *tuberosum*), the cultivar Svitez and a landrace (18,19,40).

In our *Solanum* panel, none of the *Solanum* accessions were positive for both flanking markers which can be explained by the fact that one or both of them are located outside the introgression segment. The closest marker to *Sen3* was found in the two *S. guerreroense* accessions and in an accession from the *S. tuberosum* group Andigena. This group has been reported before as being a source of potato wart disease resistance (18,19,47). CoSSA gave more robust results and clearly showed the presence of both *Sen3* and *Sen1* in one landrace from *S. tuberosum* group Chilotanum, showing that *Sen1* and *Sen3* were already present in this ancestral landrace of commercial cultivars.

***Rse-Ib-Andante* shows co-dominant resistance to pathotypes 1, 2, 6 and 18**

Several alleles on the south arm of chromosome 1 linked with P1, P2, P6 and P18 resistance or susceptibility were identified by (13) in SaKa1. *Sen2/6/18_a* linked with resistance is the same as the *Rse-Ib-Andante-a* haplotype that we identified (Supplementary table 5). The haplotype *Sen2/6/18_b* linked with susceptibility is the same as our *Rse-Ib-Andante-b* haplotype. We identified an extra haplotype at the *Rse-Ib-Andante* locus linked with resistance, *Rse-Ib-Andante-c*, which is one of the two alleles on which the marker chr01_76425362 is present. With CoSSA, we mapped *Rse-Ib-Andante* to the south arm of chromosome 1 between 70 Mb and 79.6 Mb. In the reference genome, this region contains the two NLR clusters C4 and C5 (38) and the four RLK genes *StRLK04*, *StRLK05*, *StRLK06* and *StRLK07* (39). Interestingly, the two CNL genes from the cluster C4 are homologues of NRC1, a helper NLR which is required by cell surface and intracellular immune receptors (48). The involvement of different alleles of this helper gene in resistance to wart disease mediated by *Rse-Ib-Andante* might explain the quantitative, genetic background dependent effect of *Rse-Ib-Andante*. The different resistance levels induced by *Sen* genes, which could serve as sensor NLRs, would then depend on the NLR helper they can pair with.

Identification of minor effect potato wart disease resistance QTLs

In the three populations AxV, AxD and KxA, we identified *Rse-XIc-VTN62.33.3*, which brought a quantitative resistance to pathotypes 2, 6 and 18. The recognition pattern of this haplotype depended on the population in which it segregated and on the phenotyping assay that was used: *Rse-XIc-VTN62.33.3* was associated with P2 and P6

resistance in AxV, with P6 and P18 in AxD and with P2 and P6 (Glynne-Lemmerzähl) and P18 (Spieckermann) in KxA. One reason for these differences could be that different isolates were used in the different tests. This could be the reason for the difference between the AxV and the AxD/KxA populations, but not for the difference between AxD and KxA which were phenotyped with the same isolates. A second explanation could be that several genes underlie this QTL and that recombinations happened between them in some of the populations. A third explanation could be a background dependency of this QTL. *Rse-XIc-VTN62.33.3* was mapped to the 5 first Mb of the north arm of chromosome 11. This broad region of chromosome 11 contains the three NLR genes clusters C76, C77 and C78 (38), two RLK genes and one RLP (39) in the potato reference genome.

In the SaKa1 population, (13) identified the *Sen18-IX* locus bringing resistance to P18, located on the south arm of chromosome 9. We also identified this locus in our study, and re-named it *Rse-IXa-Andante* to comply with the new naming system we introduced in this study. We mapped *Rse-IXa-Andante* between 55 Mb and 61.5 Mb on chromosome 9. This region is rich in NLR genes as it contains the C64, C65 and C66 clusters (38), and four RLK genes (39) in the reference genome. Another minor effect QTL for pathotypes 2, 6 and 18 from the variety Ludmilla was identified at this locus by (11). The Ludmilla haplotype was different from *Rse-IXa-Andante*, as determined using CoSSA, so we renamed it *Rse-IXb-Ludmilla* to avoid confusions and comply with our new naming system.

In the SaKa1 population, we identified an extra haplotype linked with P2, P6 and P18 on the north arm of chromosome 11 which was not identified previously. *Rse-XId-Andante* is different from *Rse-XIc-VTN62.33.3* identified in the AxV, AxD and KxA populations, as determined using CoSSA. With CoSSA, we mapped *Rse-XId-Andante* to a very broad region between 0 and 31 Mb.

In the KxL population (11), we identified *Rse-VIIIb-Kuba*, a minor effect locus on the south arm of chromosome 8 which improved the resistance to P18 brought by *Sen3*. The two SNPs linked with P18 resistance identified with CoSSA (chr08_44797542 and chr08_45178832) are located in a region rich in NLR gene clusters as it contains the three NLR clusters C59 (45.03 Mb - 45.12 Mb), C60 (47.54 Mb - 48.15 Mb) and C61 (48.37 Mb - 48.6 Mb) (38).

Do major and minor effect loci complete the resistance spectrum to cover intra-isolate diversity?

In this study, we observed that the stacking of major and minor effect loci was increasing the resistance levels to potato wart disease. In the AxV population, the presence of *Sen4* and *Rse-XIc-VTN62.33.3* improved the resistance to P2 and P6. Similarly, in the AxD and KxA populations, the stacking of *Sen5* and *Rse-XIc-VTN62.33.3* improved resistance to P2, P6 and P18. In the SaKa1 population, the stacking of *Rse-Ib-Andante* and *Rse-XId-Andante* for P2 and P6 and of *Rse-IXa-Andante* for P18 was required to give strong resistance. Finally, in the KxL population, we observed that the presence of *Rse-VIIIb-Kuba* improved the resistance to P18 brought by *Sen3*. An explanation for the additive effect of minor- to major effect loci in potato wart disease resistance can be sought in the genotypic diversity within *S. endobioticum* isolates used for the phenotypic assays (49). For instance, in the case of the KxL population, we can hypothesize that *Sen3* is a TNL that recognizes an effector which is present in all genotypes of the P2 and P6 isolates and in the majority of the genotypes of the P18 population. In concert, *Rse-VIIIb-Kuba* might be a CNL that recognizes an effector which is in a minority of the P18 isolate population. A similar explanation could hold for *Rse-Ib-Andante*, *Rse-IXa-Andante*, *Rse-XId-Andante* and *Rse-XIc-VTN62.33.3* which also reside in NLR clusters. Overall, the stacking of several loci to bring full resistance seems to be essential for resistance to higher pathotypes. The higher intra-isolate diversity observed in isolates from higher pathotypes such as P8, P18 and P38 would indeed explain why resistance to the higher pathotypes is more quantitative (49).

Unknown resistances remain in the breeding genepool

There are resistances present in the potato breeding germplasm which do not come from any of the identified *Sen* genes and are, therefore, still unknown. It is the case for instance, of the P2, P6 and P18 strong resistance of Belita which probably comes from *S. vernei* through the clone VTN62-33-3, donor of *G. pallida* resistance, or through the clone VTN61-13-9 (40). However, when we phenotyped VTN62-33-3, it turned out to be susceptible to P2, P6 and P18. The resistance to P6 and P18 present in Saphir is also not caused by *Sen1* 2, 3, 4, or 5. This resistance likely comes from the breeding clone MPI 44.1016/24, which shows strong resistance to pathotypes 1, 2, 6 and 18, and *S. acaule*, as it was hypothesized by Ross (19,40).

The EPPO set does not encompass the *R* loci diversity present in the potato breeding gene pool

The EPPO set is composed of differential potato varieties which are used for the identification of *S. endobioticum* pathotypes (42). The varieties from the EPPO set resistant to P1 contain either *Sen1*, or *Sen3*. *Sen1* is a TNL gene which triggers a Hypersensitive Response (HR) upon the recognition of the *S. endobioticum* effector AvrSen1 present in the pathotype 1 isolates (9). *Sen3* is also a TNL gene and we suspect it to bring resistance to P1 and P8 in addition to P2, P6 and P18 (11) as it is present in several varieties resistant to P1 that do not contain *Sen1* and in several varieties resistant to P8. Moreover, we suspect that the chromosome 11 locus segregating in the BNA2 population is *Sen3*, and it was causing P1 resistance in addition to P2, P6 and P18 in BNA2 (15). *Sen2*, *Sen3*, *Sen4* and *Sen5* are not represented in the EPPO set. *Sen5* specifically recognizes P2, P6 and P18 isolates and not P1 isolates, whereas *Sen2* and *Sen3* can recognize P1 in addition to P2, P6, P8 and P18. *Sen2* was even shown to provide P3 and P39 resistance (20). The inclusion of varieties holding the *Sen2*, *Sen4* and *Sen5* genes in the EPPO set might lead to the identification of additional *S. endobioticum* diversity or new pathotypes. Moreover, it will give direct information to inform about the varieties for quarantine deployment. As discussed previously, there are also varieties from the EPPO set for which the resistance loci have not been identified. This is the case for Talent, Delcora, Saphir, Miriam and Belita. For a more complete understanding of pathogen diversity, it is important to better characterize and update the EPPO set and base it on the *R* loci content of the varieties and the matching effectors they recognize.

CoSSA is more robust than molecular markers for the screening of specific loci

In this study, we used two different approaches to screen panels of varieties and *Solanum* accessions for the five *Sen* resistance genes. The first approach was a screening using PCR markers flanking the *Sen* genes. This approach was successful and efficient to screen the variety panel but not the wild accessions. This was not due to the different ploidy levels of these accessions as the clustering in the KASP markers output allows to mix different ploidy levels. The most probable reason is that one or all the flanking markers designed were not close enough to the resistance gene. If the SNP is not located on the introgression segment but on the *S. tuberosum* haplotype which recombined with the introgression segment, the SNP will not be linked with resistance in a panel composed of wild species and will lead to false positives. In the second approach, we used sequencing data and *k*-mers specific to the resistant

haplotypes to verify the presence of the *Sen* genes. The results were unambiguous for *Sen1*, *Sen3* and *Sen5*. We could visualize the chromosome bins of the tested varieties/accessions which contained the resistance specific *k*-mers. This visualization allows to determine unambiguously if the (fine-)mapped resistance locus is present in the tested genotype and it even allows to detect recombination events. This is the case for instance for *Sen1*. In Supplementary file 5, we can observe some varieties such as Atlantic (susceptible to P1), Kalkaska, Missaukee and Kennebec (susceptible to P1), which hold only the south part of the *Sen1* haplotype between 2.4 Mb and 4.6 Mb. It is this part of the *Sen1* haplotype which was identified in the pathotype 1 GWAS study of (10) as no markers located on the northern part of this haplotype, closer to the *Sen1* gene, were present on the 20K Infinium array. The recombination event in Atlantic and Kennebec is visible in the Supplementary file 1 of the GWAS study between PotVar0067008 and solcap_snp_c1_2314, PotVar0106272, PotVar0106247 and PotVar0105904, and showed that *Sen1* was no longer present. For *Sen4*, it was more difficult to determine the presence of the gene in the CoSSA output because the *Sen4* specific *k*-mers peak was very broad (between 10.3 Mb and 54 Mb) because insufficient recombinants were present in the resistant bulk.

Conclusion

In this study, two new dominant major effect wart resistance genes, *Sen4* and *Sen5*, bringing resistance to higher pathotypes of *S. endobioticum*, were identified and mapped. Several minor effect QTLs were also identified that are necessary to bring a full resistance to certain pathotypes. Panels of varieties and Solanum accessions were screened for the presence of the *Sen* genes. *Sen1*, highly frequent in the commercial varieties, was observed in a landrace from the *S. tuberosum* group Chilotanum and in several accessions from *S. microdontum* which might explain the presence of *Sen1* very early in the commercial varieties and the lack of identity-by-descent due to multiple introgressions. We suspect *Sen2* to be present in accessions from the Demissum series, *S. acaule*, *S. verrucosum* and/or *S. spgazzini*. *Sen3* could be traced back to the clone BRA9089 and is IBD in all the varieties from commercial germplasm. It was also observed in the Chilotanum group. *Sen4*, identical-by-descent in the Dutch varieties, could be traced back to the variety Alcmaria and was probably introduced in commercial varieties through the Andigena group and the breeding clone CPC1673-20 used to breed for nematode resistant material. *Sen5* is very rare in the commercial varieties. It could be traced back to the breeding clone VE71-105 and the wild species *S. vernei*.

Acknowledgements

We acknowledge all the partners involved in this public-private partnership (Averis seeds BV, Böhm-Nordkartoffel Agrarproduktion GmbH & Co. OHG, Danespo, HLB BV, HZPC Holland BV, C Meijer BV, SaKa Pflanzenzucht GmbH & Co. KG and Teagasc) for providing invaluable potato material, phenotypic data, KASP marker analysis, and scientific discussions. Wilbert van Workum and Limes Innovations are acknowledged for facilitating WGS at Hartwig Medical. We acknowledge Jarosław Przetakiewicz and IHAR for the phenotypic assays they performed. Fien Meijer-Dekens and Arnaud Bovy are thanked for making available DNA from the wild *Solanum* accessions.

Funding

This work was performed in a project called “An integrated genomics and effectoromics impulse for potato wart resistance management and breeding” which was financially supported by the Dutch Topsector Horticulture & Starting Materials. Within the Topsector, private industry, knowledge institutes and the government are working together on innovations for sustainable production of safe and healthy food and the development of a healthy green environment. CP was supported by a Walsh fellowship provided by Teagasc.

Authors contribution

CP, HJvE and JHV conceived and designed the experiments. CP, GVA, JET and JR performed the experiments and analyses. CP analyzed all the results with the help of GVA, JET, JR and JHV. CP wrote the manuscript. JHV and RGFV helped to write, improve and review the latest version of this manuscript. All authors read and approved the final manuscript.

References

1. Savary S, Willocquet L, Pethybridge SJ, Esker P, McRoberts N, Nelson A. The global burden of pathogens and pests on major food crops. *Nat Ecol Evol*. 2019 Mar;3(3):430–9.
2. Curtis KM. IX. The life-history and cytology of *S. endobioticum* (schilb.), perc., the cause of wart disease in potato. *Philosophical Transactions of the Royal Society of London Series B, Containing Papers of a Biological Character*. 1921;210(372–381):409.
3. Przetakiewicz J. The Viability of Winter Sporangia of *Synchytrium endobioticum* (Schilb.) Perc. from Poland. *Am J Potato Res*. 2015 Dec;92(6):704–8.
4. Hampson MC. Control of potato wart disease through the application of chemical soil treatments: a historical review of early studies (1909–1928). *EPPO Bull*. 1988 Mar;18(1):153–61.
5. Hampson MC, Proudfoot KG. Potato wart disease, its introduction to North America, distribution and control problems in Newfoundland. *FAO Plant Prot Bull*. 1974;22:53–64.
6. Maris B. Races of the potato wart causing fungus *Synchytrium endobioticum* (Schilb.) Perc. and some data on the inheritance of resistance to race 6. *Euphytica*. 1961;10(3):269–276.
7. Hehl R, Faurie E, Hesselbach J, Salamini F, Whitham S, Baker B, *et al*. TMV resistance gene *N* homologues are linked to *Synchytrium endobioticum* resistance in potato. *Theor Appl Genet*. 1999;98(3):379–386.
8. Prodhomme C, van Arkel G, Bergervoet M, Esselink D, Rijk J, Makechemu M, *et al*. *Sen1* fine-mapping and candidate gene analysis. 2020.
9. van de Vossen BTLH, Prodhomme C, van Arkel G, van Gent-Pelzer MPE, Bergervoet M, Brankovics B, *et al*. The *Synchytrium endobioticum* AvrSen1 Triggers a Hypersensitive Response in *Sen1* Potatoes While Natural Variants Evade Detection. *Mol Plant Microbe Interact*. 2019 Sep 17;MPMI-05-19-0138.
10. Prodhomme C, Vos PG, Paulo MJ, Tammes JE, Visser RGF, Vossen JH, *e.a*. Distribution of P1(D1) wart disease resistance in potato germplasm and GWAS identification of haplotype specific SNP markers. *Theor Appl Genet*. 2020 <https://doi.org/10.1007/s00122-020-03559-3>;
11. Prodhomme C, Esselink D, Borm T, Visser RGF, van Eck HJ, Vossen JH. 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* [Internet]. 2019 Dec [cited 2019 May 30];15(1). Available from: <https://plantmethods.biomedcentral.com/articles/10.1186/s13007-019-0445-5>
12. Bartkiewicz A, Chilla F, Terefe-Ayana D, Lübeck J, Strahwald J, Tacke E, *et al*. Improved genetic resolution for linkage mapping of resistance to potato wart in monoparental dihaploids with potential diagnostic value in tetraploid potato varieties. *Theor Appl Genet*. 2018 Dec;131(12):2555–66.
13. Ballvora A, Flath K, Lübeck J, Strahwald J, Tacke E, Hofferbert H-R, *et al*. Multiple alleles for resistance and susceptibility modulate the defense response in the interaction of tetraploid potato (*Solanum tuberosum*) with *Synchytrium endobioticum* pathotypes 1, 2, 6 and 18. *Theor Appl Genet*. 2011 Dec;123(8):1281–92.
14. Groth J, Song Y, Kellermann A, Schwarzfischer A. Molecular characterisation of resistance against potato wart races 1, 2, 6 and 18 in a tetraploid population of potato (*Solanum tuberosum* subsp. *tuberosum*). *J Appl Genet*. 2013 May;54(2):169–78.
15. Obidiegwu JE, Sanetomo R, Flath K, Tacke E, Hofferbert H-R, Hofmann A, *et al*. Genomic architecture of potato resistance to *Synchytrium endobioticum* disentangled using SSR markers and the 8.3k SolCAP SNP genotyping array. *BMC Genet* [Internet]. 2015 Dec [cited 2017 Aug 3];16(1). Available from: <http://www.biomedcentral.com/1471-2156/16/38>
16. Simko I, Jansky S, Stephenson S, Spooner D. Genetics of Resistance to Pests and Disease. In: *Potato Biology and Biotechnology* [Internet]. Elsevier; 2007 [cited 2019 Sep

- 29]. p. 117–55. Available from: <https://linkinghub.elsevier.com/retrieve/pii/B978044451018150049X>
17. Khiutti A, Afanasenko O, Antonova O, Shuvalov O, Novikova L, Krylova E, *et al.* Characterization of resistance to *Synchytrium endobioticum* in cultivated potato accessions from the collection of Vavilov Institute of Plant Industry: Resistance to *Synchytrium endobioticum* in cultivated potato accessions. *Plant Breed.* 2012 Dec;131(6):744–50.
 18. Bukasov SM, Kameraz AJA. Bases of potato breeding. Gosudarstvennoe Izdatel'stvo Sel'skhozjajstvennoj Literatury, Moskva, Leningrad. 1959;528.
 19. Ross H. Potato breeding - Problems and perspectives. Suppl. to *J. Plant Breeding*, Paul Parey, Berlin and Hamburg. 1986.
 20. Plich J, Przetakiewicz J, Śliwka J, Flis B, Wasilewicz-Flis I, Zimnoch-Guzowska E. Novel gene *Sen2* conferring broad-spectrum resistance to *Synchytrium endobioticum* mapped to potato chromosome XI. *Theor Appl Genet.* 2018 Nov;131(11):2321–31.
 21. Giovannoni JJ, Wing RA, Ganai MW, Tanksley SD. Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations. *Nucleic Acids Res.* 1991;19(23):6553–6568.
 22. Michelmore RW, Paran I, Kesseli RV. Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proc Natl Acad Sci.* 1991;88(21):9828–9832.
 23. Hardigan MA, Laimbeer FPE, Newton L, Crisovan E, Hamilton JP, Vaillancourt B, *et al.* Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication in the cultivated potato. *Proc Natl Acad Sci.* 2017 Nov 14;114(46):E9999–10008.
 24. Spieckermann A, Kothoff P. Testing potatoes for wart resistance. *Deutsche Landwirtschaftliche Presse* 51:114–115. 1924;
 25. Glynne MD. Infection experiments with wart disease of potatoes. *Synchytrium endobioticum* (Schilb.) Perc. *Ann Appl Biol.* 1925;12(1):34–60.
 26. Lemmerz J. A new simplified infection procedure for testing potato cultivars for wart resistance. *Zücher.* 1930;2(288):97.
 27. CGN. Plant Genetic Resources. URL = <https://www.wur.nl/en/Research-Results/Statutory-research-tasks/Centre-for-Genetic-Resources-the-Netherlands-1/Expertise-areas/Plant-Genetic-Resources.htm>. (consulted on 2019.03.12) [Internet]. 2019. Available from: <https://www.wur.nl/en/Research-Results/Statutory-research-tasks/Centre-for-Genetic-Resources-the-Netherlands-1/Expertise-areas/Plant-Genetic-Resources.htm>
 28. VanRaden PM. Genomic measures of relationship and inbreeding. *INTERBULL Bull.* 2007;(37):33.
 29. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity.* 2005;95(3):221.
 30. Potato Genome Sequencing Consortium, Pan S, Cheng S, Zhang B, Mu D, Ni P, *et al.* Genome sequence and analysis of the tuber crop potato. *Nature.* 2011 Jul 10;475(7355):189–95.
 31. Fulton TM, Chunwongse J, Tanksley SD. Microprep protocol for extraction of DNA from tomato and other herbaceous plants. *Plant Mol Biol Report.* 1995;13(3):207–209.
 32. Andrew S. FastQC: a quality control tool for high throughput sequence data. Available from: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc> [Internet]. 2010. Available from: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
 33. Jacobs MM, van den Berg RG, Vleeshouwers VG, Visser M, Mank R, Sengers M, *et al.* AFLP analysis reveals a lack of phylogenetic structure within *Solanum* section Petota. *BMC Evol Biol.* 2008;8(1):145.

34. Maddison WP, Maddison DR. Mesquite: a modular system for evolutionary analysis [Internet]. 2018. Available from: <http://www.mesquiteproject.org>
35. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*. 2001 Aug 1;17(8):754–5.
36. Obidiegwu JE, Flath K, Gebhardt C. Managing potato wart: a review of present research status and future perspective. *Theor Appl Genet*. 2014 Apr;127(4):763–80.
37. Uitdewilligen JGAML, Wolters A-MA, D'hoop BB, Borm TJA, Visser RGF, van Eck HJ. A Next-Generation Sequencing Method for Genotyping-by-Sequencing of Highly Heterozygous Autotetraploid Potato. *Lukens L, editor. PLoS ONE*. 2013 May 8;8(5):e62355.
38. Jupe F, Witek K, Verweij W, Śliwka J, Pritchard L, Etherington GJ, *et al.* Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. *Plant J*. 2013 Nov;76(3):530–44.
39. Nazarian-Firouzabadi F, Joshi S, Xue H, Kushalappa AC. Genome-wide *in silico* identification of LysM-RLK genes in potato (*Solanum tuberosum* L.). *Mol Biol Rep*. 2019 Oct;46(5):5005–17.
40. van Berloo R, Hutten RCB, van Eck HJ, Visser RGF. An Online Potato Pedigree Database Resource. *Potato Res*. 2007 Jan;50(1):45–57.
41. Gebhardt C, Bellin D, Henselewski H, Lehmann W, Schwarzfischer J, Valkonen JPT. Marker-assisted combination of major genes for pathogen resistance in potato. *Theor Appl Genet*. 2006 May;112(8):1458–64.
42. EPPO. PM 7/28 (2) *Synchytrium endobioticum*. *EPPO Bull*. 2017 Dec;47(3):420–40.
43. EPPO. *Synchytrium endobioticum*: Diagnostic protocols. *EPPO Bull*. 2004 Aug;34(2):213–8.
44. Hawkes JG. The Potato, Evolution, Biodiversity and Genetic Resources. Belhaven Press, London; 1990.
45. Bukasov SM. Les Bases de la Sélection de la Pomme de terre (suite et fin). *Rev Bot Appliquée Agric Colon*. 1940;20(223):179–89.
46. Mattheij WM, Eijlander R, de Koning JRA, Louwes KM. Interspecific hybridization between the cultivated potato *Solanum tuberosum* subspecies *tuberosum* L. and the wild species *S. circaeifolium* subsp. *circaeifolium* Bitter exhibiting resistance to *Phytophthora infestans* (Mont.) de Bary and *Globodera pallida* (Stone) Behrens. 1992;83:459–66.
47. van Soest LJM, Seidewitz L. Evaluation data on tuber-bearing *Solanum* species. *Inst. für Pflanzenbau und Pflanzenzüchtung der FAL-Stichting voor Plantenveredeling (SVP)*. 1981. 165 p.
48. Gabriëls SHEJ, Vossen JH, Ekengren SK, Ooijen G van, Abd-El-Haliem AM, Berg GCM van den, *et al.* An NB-LRR protein required for HR signalling mediated by both extra- and intracellular resistance proteins: NRC1 is required for HR signaling. *Plant J*. 2007 Mar 5;50(1):14–28.
49. van de Vossen BTLH, Brankovics B, Nguyen HDT, van Gent-Pelzer MPE, Smith D, Dadej K, *et al.* The linear mitochondrial genome of the quarantine chytrid *Synchytrium endobioticum*; insights into the evolution and recent history of an obligate biotrophic plant pathogen. *BMC Evol Biol* [Internet]. 2018 Dec [cited 2019 Mar 5];18(1). Available from: <https://bmcevolbiol.biomedcentral.com/articles/10.1186/s12862-018-1246-6>

Supplementary figures

Supplementary figure 1. Distribution of P2, P6 and P18 resistance assessed with the Spieckermann method in the AxV population.

The resistance to the three pathotypes is skewed towards resistance. Quantitative scale: 10 = strongly resistant, 1 = strongly susceptible. The parents of the population are indicated.

Supplementary figure 2. Distribution of P2, P6 and P18 resistance assessed with the Glynne-Lemmerzahl method and of P6 and P18 with the Spieckermann method in the AxD population

Resistance to P2 showed a clear 1:1 bimodal distribution ($X^2 = 0.1$, p -value = 0.75) whereas resistance to P6 did not show a clear 1:1 segregation pattern and resistance to P18 was quantitative, skewed towards susceptibility. This was not surprising since the resistant parent Aventra did not show a strong resistance (score of 4.0) to P18 with the Glynne-Lemmerzahl method either. With the Spieckermann method, 93% of the AxD descendants showed a resistance score of 8 or higher, suggesting a high number of escapes due to low virulence of the P6 isolate and the resistant scores were disregarded for further analyses. Fortunately, bimodal distribution of the resistance was observed with the Spieckermann method for P18 ($X^2 = 1.52$, p -value = 0.217). Quantitative scale for the Glynne-Lemmerzahl scores: 1 = strongly resistant, 5 = strongly susceptible. Quantitative scale for the Spieckermann scores: 10 = strongly resistant, 1 = strongly susceptible. The parents of the population are indicated when phenotyped.

Supplementary figure 3. Distribution of P2, P6 and P18 resistance assessed with the Glynne-Lemmerzahl method and of P6 and P18 with the Spieckermann method in the KxA population

With the Glynne-Lemmerzahl method, resistance to P2 was bimodal without any extreme phenotype ($X^2 = 1.4$, p -value = 0.2367). Resistance to P6 and P18 was quantitative, skewed towards susceptibility. With the Spieckermann method, resistance to P6 was skewed towards resistance with 74.3% of the descendants scored with an 8 or higher. Resistance to P18 was bimodal ($X^2 = 0.029$, p -value = 0.87). Quantitative scale for the Glynne-Lemmerzahl scores: 1 = strongly resistant, 5 = strongly susceptible. Quantitative scale for the Spieckermann scores: 10 = strongly resistant, 1 = strongly susceptible. The parents of the population are indicated when phenotyped.

Supplementary figure 4. Effect of *Rse-Xlc-VTN62.33.3* and *Sen4* on resistance in the AxV population
Effect of *Rse-Xlc-VTN62.33.3* (PotVar0066337) and of *Sen4* (PotVar0037666) on **a** P2, **b** P6 and **c** P18 resistance in the AxV population.

Supplementary figure 5. Effect of *Rse-XIc-VTN62.33.3* and *Sen5* on resistance in the AxD population

Effect of *Rse-XIc-VTN62.33.3* (PotVar0067459) and of *Sen5* (presence of both flanking markers chr05_47980695 and chr05_52004982) on **a** P2, **b** P6, **c** P18 resistance assessed with the Glynne-Lemmerzähl method and on **d** P18 resistance assessed with the Spieckermann method in the AxD population.

Supplementary figure 6. Effect of *Rse-XIc-VTN62.33.3* and *Sen5* on resistance in the AxD population

Effect of *Rse-XIc-VTN62.33.3* (PotVar0067459) and of *Sen5* (presence of both flanking markers chr05_47980695 and chr05_52004982) on **a** P2, **b** P6, **c** P18 resistance assessed with the Glynne-Lemmerzahl method and on **d** P18 resistance assessed with the Spieckermann method in the KxA population.

Supplementary figure 7. Principal Component Analysis performed in SaKa1 on the resistance to P1, P2, P6 and P18

The two first principal components explained 83,5% of the variance. Three groups of descendants are visible: group 1 contains descendants resistant to P1, P2, P6 and P18, group 2 contains descendant resistant to P1 and group 3 contains descendant susceptible to all pathotypes.

Supplementary figure 8. Comparative Subsequence Sets Analysis in the SaKa1 population.

Aa and **Ba** Number of NLR genes per bin of 1 Mb from the potato reference genome DM v4.03 according to (38). **Ab** The SaKa1_RB bulk specific *k*-mers inherited from Andante mapped to the reference genome ($y_{\max} = 661\%$). **Ac** The SaKa1_RB bulk specific *k*-mers inherited from Alegria mapped to the reference genome ($y_{\max} = 545\%$). **Bb** The SaKa1_chr01_SB specific *k*-mers inherited from Andante mapped to the reference genome ($y_{\max} = 356\%$). **Bc** The SaKa1_chr01_SB specific *k*-mers inherited from Alegria mapped to the reference genome ($y_{\max} = 250\%$). **Bd** The SaKa1_chr01_RB specific *k*-mers inherited from Andante mapped to the reference genome ($y_{\max} = 321\%$). **Be** The SaKa1_chr01_RB specific *k*-mers inherited from Alegria mapped to the reference genome ($y_{\max} = 292\%$). For each chromosome, the average of the number of R-bulk specific *k*-mers mapped has been calculated and for each 1 Mb bin, the number of *k*-mers mapped from each set has been transformed to a percentage of the chromosome average.

Supplementary figure 9. Effect of the QTLs identified in SaKa1 on resistance

Effect of the QTLs identified in SaKa1 on **a** P1, **b** P2, **c** P6 and **d** P18 resistance. QTL1s = *Rse-Ib-Andante-a* in one copy (chr01_76425362 in simplex), QTL1d = chr01_76425362 in duplex, QTL2 = *Rse-XId-Andante* (chr11_11859618), QTL3 = *Rse-IXa-Andante* (chr09_61085336). y axis: 1 = resistant and 5 = susceptible.

Supplementary figure 10. Principal Component Analysis performed in the KxL population on P2, P6 and P18 resistance

The two first principal components explained 97.8% of the variance. The descendants could be divided in three groups: group 1 was composed of descendants resistant to P2, P6 and P18, group 2 was composed of descendants resistant to P2 and P6 but slightly susceptible to P18 and group 3 was composed of descendants susceptible to the four pathotypes.

Supplementary figure 11. Comparative Subsequence Sets Analysis performed in the KxL population

a Number of NLR genes per bin of 1 Mb from the potato reference genome DM v4.03 according to (38). **b** KxL_RB specific *k*-mers inherited from Kuba mapped to the reference genome ($y_{\max} = 420\%$). **c** KxL_RB specific *k*-mers inherited from Ludmilla mapped to the reference genome ($y_{\max} = 452\%$). **d** KxL_RB specific *k*-mers inherited from both parents mapped to the reference genome ($y_{\max} = 444\%$). An arrow shows *Rse-VIIIb-Kuba* which is linked with P18 resistance.

Supplementary figure 12. Effect of *Sen3* and *Rse-VIIIb-Kuba* on resistance in KxL

Effect of *Sen3* and of *Rse-VIIIb-Kuba* (chr08_44797542 and chr08_45178832) on P18 resistance in the KxL population.

Supplementary figure 13. Gel picture of the digestion product of the CAPS marker linked with *Sen2* for several varieties and accessions

The positive control (DG 97-264) is indicated by the mean of stars. To score *Sen2* as being present or absent in the variety and Solanum panels, we looked at the presence of the 250 bp band (arrow), its intensity compared to DG 97-264 considering the ploidy level of the genotype, and the intensity of the smaller bands from DG 97-264. If the 1046 bp fragment is not amplified, or the 250 bp fragment (arrow) is absent, the genotype is scored negative for the marker. The genotypes scored as positive for *Sen2* are indicated.

Supplementary tables

Supplementary table 1. Summary of the phenotypic assays performed in this study

Popula- tion	Method	Year	Place	Patho- types	Isolates	# genotypes assessed	# tubers
AxV	Spieck-ermann	2015	HLB BV	P2	HLB P2 (G1)-03-08	100	8
				P6	HLB 6 (O1) 10-11	100	8
				P18	HLB 18 (T1) 05-08	100	8
		2016	Averis seeds BV	P2	HLB P2 (G1)-01-05	45	15
				P6	HLB P6 (O1) 07-09 + HLB 14-13	75	15
				P18	HLB P18 (T1) 09-11 + HLB 04-07	75	15
		2017		P2	HLB P2 (G1)-03-08	28	15
				P6	HLB 6 (O1) 10-11 + MB57	2	30
				P18	HLB 18 (T1) 09-11 + HLB05-08	2	30
AxD & KxA	Glynn- Lemmerz zahl	2015	IHAR- PIB	P6	JKI P6(O1)-2009	77	3
				P18	JKI P18(T1)-2009	77	3
		2016		P2	JKI P2(G1)-2009	75	6
				P6	JKI P6(O1)-2009	46	6
		2017		P2	JKI P2(G1)-2009	27	6
				P6	JKI P6(O1)-2009	16	6
	Spieck- ermann	2016	HLB BV	P6*	HLB 6 (O1) 10-11*	77*	8*
				P18	HLB 18 (T1) 05-08	77	8
		2017		P6*	HLB P6 (O1) 10-11*	58*	12*
				P18	HLB P18 (T1) 05-08	27	12

* The results from these phenotypic tests were not considered for further analyses as the infection level was very low and led to a high number of escapes.

Supplementary table 2. Validation of the GWAS and CoSSA SNPs in the AxV population

Marker	Chr	Position	Expect ed ratio ^a	X2 p- value ^b	Spieck P2 ^c	Spieck P6 ^c	Spieck P18 ^c
PotVar001925	3	2,233,03	1:1	1.00	0.52	0.93	0.89
PotVar001950	3	2,915,21	1:1	1.00	0.52	0.93	0.89
PotVar001955	3	3,006,81	1:1	0.76	0.76	0.96	0.86
PotVar010642	3	14,144,6	1:1	0.68	0.83	0.52	0.56
PotVar003959	4	8,362,27	1:4:1	0.00	0.63	0.21	0.36
PotVar002605	5	3,876,48	1:1	0.10	0.42	0.09	0.57
PotVar008381	5	7,541,45	1:1	0.04	0.83	0.24	0.54
PotVar008417	5	7,670,81	1:1	0.18	0.24	0.86	0.18
PotVar012581	5	8,133,20	1:1	0.10	0.71	0.27	0.72
PotVar008553	5	8,807,47	1:1	0.04	0.36	0.26	0.99
PotVar008540	5	8,808,87	1:1	0.01	0.71	0.45	0.40
PotVar008316	6	378,722	1:1	0.92	0.49	0.06	0.46
PotVar008316	6	378,767	1:1	0.92	0.40	0.02	0.21

PotVar008288	6	388,721	1:1	0.84	0.44	0.05	0.40
PotVar002703	6	3,213,69	1:1	0.84	0.83	0.18	0.61
PotVar006948	6	29,621,9	1:1	0.30	0.47	0.34	0.75
PotVar006590	6	58,462,3	1:1	0.30	0.51	0.94	0.91
PotVar011537	7	44,691,9	1:1	0.47	0.67	0.12	0.54
PotVar009367	7	45,134,8	1:1	0.47	0.67	0.12	0.54
PotVar013408	7	49,456,6	1:1	1.00	0.69	0.53	0.62
PotVar012542	8	45,873,4	1:1	0.41	0.93	0.49	0.37
PotVar010728	9	11,227,0	1:1	0.29	0.71	0.83	0.26
PotVar006451	11	808,641	1:1	0.47	1.26E-02	0.33	0.42
PotVar006462	11	810,912	1:1	0.47	1.26E-02	0.33	0.42
PotVar006477	11	812,865	1:1	0.54	8.24E-03	0.25	0.47
PotVar006478	11	813,322	1:1	0.54	8.24E-03	0.25	0.47
PotVar006621	11	2,037,45	1:1	0.14	6.67E-04	3.33E-02	0.92
PotVar006633	11	2,261,65	1:1	0.22	7.48E-04	9.42E-03	0.62
PotVar006701	11	2,773,88	1:1	0.18	1.42E-03	2.23E-02	0.69
PotVar006745	11	3,073,73	1:1	0.06	6.21E-03	1.31E-02	0.60
PotVar011043	11	3,658,07	1:1	0.06	6.21E-03	1.31E-02	0.60
PotVar010605	11	4,346,44	1:1	0.10	1.05E-02	1.95E-02	0.67
PotVar010601	11	4,347,63	1:1	0.10	1.05E-02	1.95E-02	0.67
PotVar005928	11	8,816,27	1:4:1	0.92	4.40E-03	0.10	6.48E-02
PotVar003191	12	43,326,4	1:1	0.17	1.61E-02	6.73E-11	5.73E-12
PotVar003648	12	45,970,9	1:1	0.18	5.83E-04	3.72E-14	1.19E-13
PotVar003632	12	45,974,0	1:1	0.18	9.51E-03	6.98E-11	7.04E-12
chr12_485014	12	48,501,4	1:1	0.20	1.69E-03	2.80E-13	6.35E-13
PotVar003768	12	50,379,0	1:1	0.04	2.01E-02	3.54E-13	1.26E-13
PotVar003766	12	50,379,3	1:4:1	0.00	9.85E-04	2.14E-14	1.65E-15
PotVar003740	12	50,392,2	1:1	0.08	1.58E-03	1.98E-14	7.02E-15
chr12_514990	12	51,499,0	1:1	0.09	1.13E-03	9.37E-14	9.91E-15

^a Expected segregation ratio of the marker minor allele in the offspring in function of the observed dosage in the population parents, ^b p -value of the X^2 test performed to validate the goodness of fit of the segregation ratio of the tested markers with the expected segregating ratio, ^c p -values of the association tests performed between the marker results and the phenotypic data (Kruskal-Wallis tests). The average value of the 2015, 2016 and 2017 phenotypic assays was used for the association tests.

Supplementary table 3. Validation of the GWAS and CoSSA SNPs in the AxD population

Marker	Chr	Position	Expected ratio ^a	X ² p-value ^b	GL P1 ^c	GL P2 ^c	GL P6 ^c	GL P18 ^c	Spieck P1 ^c	Spieck P18 ^c
PotVar0039597	4	8,362,270	1:1	0.03	0.46	0.30	0.74	0.97	0.76	0.44
PotVar0026051	5	3,876,481	1:1	1.00	0.48	0.81	0.32	0.48	0.44	0.81
PotVar0083817	5	7,541,453	1:1	0.35	0.60	0.50	0.80	0.52	0.79	0.92
PotVar0084178	5	7,670,810	1:1	0.54	0.53	0.79	0.15	0.88	0.66	0.72
PotVar0125819	5	8,133,203	1:1	0.64	0.87	0.92	0.19	0.78	0.81	0.95
PotVar0085531	5	8,807,478	1:1	0.26	0.42	0.51	0.12	0.56	0.20	0.98
PotVar0085405	5	8,808,875	1:1	0.19	0.56	0.78	0.26	0.96	0.24	0.96
chr05_47980695	5	47,980,695	1:1	0.34	0.17	8.01E-03	1.40E-02	5.40E-02	0.42	1.72E-05
PotVar0081757	5	48,568,431	1:1	1.00	0.11	1.41E-02	3.97E-02	6.70E-02	0.11	2.62E-04
PotVar0082051	5	48,659,924	1:1	0.35	0.08	3.37E-03	8.32E-03	2.59E-02	0.27	7.14E-06
PotVar0123123	5	49,203,506	1:1	0.35	0.08	3.37E-03	8.32E-03	2.59E-02	0.27	7.14E-06
PotVar0127943	5	49,733,240	1:1	0.35	0.08	3.37E-03	8.32E-03	2.59E-02	0.27	7.14E-06
PotVar0035046	5	50,250,772	1:1	0.35	0.08	3.37E-03	8.32E-03	2.59E-02	0.27	7.14E-06
PotVar0035016	5	50,251,077	1:1	0.35	0.08	3.37E-03	8.32E-03	2.59E-02	0.27	7.14E-06
PotVar0034831	5	51,163,965	1:1	0.54	0.03	1.48E-03	4.35E-03	6.26E-03	0.20	6.74E-06
PotVar0034727	5	51,320,561	1:1	0.35	0.02	1.42E-03	3.65E-03	1.24E-02	0.08	2.86E-05
chr05_52004982	5	52,004,982	1:1	0.35	0.02	1.42E-03	3.65E-03	1.24E-02	0.08	2.86E-05
PotVar0083164	6	378,767	1:4:1	0.00	0.92	0.74	0.40	0.98	0.78	0.49
PotVar0027035	6	3,213,693	1:4:1	0.00	0.69	0.99	0.99	0.50	0.43	0.85
PotVar0069488	6	29,621,950	1:1	0.22	0.86	0.38	0.92	0.70	0.92	0.07
PotVar0065903	6	58,462,327	1:1	0.27	0.99	0.62	0.82	0.24	0.67	0.21
PotVar0107285	9	11,227,013	1:1	1.00	0.66	0.96	0.64	0.69	0.68	0.07
PotVar0005344	10	55,856,581	1:1	0.88	0.41	0.78	0.10	0.35	0.14	0.17
PotVar0005531	10	56,583,884	1:1	0.76	0.68	0.66	0.07	0.35	0.12	0.17
PotVar0057428	10	58,609,544	1:1	0.87	0.29	0.15	0.14	0.31	0.01	0.29
PotVar0122826	10	59,471,325	1:1	0.88	0.30	0.16	0.23	0.24	0.01	0.54
PotVar0122661	10	59,673,263	1:1	0.88	0.33	0.19	0.33	0.25	0.04	0.40
PotVar0064625	11	810,912	1:4:1	0.69	0.09	0.11	0.04	0.14	0.97	0.13

PotVar0064771	11	812,865	1:4:1	0.91	0.10	0.05	0.01	0.05	0.88	0.05
PotVar0064787	11	813,322	1:4:1	0.87	0.11	0.06	0.01	0.04	0.88	0.06
PotVar0066337	11	2,261,657	1:4:1	1.00	0.10	0.03	0.02	0.04	0.80	0.07
PotVar0067017	11	2,773,883	1:1	0.76	0.54	0.74	0.02	0.02	0.44	0.12
PotVar0067459	11	3,073,732	1:1	0.76	0.54	0.82	2.64E-03	1.78E-03	0.44	0.04
PotVar0106057	11	4,346,448	1:1	0.76	0.54	0.82	2.64E-03	1.78E-03	0.44	0.04
PotVar0106019	11	4,347,636	1:1	0.76	0.54	0.82	2.64E-03	1.78E-03	0.44	0.04

^a Expected segregation ratio of the marker minor allele in the offspring in function of the observed dosage in the population parents, ^b *p*-value of the χ^2 test performed to validate the goodness of fit of the segregation ratio of the tested markers with the expected segregating ratio, ^c *p*-values of the association tests performed between the marker results and the phenotypic data (Kruskal-Wallis tests). If several years of phenotyping were performed, the average value across years was used for the association tests.

Supplementary table 4. Validation of the GWAS and CoSSA SNPs in the KxA population

Marker	Chr	Position	Expected ratio ^a	χ^2 <i>p</i> -value ^b	GL P1 ^c	GL P2 ^c	GL P6 ^c	GL P18 ^c	Spieck P1 ^c	Spieck P18 ^c
PotVar0039597	4	8,362,270	1:1	0.29	0.84	0.66	0.47	0.78	0.96	0.16
PotVar0026051	5	3,876,481	1:1	0.17	0.89	0.97	0.37	0.03	0.41	0.50
PotVar0083817	5	7,541,453	1:1	0.09	0.56	1.00	0.19	0.02	0.86	0.15
PotVar0084178	5	7,670,810	1:1	0.24	0.88	0.69	0.20	0.01	0.25	0.20
PotVar0125819	5	8,133,203	1:1	0.30	0.86	0.48	0.05	0.00	0.60	0.05
PotVar0085531	5	8,807,478	1:1	0.38	0.60	0.66	0.12	0.02	0.17	0.15
PotVar0085405	5	8,808,875	1:1	0.21	0.54	0.67	0.16	0.02	0.20	0.15
chr05_47980695	5	47,980,695	1:1	0.49	0.31	4.99E-04	3.49E-05	3.56E-04	0.18	2.45E-06
PotVar0081757	5	48,568,431	1:1	0.49	0.31	4.99E-04	3.49E-05	3.56E-04	0.18	2.45E-06
PotVar0082051	5	48,659,924	1:1	0.49	0.31	4.99E-04	3.49E-05	3.56E-04	0.18	2.45E-06
PotVar0123123	5	49,203,506	1:1	0.49	0.31	4.99E-04	3.49E-05	3.56E-04	0.18	2.45E-06
PotVar0127943	5	49,733,240	1:1	0.38	0.16	5.48E-04	4.29E-05	9.15E-04	0.12	2.45E-06
PotVar0035046	5	50,250,772	1:1	0.30	0.15	6.16E-04	5.87E-05	8.57E-04	0.09	1.83E-06
PotVar0035016	5	50,251,077	1:1	0.30	0.15	6.16E-04	5.87E-05	8.57E-04	0.09	1.83E-06
PotVar0034831	5	51,163,965	1:1	0.49	0.11	2.92E-03	1.72E-04	3.19E-03	0.05	2.91E-06
PotVar0034727	5	51,320,561	1:1	0.49	0.11	2.92E-03	1.72E-04	3.19E-03	0.05	2.91E-06

chr05_52004982	5	52,004,982	1:1	0.49	0.11	2.92E-03	1.72E-04	3.19E-03	0.05	2.91E-06
PotVar0083164	6	378,767	1:4:1	0.00	0.60	0.08	0.10	0.28	0.66	0.36
PotVar0027035	6	3,213,693	1:4:1	0.04	0.67	0.48	0.17	0.87	1.00	0.43
PotVar0069488	6	29,621,950	1:1	0.17	0.68	0.20	0.63	0.23	0.91	0.28
PotVar0065903	6	58,462,327	1:1	0.16	0.67	0.62	0.16	0.24	0.20	0.24
PotVar0107285	9	11,227,013	1:1	0.09	0.26	0.26	0.29	0.34	0.49	0.68
PotVar0005344	10	55,856,581	1:1	0.87	0.55	0.65	0.73	0.17	0.99	0.97
PotVar0005531	10	56,583,884	1:1	0.49	0.66	0.33	0.64	0.18	0.83	0.94
PotVar0057428	10	58,609,544	1:1	0.60	0.25	0.34	0.81	0.54	0.21	0.83
PotVar0122826	10	59,471,325	1:1	0.48	0.19	0.34	0.79	0.57	0.17	0.79
PotVar0122661	10	59,673,263	1:1	0.73	0.32	0.31	0.80	0.51	0.27	1.00
PotVar0064625	11	810,912	1:4:1	0.56	0.62	9.43E-03	3.95E-02	0.54	0.29	1.03E-01
PotVar0064771	11	812,865	1:4:1	0.31	0.40	1.47E-02	8.44E-02	0.55	0.15	2.46E-01
PotVar0064787	11	813,322	1:4:1	0.03	0.17	2.62E-02	6.48E-02	0.46	0.02	2.41E-01
PotVar0066337	11	2,261,657	1:4:1	0.42	0.60	2.72E-02	8.56E-02	0.77	0.09	1.95E-01
PotVar0067017	11	2,773,883	1:1	0.49	0.17	1.73E-03	5.74E-04	0.03	0.40	8.97E-04
PotVar0067459	11	3,073,732	1:1	0.73	0.08	6.91E-03	1.37E-03	0.11	0.14	1.29E-03
PotVar0106057	11	4,346,448	1:1	1.00	0.11	1.22E-02	2.97E-03	0.12	0.24	3.33E-03
PotVar0106019	11	4,347,636	1:1	0.86	0.14	5.09E-03	2.02E-03	0.06	0.37	2.56E-03

^a Expected segregation ratio of the marker minor allele in the offspring in function of the observed dosage in the population parents, ^b p -value of the χ^2 test performed to validate the goodness of fit of the segregation ratio of the tested markers with the expected segregating ratio, ^c p -values of the association tests performed between the marker results and the phenotypic data (Kruskal-Wallis tests). If several years of phenotyping were performed, the average value across years was used for the association tests.

Supplementary table 5. Validation of the CoSSA SNPs in the SaKa1 population

Marker	Chr/haplo-type	Position	Experiment ^a	Parental origin	Expected ratio ^b	X ² p-value ^c	P1	P2	P6	P18
Sen1	chr11_130892	1,308,92	1	Both	1:2:1	0.77	2.11E-14	0.95	0.47	0.92
	chr11_166635	1,666,35	1	Both	1:2:1	0.22	6.47E-13	0.54	0.50	0.98
	chr01_700666	70,066,6	2	Andante	1:1	0.06	4.75E-02	0.05	0.04	0.06
	chr01_727740	72,774,0	3	Andante	1:1	0.05	0.03	7.55E-06	9.24E-06	1.89E-06
Rse-Ib-Andante	chr01_735270	73,527,0	2	Andante	1:1	0.17	0.07	0.03	3.46E-02	4.49E-02
	chr01_741485	74,148,5	3	Andante	1:1	0.14	2.77E-03	5.74E-06	4.69E-07	2.17E-08
	chr01_741626	74,162,6	3	Andante	1:1	0.26	2.88E-03	1.21E-04	2.48E-06	3.54E-07
	chr01_764253	76,425,3	3	Andante	1:4:1	0.04	5.22E-04	3.92E-07	2.02E-10	9.42E-10
	chr01_773299	77,329,9	4	Andante	1:1	0.19	0.02*	1.46E-04*	0.02*	0.08
	chr01_777502	77,750,2	3	Andante	1:1	0.36	5.75E-03	2.52E-03	8.74E-05	3.37E-04
	chr01_778012	77,801,2	3	Andante	1:1	0.03	3.29E-03	2.30E-05	2.80E-05	2.76E-06
	chr01_790268	79,026,8	3	Andante	1:1	0.06	3.68E-03	2.95E-05	1.10E-05	1.20E-05
Rse-IXa-Andante	chr01_796946	79,694,6	3	Andante	1:1	0.31	3.37E-03	5.63E-04	3.14E-05	2.09E-04
	chr01_797515	79,751,5	4	Andante	1:1	0.07	0.04*	7.18E-05*	0.01*	0.05
	chr09_526587	52,658,7	3	Andante	1:1	0.22	0.40	0.58	0.76	8.4E-03
	chr09_550284	55,028,4	3	Andante	1:1	0.37	0.49	0.17	0.40	2.7E-04
Rse-IXd-Andante	chr09_610853	61,085,3	3	Andante	1:1	0.59	0.61	0.11	0.38	1.5E-05
	chr09_614824	61,482,4	3	Andante	1:1	0.72	0.76	0.23	0.71	2.5E-04
	chr09_614838	61,483,8	3	Andante	1:1	0.52	0.44	0.38	0.65	2.0E-04
	chr11_916612	9,166,12	3	Andante	1:1	0.31	0.80	0.06	0.15	0.06
Rse-IXd-Andante	chr11_118596	11,859,6	3	Andante	1:1	0.11	0.28	0.03	0.05	0.02
	chr11_299765	29,976,5	3	Andante	1:1	0.05	0.47	0.04	0.09	0.03
	chr12_572408	57,240,8	3	Alegria	1:1	0.54	0.23	0.79	0.93	0.32
	chr12_598390	59,839,0	3	Alegria	1:1	0.19	0.29	0.03	0.11	0.38

^a 1: (8), 2: SaKa1_chr01_RB minus SaKa1_chr01_SB, 3: SaKa1_RB minus SaKa1_P1SB, 4: SaKa1_chr01_SB minus SaKa1_chr01_RB, ^b Expected segregation ratio of the marker minor allele in the offspring in function of the observed dosage in the sequencing data of the parents, ^c p-value of the X² test performed to validate the goodness of fit of the segregation ratio of the tested markers with the expected segregating ratio, ^c p-values of the association tests performed between the marker results and the phenotypic data (Kruskal-Wallis tests). * Significant markers with a negative effect on resistance (associated with susceptibility).

Supplementary table 6. Validation of the CoSSA SNPs in the KxL population

Marker	Chromosome	Position	Parental origin	Expected ratio ^a	χ^2 <i>p</i> -value ^b	GL P2 ^c	GL P6 ^c	GL P18 ^c
chr01_8093926	1	80,939,26	Ludmilla	1:4:1	8.4E-47	0.41	0.21	0.56
chr02_3031147	2	30,311,47	Ludmilla	1:1	2.3E-03	0.91	0.85	0.56
chr02_3241076	2	32,410,76	Ludmilla	1:1	0.20	0.83	0.80	0.61
chr03_5806659	3	58,066,59	Ludmilla	1:1	0.82	0.22	0.22	0.22
chr03_5807684	3	58,076,84	Ludmilla	1:1	0.66	0.54	0.67	0.95
chr05_8126924	5	8,126,924	Kuba	1:1	0.20	0.36	0.43	0.89
chr06_2500048	6	25,000,48	Kuba	1:1	0.22	0.16	0.81	0.37
chr07_1021494	7	10,214,94	Ludmilla	1:1	0.37	0.70	0.89	0.22
chr07_4108450	7	41,084,50	Ludmilla	1:1	0.25	0.91	0.48	0.37
chr07_4523314	7	45,233,14	Ludmilla	1:1	0.20	0.93	0.47	0.47
chr08_4479754	8	44,797,54	Kuba	1:1	2.4E-03	0.18	0.54	0.03
chr08_4517883	8	45,178,83	Kuba	1:1	1.9E-03	0.33	0.85	0.05

^a Expected segregation ratio of the marker minor allele in the offspring in function of the observed dosage in the sequencing data of the parents, ^b *p*-value of the χ^2 test performed to validate the goodness of fit of the segregation ratio of the tested markers with the expected segregating ratio, ^c *p*-values of the association tests performed between the marker results and the phenotypic data (Kruskal-Wallis tests). If several years of phenotyping were performed, the average value across years was used for the association tests.

Supplementary files list

Supplementary files can be accessed at <https://doi.org/10.17026/dans-2zf-pnd3>

> Supplementary file 1 (Chapter5_S_file_1.xlsx)

Overview of the historical sources that provided phenotypic data on wart disease resistance in the potato varieties used in our GWAS panels. These sources include documents for National listing where VCU data were recorded, marketing brochures from commercial breeders as well as scientific publications. Several sources use different scales to describe the level of wart disease resistance. Quantitative scales were harmonised to a common scale ranging from 1 = susceptible to 10 = resistant. Qualitative data R and S were converted into 10 and 1 respectively. This allowed compilation of the second tab sheet showing all records used to calculate a single phenotypic score using REML. The third, fourth and fifth tab sheets contain the genotypic data of the P2, P6 and P18 GWAS panels.

> Supplementary file 2 (Chapter5_S_file_2.xlsx)

Phenotypic and genotypic data of the five full-sib populations used in this study. Each Excel sheet comprises the phenotypic data of all the tests performed, the genotypic data of all the markers screened and information about the bulks composition.

> Supplementary file 3 (Chapter5_S_file_3.xlsx)

Composition on the variety and the *Solanum* panels. For each variety/accession, the species, the ploidy, the year and the origin are given. For the accessions, the series (according to (44)) are also given. The genotypic data of the markers screened in the two panels are also given. When resistance to P1, P2, P6, P8 or P18 is known, it is indicated (R = resistant, M = medium, S = susceptible).

> Supplementary file 4 (Chapter5_S_file_4.xlsx)

Are given in this file: the sequencing depth of all the sequenced samples in this study, a description of the different CoSSA workflows and a summary of the CoSSA *k*-mers subsets for all the workflows.

> Supplementary file 5 (Chapter5_S_file_5.xlsx)

Screening of the potato wart disease *Sen* genes specific *k*-mers in the panel from (23) using CoSSA. ^a Proportion of the *Sen* genes specific *k*-mers shared between the positive control and the variety/accession from (23). The positive controls for *Sen1*, *Sen3*, *Sen4* and *Sen5* are respectively Desiree and Kuras, Kuba, Axion and Avenra. In bold are indicated the accessions that we consider containing the *Sen* gene according to this table and the extra sheets.

Extra-sheets: *K*-mers specific to each of the four *Sen* genes (*Sen1*, *Sen3*, *Sen4* and *Sen5*) present in the varieties and accessions of (23) mapped to the potato reference genome. The chromosomes were cut in bins of 100 Kb and the number of *Sen* specific *k*-mers per bin was counted. The varieties positive for a specific *Sen* gene are highlighted in green. The (fine)-mapped region of each *Sen* gene is highlighted in yellow.

Chapter 6

General discussion

Phytopathologists and breeders aiming at disease resistance in their crops will agree that “betting against the pathogen” is an error not to be committed while studying plant-pathogen interactions and breeding for resistance. Knowledge about the pathogen infection process and about the host defence systems is the best weapon we have against plant pathogens. My thesis is part of a larger project which main objectives were the identification of *S. endobioticum* Avr genes and the identification of the matching potato *R* genes, responsible for pathotype specific wart resistance. In this thesis, we set out to identify the different wart disease resistance sources present in the potato breeding gene pool by means of diversity panels and bi-parental mapping populations. We implemented a novel approach, Comparative Subsequence Sets Analysis (CoSSA), to identify haplotype specific variants and to design diagnostic markers for marker assisted selection (MAS), an approach which can be applied to any crop and any trait. CoSSA is of special value for heterozygous polyploids for which it would otherwise be difficult to select for the right haplotype. We used CoSSA to map and fine-map several major and minor effect resistance loci in different genetic backgrounds. Markers diagnostic for resistance were designed and can be directly integrated in potato breeding programs. We traced back the identified resistance loci in pedigrees and distinguished different resistance sources. The knowledge acquired in this thesis is essential for a better understanding of the mechanisms behind the pathotype specific resistance, for the breeding of more efficient and durable resistances, and for an improved management of the disease.

Sources of potato wart disease resistance

In **chapters 2, 3 and 5**, we set out to identify the different potato wart disease resistance sources present in the breeding genepool. Thanks to the screening of variety panels, we could identify the main resistance sources present in the Dutch, German and Polish material. Knowing which varieties carry which resistance gene(s) is crucial for breeders as it will help them making choices for their crosses and avoiding redundancy. Without knowing which *S. endobioticum* *R* genes are present in which variety, breeders can only rely on pedigree information to try to combine different resistant sources in their breeding material.

Potato wart disease resistance sources in the breeding genepool

In my thesis, several potato wart disease resistance loci were identified in commercial varieties and breeding clones. In **chapter 2**, we observed that *Sen1*, first identified by (1) in a diploid population, was present in 56% of the GWAS panel varieties and in

chapter 5, it was observed in 69% of the variety panel. Overall, in these two panels, we observed that *Sen1* was present in recent and old varieties, present in the breeding programs of many countries (in Europe and America) and present in varieties bred for different potato markets. *Sen1* is, therefore, the main source of pathotype 1 (P1) resistance in the potato breeding genepool. It is, however, not the only source of P1 resistance. Previously, a pathotype 1 resistance locus was identified on the south arm of chromosome 4 (2), but we did not re-identify it with the GWAS, probably because it is a rare source of P1 resistance. Other loci with a broader resistance pattern, such as *Sen2* (3), *Sen3* (4) and *Sen4* (**chapter 5**), might represent alternative sources of P1 resistance. In **chapter 2**, we hypothesised that another source of P1 resistance might be present in the structure group composed of varieties with British origins. Indeed, 27 varieties from this group were resistant to P1 but did not contain the *Sen1* associated GWAS marker PotVar0067008. The design of the new *Sen1* flanking markers in **chapter 4** and the variety panel screening in **chapter 5** showed that five of these varieties hold *Sen1*. Analysis of the 22 remaining varieties may reveal if there are one or more additional sources of P1 resistance in this material.

Resistance to the higher pathotypes is less frequent in the breeding material. For instance, in the GWAS panel of **chapter 5**, we could find only 15 genotypes out of the 569 genotypes from (5) which were referenced as resistant to P18. Eight of these varieties get their resistance from *Sen4*, one from *Sen3* and one from *Sen5*. *Sen4* is the main source of resistance in the Dutch breeding material (Figure 1). Most of the Dutch varieties holding *Sen4* are varieties for starch production. Resistance to the higher pathotypes of *S. endobioticum* is required for Dutch starch varieties as they are grown in areas under quarantine surveillance. In the Dutch breeding material, we could trace *Sen4* back to AM78-3704 and Alcmaria as a source. In the German material, we found *Sen4* in Panda and in Hindenburg, which is in the pedigree of Panda. We hypothesized that the German source of *Sen4* is different from the Dutch source. Indeed, Alcmaria and Hindenburg share only one ancestor in common far in their pedigree: Erste von Fromsdorf, an old (1884) German variety (6) which was intensively used in breeding and is, therefore, unlikely a source of wart disease resistance. If a sample mistake for Hindenburg could be ruled out, we could make the hypothesis that the German *Sen4* source is Jubel, the male parent of Hindenburg, which was reported as being resistant to wart disease and intensively used in German breeding programs to create wart resistant varieties (7,8). It is not known, however, if Jubel and Hindenburg are resistant to the higher pathotypes, so these hypotheses have to be taken carefully and should be validated with phenotypic and genotypic assays. The

main source of resistance to wart disease of the German and Polish varieties is *Sen3* (Figure 1), which could be traced back to the variety Ora in **chapter 3**. The clone BRA9089, the male parent of Ora (6), was reported by different authors as being resistant to the higher pathotypes and used to breed resistant varieties (10–13).

Figure 1. Putative sources of potato wart disease resistance in the potato breeding gene pool

The map shows the European countries in which *S. endobioticum* is present (orange). Yellow dots = present, pink dot in Norway = transient. Adapted from (9) (updated 02-09-2019).

Despite the fact that we found the genotype labelled today as being BRA9089 susceptible to pathotypes 1, 2, 6 and 18 (P1, P2, P6 and P18) in **chapter 3**, I still suspect the old BRA9089 to be the ancestor donor of *Sen3*. Surprisingly, although *S. endobioticum* is not an issue in the USA, we found *Sen3* in two American varieties (Defender and Goldrush), both holding BRA9089 in their pedigree. Another source of resistance to the higher pathotypes in the Dutch breeding material is *Sen5* which was traced back to the breeding clone VE71-105. This gene is very rare in the breeding gene pool as we found it only in Aventra, Kanjer and VE71-105. The *Sen2* gene

identified by (3) has not been introduced yet in the potato breeding genepool. To conclude, we identified different resistance sources present in the breeding gene pool. This information is highly valuable for breeders to select non-overlapping sources of resistance as well as to combine multiple sources to broaden the resistance spectrum, and hence, to enhance the resistance levels against the complex higher pathotypes.

Origin of potato wart disease resistance sources

Because of the high frequency of *Sen1* in the potato breeding genepool, it is not possible to identify a common ancestor that introduced *Sen1* in the pedigrees of *Sen1* plants. In a previous study (14), 52 accessions from *S. phureja*, *S. stenotomum*, *S. tuberosum* ssp. *andigenum* and *S. tuberosum* ssp. *tuberosum* were screened with the *Sen1* linked marker NI25 (1). No association was observed of the marker with taxonomy or geography. The two alleles of NI25 were found in susceptible as well as resistant accessions which shows the marker was not diagnostic in this material. Resistance to P1 was found in all the species tested which were all cultivated species. A broad panel of tuber-bearing *Solanum* species was phenotyped with P1 by (15) and resistance was found in more than 18 different species. The abundance of *Sen1* can explain why we could not find proof for identity-by-descent (IBD) in the commercial varieties. Also in our work, we found that *Sen1* was already widely spread in the cultivated potato ancestors. In **chapter 5**, we found *Sen1* in one landrace from *S. tuberosum* Chilotanum group (16) and in several accessions mainly from Argentina and Bolivia, including five accessions from *S. microdontum*. This is interesting as group Chilotanum is the ancestor of cultivated potato and *S. microdontum* was intensively used to adapt cultivated potatoes to long days conditions (16). We can hypothesize that *Sen1* was introgressed “by chance” multiple times in the cultivated potato ancestors, which explains its presence already in very old material and why it is not possible to trace it back to a common ancestor.

The introgression of the genes bringing resistance to the higher pathotypes in the breeding gene pool may have been deliberate, or a lucky coincidence that happened during the introgression of resistances to other pathogens, or other traits like long day adaptation. It is not clear yet from which species in its pedigree DG 97-264 inherited *Sen2* (3). According to the results obtained in **chapter 5**, I think *Sen2* might be present in the Acaulia, Demissa and Tuberosa series (17). The screening of *Sen3* in the accession panels in **chapter 5** did not give clear insights about how *Sen3* was introgressed. *Sen3* is IBD in all the varieties and its source is probably a *S. tuberosum* landrace from the Chiloe island (ancestor of BRA9089) (6,11). *Sen4*, IBD in all the

Dutch starch varieties, was found in four accessions from *S. tuberosum* ssp. *andigenum*. The Dutch *Sen4* could be traced back to Alcmaria which comes from a cross between Sirtema and a descendent from Saskia and CPC1673-20 (adg) (6). We showed that Sirtema does not hold *Sen4* so CPC1673-20 is likely the *Sen4* donor. Resistance has been reported several times in the Andigenum group (10,11,15,18) and wart resistance was observed in the clone CPC1673 which was used to introgress resistance to nematodes (11). Several other varieties resistant to higher pathotypes possess andigenum and CPC1673 background, i.e. Amalia (adg, CPC1673-24), Gelda, Maja and Titana (11). It is not known if these varieties got their resistance through *Sen4* or another andigenum resistance locus. *Sen5* could be unambiguously traced back to VE71-105 and *S. vernei* which was already reported before as being a potential source of wart disease resistance (18). Among the five *S. vernei* accessions from our **chapter 5** panel which were positive for the flanking marker of *Sen5*, two showed a certain level of resistance to P2 and three to P18.

Unknown resistance sources in the breeding genepool

The GWAS panel from **chapter 2** and the variety panel from **chapter 5** revealed the presence of additional unknown resistance sources already present in the breeding material. For some of these varieties, intermediate or weak resistance can be due to one of the QTLs that were identified in **chapters 3 and 5** or from previous studies (19–21). For the stronger resistances for which the resistance source is still unknown, I performed a pedigree analysis (6) of the varieties and provided an overview of them in Table 1. The species *S. tuberosum* ssp. *andigena*, *S. demissum*, *S. edinense*, *S. oplocense*, *S. phureja*, *S. tuberosum* and *S. vernei* contributed greatly to the genetic background of the varieties from Table 1. As mentioned previously, resistance to the higher pathotypes was already observed in the Andigena group and in *S. vernei*.

As being part of the EPPO set, Belita harbours a strong resistance to pathotypes 1, 2, 6 and 18 but does not contain any of the known *Sen* genes. The resistance from Belita and Plasettie (descendant of Belita) might come from *S. vernei*. Also, Saphir is well known for its P1, P6 and P18 resistance. However, Saphir was negative for all tested markers making it an additional resistance source. Saphir's resistance was probably inherited from the breeding clone MPI 44.1016/24 which harbours a strong resistance to pathotypes 1, 2, 6 and 18 (11). The resistance of MPI44.1016/24 and Saphir probably comes from *S. acaule*, as it was hypothesized by Ross (11). The susceptibility of Saphir for P2 suggests that additional gene(s) from MPI 44.1016/24 are missing.

Also, other varieties from the EPPO set, such as Delcora and Miriam contain resistance loci which have not yet been mapped.

Several varieties which show resistance to pathotypes 2 and/or 6 are descendants of the Dutch variety Astarte (Table 1) which shows resistance to pathotypes 1 and 18 and a weak resistance to P2 and P6. Astarte might therefore be another source of wart resistance in the breeding gene pool. Other varieties that are negative for the described markers but expressing strong resistance to higher pathotypes are Caruso, Euroviva and Otolia. The pedigrees of these varieties are not publicly available (6) so no hypothesis can be made about their putative resistance source.

Table 1. Wart disease resistance sources that are not based on the *Sen1*, 2, 3, 4, 5 genes

Variety	Year	Country	P1	P2	P6	P8	P18		adl	adg	chc	dms	EF XII-2	hit	mlt	opl	phu	spg	sto	tbr	vrn
Antares	1961	DDR	R	R			R														
Apollo	1956	DDR	R	R			R														
Astarte	1976	HOL	R	M	M		R														
Avenance	2005	HOL	R	S	R		M	Astarte													
Aviala	2004	HOL	R	R	R		S	Astarte													
Belita	1986	HOL	R	R	R		R	S. vernei													
Carrera	1999	HOL	R	S	R		S														
Caruso*	2004	BRD	R	R	R		S														
Delcora	1988	HOL	R	R	R		S														
Eurobona		BRD	R		R			Astarte													
Euroviva*		BRD	R	R	R	R	R														
Festien	2000	HOL	S	R	R		M	Astarte													
Hansa	1957	BRD	S	R	S																
Kantara	1997	HOL	R	R	R		S	Astarte													
Miriam	1988	BRD	R	R	R		S														
Monalisa	1982	HOL	R	R																	
MPI 44.1016/24		GER	R	R	R		R	S. acaule													
Nomade	1995	HOL	R	R	M		S														
Otolia*		BRD	R	R	R		R														
Parella		HOL	R	R	R		S														
Plasettie	2009	HOL	R		R		R	Belita/S. vernei													
Provento	1994	HOL	R	R	R																
Royal	2008	DEN	R	R	R		S														
Saphir	1960	BRD	R	S	R	R	R	MPI44.1016/24													
Sarpo Mira		HUN	S		R																
Simply red	2001	HOL	R	S			R														
Smart	2008	HOL	R	S	S		R														
Starga	2000	HOL	R	R	M																
Talent	2006	BRD	R	R	R		S														
Tivoli	2002	DEN	R	R	R		R?														
Ultra	1999	HOL	S		R																

* pedigree unknown. acl: *S. acaule*, adg: *S. tuberosum* ssp. *andigena*, chc: *S. chacoense*, dms: *S. demissum*, edn: *S. edinense*, hjt: *S. hjertingii*, mlt: *S. multidissectum*, opl: *S. oplocense*, phu: *S. phureja*, spg: *S. spgazzinii*, sto: *S. stoloniferum*, tbr: *S. tuberosum*, vrn: *S. vernei*.

Genes involved in potato wart disease resistance

Synchytrium endobioticum is an obligate biotrophic Chytrid fungus which needs to enter the host cell to complete its life-cycle (18). The intracellular localisation of the pathogen and the specific recognition of certain pathotypes in different host genotypes evokes a gene-for-gene recognition system (22). *S. endobioticum* is probably sensed by cytoplasmic immune receptors, better known under the name NLR (Nucleotide binding and Leucine-rich Repeat) proteins. This is for instance the case for another obligate biotrophic fungus, *Melampsora lini*, responsible for flax rust. Through its specialized structures (haustoria), *M. lini* secretes *Avr* proteins which are translocated into the cell and are recognised by the NLR *L* proteins (23,24). Recently, the first *S. endobioticum* effector *AvrSen1* (25), which triggers Hypersensitive Responses (HR) in potato genotypes containing *Sen1*, has been identified. Its transient expression in *Sen1* plants with or without a signal peptide showed that the recognition of *AvrSen1* occurs intracellularly.

Importance of TNL genes for potato wart disease resistance

In **chapters 2 and 4**, we mapped and fine-mapped *Sen1* to a region corresponding to the C76 *R* gene cluster (26) of the reference genome DM1-3 (DM) v4.03 (27,28) on the north arm of chromosome 11 (Table 2). C76 contains eight putative TNL genes (26). In **chapter 3**, we anchored the scaffold DMB734, containing three other putative TNL genes, to this chromosome 11 region. Therefore, we early suspected *Sen1* was a TNL gene. We could confirm this hypothesis in **chapter 4** with RNAi plants silenced for these TNLs, which showed a significant reduction of the HR triggered by *AvrSen1*. The candidate gene search resulted in the identification of two candidate TNL genes. In **chapter 3**, we mapped *Sen3* to the same TNL cluster as *Sen1* and made the hypothesis that *Sen1* and *Sen3* are very likely TNL genes, paralogs or even allelic variants of the same TNL gene. In **chapter 5** and in (20), the resistance locus with the strongest effect in the SaKa1 population was mapped to the long arm of chromosome 1, between 72 Mb and 79 Mb. In DM, this region contains the cluster C5 (26) which is composed of seven TNL genes.

Table 2. Comparison of the different potato wart disease resistance haplotypes mapped in this thesis with previous studies

Haplotype		Type of gene	Chr	Position**	Resistance pattern	R parent donor
This study	Previous study*					
<i>Rse-Ib-Andante-a</i>	<i>Rse-Ib-a^a</i>	QTL	1	~ 72 - 80 Mb	P1, P2, P6, P18	Andante
<i>Rse-Ib-Andante-c</i>		QTL	1	~ 70 - 76 Mb	P1, P2, P6, P18	Andante
<i>Sen5</i>		Major gene	5	[49.2 - 51.16 Mb]	P2, P6, P18	Aventra
<i>Rse-VIIIb-Kuba</i>		QTL	8	~ 44 - 47 Mb	P18	Kuba
	<i>Rse-VIIIa^c</i>	QTL	8	south arm chr 8	P1, P2, P6, P18	Saturna, Panda
<i>Rse-IXa-Andante</i>	<i>Rse-IXa-a^a</i>	QTL	9	~ 55 - 61.5 Mb	P18	Andante
<i>Rse-IXb-Ludmilla</i>		QTL	9	~ 55 - 57 MB	P2, P6, P18	Ludmilla
<i>Sen1</i>	<i>Sen1^d</i>	Major gene	11	[1.31 - 1.67 Mb]	P1	Desiree, Kuras
<i>Sen3</i>	<i>Rse-XIa^{ef}</i>	Major gene	11	[1.26 - 1.52 Mb]	P1?, P2, P6, P8?, P18	Kuba
<i>Rse-XIc-VTN62.33.3</i>		QTL	11	~ 0 - 5 Mb	P2, P6, P18	Axion, Aventra
	<i>Sen2^g</i>	Major gene	11	[33.7 - 35.06 Mb]	P1, P2, P3, P6, P8, P18, P39	DG 97-264
<i>Rse-XId-Andante</i>		QTL	11	~ 0 - 31 Mb	P2, P6, P18	Andante
<i>Sen4</i>		Major gene	12	[48.5 - 51.5 Mb]	P2, P6, P18	Axion
	<i>Rse-XIIa^e</i>	QTL	12	full chr 12	P2, P6, P18	Karolin, Ps-354

* The "Rse-chromosome number" naming is according to (21). A letter was added at the end of the locus name to differentiate the different haplotypes identified. ** [-]: interval of the closest flanking markers, ~ : peak interval in CoSSA results. Loci previously identified by ^a (20), ^b (2), ^c (19), ^d (1), ^e (21), ^f (29), ^g (3).

TNL genes seem, therefore, to play an important role in resistance to *S. endobioticum* and one could wonder if there is something in the lifestyle of *S. endobioticum* that makes TNLS more effective compared to other *R* gene types. TNLS have been shown to be involved in resistance to several pathogens such as bacteria, fungi, nematodes, oomycetes and viruses. In potato, the three TNLS cloned so far are involved in resistance to the potato virus Y (PVY) (30,31) and to *G. rostochiensis* (32). Also *R* genes against other pathogens, such as *P. infestans* (33,34) or *G. pallida* (35), have

been mapped to TNL clusters as well. So, at a first glance, it is not possible to highlight a correlation between the NLR type and the recognition of a specific pathogen group. The *L* locus of flax gives resistance to *M. lini*, an obligate biotrophic fungus (36), and so far, all the cloned NLR genes involved in resistance to flax rust are TNLs (37). TNL genes are also involved in resistance to the obligate biotrophic oomycete *P. parasitica* in *A. thaliana*, however, CNLs and other gene classes such as receptor-like proteins (RLPs) are also involved (38).

TNLs versus CNLs: Requirements for downstream signalling

If TNLs are more efficient than CNLs for the resistance to *S. endobioticum*, what could be the reason? A substantial difference between TNLs and CNLs is the signalling pathways recruited by the N-terminal domains. Since the first studies on the signalling components of NLRs in the 1990's, it has been shown that most CNLs require the NDR1 (Non-Race Specific Disease Resistance-1) protein (39,40) to activate immunity whereas most TNLs require the EDS1 (Enhanced Disease Susceptibility 1) protein (39,41). Recently, it has been shown that some NLR proteins require other NLR proteins, called "helpers", to transduce the defence signals. These NLR proteins have evolved to work in pairs, one partner being the sensor NLR and the other the helper NLR, or even in more complex networks involving more partners (42,43). The helper proteins required by CNLs and TNLs can differ.

In Solanaceae, it has been shown that an important number of CNLs, such as for example the *P. infestans* *R* genes *R1*, *R8* and *Rpi-blb2*, require other CNL helper proteins from the NRC (NLR-Required for Cell death) family (44–47). It is not yet known if Solanaceae TNLs require these NRC helpers. In *A. thaliana* and in *N. benthamiana*, the NRG1 (*N* Requirement Gene 1) protein from the RPW8-NLR type is required by several TNLs for defence signalling (45,48,49). The ADR1 (Activated Disease Resistance 1) type of CNLs are also NLR helpers having a role downstream several TNLs (45,49). Interestingly, it was shown recently that NRG1s and ADR1s are involved in two different parallel pathways downstream TNLs (45). Some CNLs showed also an ADR1 (50,51) dependency or a weak NRG1 dependency (48). In potato, the only study of the signalling pathway of a TNL gene showed that the PVY resistance TNL gene *Rysto* is EDS1 and NRG1 dependent (52).

Recently, a great discovery has been made about the N-terminal CC domain function and structure of the *A. thaliana* ZAR1 (Hopz-Activated Resistance 1) CNL (53). Upon activation by the *Xanthomonas campestris* effector AvrAC, an inactive form of the CNL

ZAR1 will undergo a structural remodelling and will form the wheel-like pentameric ZAR1 resistosome with RKS1 (Resistance-related Kinase 1) and PBL2 (PBS1-like protein 2). The five N-terminal $\alpha 1$ helices of the ZAR1 CC domains will form a funnel which is required for cell death through its insertion in the plasma membrane and the formation of a pore, toxic for the cell. There are other CNLs which are located at the plasma membrane for which a similar mode of action is suspected. The N-terminal domains of CNLs and TNLs are very different so TNLs likely trigger HR through a very different system, or through an association with CNL helpers to form a similar cell-death triggering structure as the ZAR1 resistosome.

It would be of great interest to investigate the helper proteins required by the TNLs involved in resistance to *S. endobioticum*. The recruitment of a certain type of helper proteins might explain an advantage of TNLs over CNLs for resistance to the wart disease. The recruitment of helpers from another genomic region might also explain the background dependency of some resistance loci such as the *Rse-Ib-Andante* chromosome 1 resistance locus from Andante in SaKa1 (**chapter 5**).

TNLs versus CNLs: subcellular localisation

Different compartmentalization of the plant NLRs might also explain why certain TNLs and/or certain helpers are more efficient in resistance to *S. endobioticum* than others. Indeed, it has been observed in several studies that some NLRs such as MLA10 (CNL) in barley and RPS4 (TNL) in *Arabidopsis*, accumulate in the nucleus after activation by their matching effector and this nuclear localization is necessary for the activation of defences (54). On the other hand, *Rx*, a CNL giving extreme resistance to PVX in potato, is inactivated when localized in the nucleus, the pathogen recognition and the signalling pathways activation happen in the cytoplasm (55,56). Interestingly, a localization at the plasma membrane is required for some NLRs to function (54). This is the case for RPS5 and RPM1, two CNL proteins of *Arabidopsis*. This plasma membrane localization is likely required because the matching pathogen effector proteins also localize there and, in the case of CNLs such as ZAR1, because they form a resistosome which triggers HR through the formation of a pore in the plasma membrane. The protein sequence of AvrSen1 does not contain subcellular localization signals (25). Therefore, the recognition of AvrSen1 by Sen1 and the recruitment of the downstream signalling pathways probably occur in the cytoplasm, where AvrSen1 is secreted by *S. endobioticum*. Interestingly, *L6* and *M*, two TNL genes involved in resistance to *M. lini*, localize respectively to the Golgi apparatus and to the tonoplast membranes (57). It was shown that it is the N-terminal domain which is responsible

for this localization, and the anchoring to the endomembranes which was required for the protein stability. Is it necessary for the TNLs involved in *S. endobioticum* resistance to possess N-terminal signal anchoring sequences for a better efficiency through an improved stability? The cloning of the *Sen* genes, their (alternatively spliced) mRNAs and a study of the proteins localization in the cell and their domains function should answer these questions.

Recognition of *S. endobioticum* effectors by potato wart disease R genes

Sen1 and *Sen3* are likely paralogs from similar TNLs clusters or allelic variants of the same gene (**chapters 3 and 4**). *Sen1* brings specific resistance to P1 isolates through the recognition of AvrSen1. *Sen3*, however, possesses a broader recognition spectrum as we showed it brings resistance to P2, P6 and P18, and most likely to P1 as well. The recognition pattern of *Sen3* could be due to the recognition of an effector highly conserved between the pathotypes populations. Another possibility could be that *Sen3* can bring resistance to P2, P6 and P18 through the recognition of an effector common to these three pathotypes and that both *Sen1* and *Sen3* can bring resistance to P1 upon the recognition of AvrSen1. There are some examples in literature of different R proteins which are able to detect the same effector protein. This is the case in the pathosystem composed of *B. napus* and *L. maculans*, the hemibiotrophic ascomycete responsible for blackleg. The two genes *Rlm1* (not cloned) and *LepR3* (RLP), located on different chromosomes, can recognise the effector *AvrLm1* (58). The effector protein AvrLm4-7 is also detected by the two *R* genes *Rlm4* and *Rlm7* (59) which are located in a same cluster (60). More recently, it was observed that Avr2, an effector protein from *P. infestans*, can be detected by potato *R* genes from the R2 family (CNLs), such as for instance *Rpi-blb3*, and *Rpi-mcq1* (CNL) (61). The agroinfiltration of *AvrSen1* in *Sen3* and non-*Sen3* plants could confirm the hypothesis of *Sen3* being able to recognize AvrSen1. If this hypothesis is true and if *Sen3* is a unique gene (not closely linked NLR paralogues), it would mean that *Sen3* enables “dual specificity” recognition of at least two different effectors from *S. endobioticum*. One well known example of a gene enabling dual specificity is *RPM1* which recognizes both AvrRpm1 and AvrB effectors from *Pseudomonas syringae* (62). This dual specificity is enabled by the “guardee” protein RIN4 which is required by *RPM1* to bring resistance (63). Another example of this dual specificity enabled by a guardee protein is *Cf-2*, in tomato, which brings resistance to *C. fulvum* and *G. rostochiensis*. The apoplastic tomato protein *Rcr3* is targeted by the *C. fulvum* effector Avr2 and the *G. rostochiensis* effector Gr-VAP1 (64). The activity of these effectors on *Rcr3* is detected by *Cf-2*.(65,66). Again, the

cloning of the *Sen1* and *Sen3* genes would provide very useful tools to further study the recognition mechanism.

Also, knowing the recognition pattern of the different *Sen* genes will facilitate the identification of the matching *S. endobioticum* *Avr* genes. *AvrSen1* was identified through comparative genomics between P1 and the higher pathotypes, the basic premise being that the gene encoding *AvrSen1* should be present in the sequenced P1 genomes but functionally or structurally absent from the higher pathotypes genomes (25). I expect a similar comparative genomics strategy to work for *AvrSen5* as well (**chapter 5**). *Sen5* is involved in resistance to P2, P6 and P18 but does not provide resistance to P1. This strategy would also work to identify the effectors matching the resistance genes present in Saphir (P1, P6, P18 resistant), Delcora and Miriam (both P1, P2, P6 resistant) (Table 1), if their resistance is brought by a single gene. Regarding *Sen4*, the complete resistance pattern is not known yet as only P2, P6 and P18 were tested (**chapter 5**). We identified in **chapter 5** two varieties which carry *Sen4* without *Sen1* but one is scored as susceptible to P1 (Celandine) and the other resistant (Panda). For *Sen2* and *Sen3*, which provide a broad resistance spectrum, this comparative genomics approach would not work as all isolates are expected to harbour those *Avrs*. Novel approaches need to be designed to identify these *Avrs* conserved between pathotypes.

In **chapter 4**, we used *AvrSen1* to phenotype the *Sen1* recombinants and the RNAi plants silenced for the TNL clusters C76 and C77. This phenotyping method gave unambiguous results and did not require the multiplication of tubers, while the Spieckermann (67) and Glynne-Lemmerzähl (68,69) assays require a significant number of tubers and at least two seasons of phenotyping to give reliable results. Therefore, agroinfiltration of the *Avrs* as a phenotyping tool can speed up drastically the wart disease resistance loci mapping and fine-mapping experiments. It could as well be used to screen for resistant varieties, or wild potato relatives in order to select for non-redundant resistance sources. *Avr* response screens are more cost intensive than molecular markers, but can be considered in case molecular markers are not available, not specific enough or too distant from the matching *R* gene.

Potato wart disease Quantitative Trait Loci: NLRs or different gene types involved?

In this thesis, we identified qualitative effect loci such as *Sen1* and *Sen3*, which are members from the NLR family, and QTLs with different ranges of effects on resistance. We can speculate about the kind of genes behind these QTLs. *Sen4* and *Sen5*, identified

in **chapter 5**, seem to have a more quantitative effect on resistance compared to *Sen1*, *Sen2* or *Sen3*, although they bring a good level of resistance. *Sen4* was mapped to the south arm of chromosome 12, between 48.5 Mb and 51.5 Mb. In the DM reference genome, this region contains only one CNL gene (26) and one RLK (Receptor-Like Kinase) (70). I suspect however the *Sen4* haplotype to be quite different from the reference genome. The CoSSA performed in the AxV population resulted in a very broad peak of *k*-mers between 10 Mb and 54 Mb. This broad peak can be explained by a small number of recombination events in the resistant bulk in the haplotype in which *Sen4* was introgressed. We observed a low recombination frequency of the *Sen4* haplotype also in the entire AxV population which might be explained by suppression of recombination due to the introgression of a segment quite different from the other haplotypes. This hypothesis must be tested in larger populations. It is therefore not possible to make assumptions about the *Sen4* candidate genes using DM, a *de novo* assembly of this *Sen4* haplotype would be necessary. *Sen5* also showed a quantitative effect on resistance in the AxD and KxA populations (**chapter 5**). We mapped *Sen5* to the south arm of chromosome 5, between 49.2 Mb and 51.16 Mb. In DM, this region holds the NLR clusters C40 and C41 (26). The DM assembly of this region is sub-optimal as it contains four gaps. *Sen5* was probably introgressed in the breeding genepool from *S. vernei* and only the *de novo* assembly of this haplotype would allow to identify the candidate genes. In the SaKa1 population, two haplotypes from the south arm of chromosome 1 were involved in resistance to P1, P2, P6 and P18 (**chapter 5**). This could be due to different genes involved in resistance to wart disease or to several alleles from the same gene. In DM, this region contains the two *R* gene clusters C4 and C5 (26), composed of two CNLs and seven TNLs respectively, and four RLKs (70). The two CNL genes from cluster C4 are homologues of the NRC1 helper gene which has been shown to be a signalling component required by cell surface immune receptors, such as *Cf-4* and *Cf-9* (two RLPs), and by intracellular immune receptors such as *Rx* and *Mi-1.2* (two CNLs) (46). Silencing these NRC1 homologues in Saka1 descendants would reveal if they also have a “helper” role in resistance to *S. endobioticum* and if they can explain the background dependency of the chromosome 1 main resistance locus from Andante.

In **chapters 3 and 5**, several minor effect QTLs which had an additive effect to the major effect gene were identified. In the AxV, AxD and KxA populations of **chapter 5**, one QTL, *Rse-IXc-VTN62.33.3*, on the north arm of chromosome 11, brought resistance to some of the higher pathotypes. The recognition pattern was different between the populations which could be due to recombinations, to a different genetic

composition of the isolates used for the phenotyping, or to dependency on genetic factors in the background of the plant. Indeed, *Sen4* or *Sen5* are also segregating in these populations and the effect of these genes might be epistatic. The markers designed on the *Rse-IXc-VTN62.33.3* haplotype spanned a huge region so it is difficult to make assumptions about the candidate genes behind it. In the SaKa1 population, we found on the south arm of chromosome 9, around 60-61 Mb, *Rse-IXa-Andante* which is involved in resistance to P18 only. The reference genome in this region is rich in *R* gene clusters as clusters C64, C65 and C66, composed of CNLs and TNLs, are located between 59.3 Mb and 61 Mb (26). Many *R* genes involved in resistance to different pathogens such as *P. infestans*, nematodes or viruses, have been mapped and/or cloned in this region. It is likely that one NLR gene from these clusters, introgressed by chance while introgressing resistance to other pathogens, recognises specifically an effector present in P18 isolates. In KxL (**chapter 5**), we identified *Rse-VIIIb-Kuba* on the south arm of chromosome 8 around 45 Mb which was also specifically associated with P18 resistance. In DM, this region contains the NLR clusters C59, C60 and C61 (26). Overall, these observations suggest that NLR genes might also be candidate genes for the minor effect QTLs.

NLR genes are usually known to give a qualitative resistance upon specific recognition of the matching pathogen Avr protein. There are, however, exceptions and it has already been observed that NLR genes can bring incomplete or quantitative resistance (71,72). If NLR genes are truly responsible for *Rse-IXa-Andante* and *Rse-VIIIb-Kuba* minor effect QTLs giving P18 resistance, how to explain their weak effects compared to *Sen1* and *Sen3*? In the *AvrSen1* paper (25), the *AvrSen1* gene from different isolates was PCR amplified and sequenced with long read technology to detect the proportion of the different *AvrSen1* variants in each isolate. All the reads obtained from two P1 isolates contained the wild-type of *AvrSen1* which explains the qualitative resistance brought by *Sen1*. To explain the major effect of *Sen3* on P2, P6 and P18 resistance and the minor effect of *Rse-VIIIb-Kuba* on P18, we can imagine that *AvrSen3* is present in a high frequency in the *S. endobioticum* population from the P2, P6 and P18 isolates that were used for the phenotyping. The Avr matching *Rse-VIIIb-Kuba*, however, is then expected to be present in a low frequency in the P18 isolate and absent from the P2 and P6 isolates used. This suggests that P18 isolates possess a more diverse repertoire of effectors than P2 and P6 isolates. Indeed, *S. endobioticum* mitochondrial haplotype diversity within-isolates was much higher in P8, P18 and P38 isolates than in isolates of P2 and P6 (73).

Perspectives for potato wart disease resistance breeding and management

S. endobioticum, being one of the most devastating quarantine diseases of potato, requires that strict quarantine measures and clever management strategies are set. The knowledge acquired in this thesis about the specific Avr proteins/R proteins interaction responsible for pathotype specific resistances, the major and minor effect resistance loci and the different resistance sources are strong weapons to improve wart disease resistance breeding programs and the management of the pathogen.

The EPPO set does not encompass the R loci diversity present in the potato breeding genepool

So far, there are no molecular diagnostic tools to discriminate between *S. endobioticum* pathotypes. The pathotype identification is therefore made by means of biotests using a differential set of potato varieties. For decades, European countries used different sets of varieties for the pathotyping and used different naming systems for the pathotypes which made the comparison between pathotypes impossible (74). In 2006, Baayen proposed a standardisation of the pathotypes naming system and a harmonisation for the set of varieties used, which is now called the EPPO set. The EPPO

Table 3. Differential set for R gene/Avr gene specific interactions.

Variety	AvrSen1	AvrSen2	AvrSen3	AvrSen4	AvrSen5	AvrSenDelcora	AvrSenTalent	AvrSenMiriam	AvrSenSaphir	AvrSenBelita	P1	P2	P6	P8	P18	R gene content
Tomensa/Deodara	-	-	-	-	-	-	-	-	-	-	S	S	S	S	S	none
Producent	+	-	-	-	-	-	-	-	-	-	R	S	S	S	S	Sen1
Delcora	+	-	-	-	-	+	-	-	-	-	R	R	R	S	S	Sen1, SenDelcora
Talent	+	-	-	-	-	-	+	-	-	-	R	R	R	?	S	Sen1, SenTalent
Miriam	+	-	-	-	-	-	-	+	-	-	R	R	R	R	S	Sen1, SenMiriam
Saphir	+	-	-	-	-	-	-	-	+	-	R	S	R	R	R	Sen1, SenSaphir
Aventra	-	-	-	-	+	-	-	-	-	-	S	R	R	?	R	Sen5
Axion	+	-	-	+	-	-	-	-	-	-	R	R	R	?	R	Sen1, Sen4
DG97-264	+	+	-	-	-	-	-	-	-	-	R	R	R	R	R	Sen1, Sen2

Ikar/Gawin / Karolin	-	-	+	-	-	-	-	-	-	-	R	R	R	R	R	<i>Sen3</i>
Belita	+	-	-	-	-	-	-	-	-	+	R	R	R	R	R	<i>Sen1</i> , <i>SenBelita</i>

- : no hypersensitive response, +: hypersensitive response, R = resistant, S = susceptible

set was formerly composed of eleven varieties (75). Later, the EPPO set was updated: the varieties Eersteling, Combi, Delcora, Miriam and Ulme were removed and the varieties Evora, Irga, Talent, Ikar and Gawin were added (76). In the new EPPO set, Ikar, Gawin and Karolin are redundant as they all possess *Sen3* which confers them resistance to P1, P2, P6 and P18 (**chapter 5**). Varieties Producent, Talent, Saphir and Belita possess *Sen1*. The resistance source of Talent, Saphir and Belita is unknown and does not come from *Sen2*, *Sen4* or *Sen5*. Therefore, *Sen2*, *Sen4* and *Sen5* are not represented in the EPPO set, neither the resistances present in Andante. The inclusion of varieties holding these *R* genes would probably lead to the identification of new pathotypes, thereby better reflecting the pathogen diversity that is around.

The identification of *S. endobioticum* *Avr* genes (25) could lead to a new pathotyping system based on the *R* gene/*Avr* gene composition, such as for *P. infestans* (77). In Table 3, I made an attempt for such a differential set. Ideally, each potato variety should contain only one *R* gene. Based on this system, the isolates would be characterised by a virulence pattern that better reflects *Avr* contents. Besides inoculation tests, the presence of *Avrs* could be tested through PCR assays (Table 4).

Table 4. Pathotype definition according to their *Avr* content

Pathotype/ Isolate	<i>AvrSen1</i>	<i>AvrSen2</i>	<i>AvrSen3</i>	<i>AvrSen4</i>	<i>AvrSen5</i>	<i>AvrSenDelcora</i>	<i>AvrSenTalent</i>	<i>AvrSenMiriam</i>	<i>AvrSenSaphir</i>	<i>AvrSenBelita</i>
1	+	+	+	-?	-	-?	-?	-?	-?	-?
2	-	+	+	+	+	+	+	+	-	+
6	-	+	+	+	+	+	+	+	+	+
8	-	+	+	?	?	-	?	+	+	+
18	-	+	+	+	+	-	-	-	+	+

- : absent, +: present

Such a pathotyping system would fine-tune the management of the disease: knowing the *Avr* composition of the isolates present in the field will allow to deploy an adapted set of varieties containing the right combination of *R* genes and QTLs. Tables 3 and 4 contain only the major *R* loci and the corresponding most frequent *Avrs*, but minor

effect loci and less frequent *Avrs* should be included as well and taken into account for the choice of resistant varieties.

Stacking of major and minor effect loci

In **chapter 5**, we highlighted the importance of minor effect loci to bring full resistance to the higher pathotypes, especially for P18. We ventured the hypothesis that the small additive effect on resistance brought by some loci might be due to NLR genes which detect an effector present in a low frequency in the *S. endobioticum* isolate used for the phenotyping. Van de Vossen *et al.* (73) showed that the higher pathotypes such as P8, P18 and P38 had a higher intra-isolate diversity than P1 or P2. So for higher pathotypes, stacking of major and minor effect loci in the resistant varieties is essential to bring full field resistance and to avoid a shift in the *S. endobioticum* populations in the field. Indeed, if an isolate holds one *Avr* gene with a high frequency and another *Avr* with a low frequency, and if potato varieties possessing only the *R* gene matching the most frequent *Avr* are deployed, the *S. endobioticum* genotypes that do not possess the detected *Avr* protein will multiply. After a few multiplication cycles, the *S. endobioticum* population will become virulent on the potato varieties deployed. In the *AvrSen1* paper, such a shift was observed with a P1 isolate multiplied on a partially resistant variety (25). Only two multiplications were necessary to obtain a shift of the isolate population from a P1 to a P6. To avoid this kind of population shifts and to guarantee an efficient and durable field resistance, breeders should focus on the stacking of several *R* loci in their material, combining major and minor effect loci.

R genes pyramiding is a strategy which has been known for decades to enhance resistance and resistance durability in crops. Few studies however have focused on the interaction between *R* genes and the benefits or costs of these inter-genic interactions. Our former hypothesis was that the stacking of major and minor effect loci would have an additive effect thanks to the recognition of different *Avrs* present in the *S. endobioticum* isolate. Another explanation of the positive effect of combining major and minor effect loci could be inter-genic complementary interactions. It might be the case for *Rse-VIIIb-Kuba* segregating in the KxL population (**chapter 5**). Indeed, the effect of this locus on P18 resistance seems stronger when *Sen3* is also present. I performed an ANOVA to test the interaction between the two loci and this interaction was significant (p -value < 0.01) which could reveal a beneficial interaction between the two *R* genes. Such a beneficial inter-genic interaction has been recently studied in *Medicago truncatula* between two *R* genes, *AKR* and *AIN*, bringing resistance to the bluegreen aphid (78). Plants holding both genes show an additive effect on resistance.

Moreover, *AKR* is epistatic over *AIN* as it reduces the negative growth phenotype which is associated with *AIN*. This interaction between the two genes is mediated by a crosstalk at the phytohormone level (78,79). Positive interactions between *R* genes have also been observed in potato for resistance to nematodes (80) although the molecular mechanisms behind that has not been described, but most likely has its origin in the diversity within pathogen populations, similarly as my hypothesis for wart disease resistance.

How to design diagnostic markers for the minor effect loci

In **chapter 5**, we designed haplotype specific markers associated with minor effect loci. Because the effect of these loci is small, their fine-mapping and the design of flanking diagnostic markers is difficult. A traditional strategy to fine-map such QTLs would be to increase the size of the mapping population. Another strategy I propose would be to enrich the *S. endobioticum* population used for the phenotyping for the *Avr* matching the minor effect locus of interest. For instance, to fine-map *Rse-VIIIb-Kuba* improving resistance to P18 in the KxL population (**chapter 5**), I would recommend to select KxL descendants that hold *Sen3* but do not hold *Rse-VIIIb-Kuba*. After a few multiplication rounds on these selected descendants of the P18 isolate we initially used for the phenotyping in **chapter 3**, only the *S. endobioticum* genotypes that do not hold *AvrSen3* will be able to reproduce and there will be an enrichment of the effector matching *Rse-VIIIb-Kuba*. A new phenotyping round of the KxL population with the “trained” P18 isolate will end up in a stronger effect of this locus and an easier mapping and fine-mapping. This strategy can be combined with the sequencing of the genomes of initial and trained isolates to identify the *Avrchr8* gene matching *Rse-VIIIb-Kuba* candidate *R* gene, which can then be applied through agroinfiltration, resulting in unambiguous phenotypes.

R gene mapping and cloning strategies

In this thesis, we combined several tools such as GWAS, BSA and *in silico* BAC walking for the mapping of several potato wart disease *R* genes, and for the fine-mapping and candidate gene search of *Sen1*. Across these different studies, I encountered some pitfalls such as ascertainment bias of SNP arrays and used efficient tools such as *k*-mers. Here, I will address these different aspects and will discuss current strategies to map and clone *R* genes in plants.

The use of SNP arrays: limited genome coverage and ascertainment bias

In **chapter 2** of this thesis, we performed a Genome Wide Association Study of potato wart disease P1 resistance in the potato breeding genepool. The genotypic data used in this GWAS was obtained by (5) using the 20K SNP SolSTW array. We identified a peak of significantly associated markers on the north arm of chromosome 11, where *Sen1* had been previously identified (1). The significant markers were validated in two populations. The most closely linked GWAS marker was however still 10 cM away from *Sen1*. We inquired why we could not identify markers closer to *Sen1* and why we did not identify any *Sen1* haplotype specific marker on the telomeric side of *Sen1*. We were not hampered by ascertainment bias with the 20K SNP array as P1 resistance is common in the potato breeding gene pool and several varieties used to identify the 130K variants in (81) are resistant to P1. Moreover, at least nine of these 84 varieties hold *Sen1*: Adretta, Agria, Aveka, Avenance, Innovator, Kuras, Nomade, VE71-105 and VTN62-33-3 (see **chapter 5**). Still, among the 130K variants, none were located between 814,888 bp and 1,416,133 bp which explains why we could not identify any SNP marker upstream the *R* gene cluster C76 (26). Moreover, among the markers designed upstream 2.7 Mb, none were from the *Sen1* haplotype, which explains why our closest marker was still 10 cM away from *Sen1*. So the obvious reason why our markers were not fully diagnostic is because of the bait strategy used in (81). No baits were designed in the *Sen1* region, probably because the authors' focus was not on resistance genes and because they avoided to design baits in regions exhibiting paralogous sequences in the genome, which is the case for *R* gene rich regions.

In **chapter 5**, we performed a GWAS as well in the potato breeding genepool but this time for P2, P6 and P18 resistance. This GWAS allowed us to identify *Sen4* and *Rse-IXc-VTN62.33.3*. The *Sen4* SNPs were identified thanks to the cultivars Hindenburg and VE70-9 in the set from (81). We know that other genes than *Sen4* are present in the breeding genepool so why could not we identify these other loci? In the 130K variants set, markers were designed on the *Sen5* haplotype thanks to the presence of VE 71-105 in the discovery panel. We were, however, unable to identify *Sen5* in the GWAS, because *Sen5* was too rare in the GWAS panels (only present in Avenra and VE71-105). In **chapters 3 and 5**, we observed that *Sen3* was the main wart disease resistance source in the German and Polish breeding material. *Sen3* was however not found with the GWAS. This is obviously due to ascertainment bias. Indeed, no East European variety was included in the discovery panels of (81,82). Therefore, the 20K SNP array does not cover any East European haplotype and is unable to detect *Sen3*.

Moreover, Ulme is the only *Sen3* variety which was present in the GWAS panel so even if the 20K SNP array was covering East European haplotypes, *Sen3* would most probably not have been found.

Towards an optimal R gene mapping and cloning strategy

To prevent ascertainment bias, we decided to use Whole Genome Sequencing (WGS) data in further mapping studies performed in this thesis. In **chapter 3**, we developed CoSSA (4), a novel set of workflows for genetic analyses of a trait of interest, that was used to (fine-)map *Sen1*, *Sen3*, *Sen4*, *Sen5* and minor effect QTLs in **chapters 3, 4 and 5**. In Figure 3, I made an overview of different strategies which can be used for R gene mapping and cloning in plants. In this part of the discussion, I want to compare these strategies with the one we used in **chapter 4** to fine-map *Sen1* and identify candidate genes.

Overall, all the mapping strategies described in Figure 3 comprise the same steps which are the mapping of the locus, the generation of a *de novo* assembly, the fine-mapping and the candidate gene search. The different strategies can be divided in two categories: biased for a specific gene family (e.g. NLRs, RLKs, RLPs etc.) and the unbiased strategies which do not target any specific gene family. With respect to ascertainment bias, the AgRenSeq (83), SMRT-RenSeq (84), HAWK (85) and CoSSA (4) strategies, seem equally efficient because they are based on sequencing data. Although, the complexity reduction in AgRenSeq and SMRT-RenSeq may introduce a novel bias, in case the causal gene is not part of the targeted gene family. However, the complexity reduction does offer the opportunity to overcome sequence coverage limitations, especially for big or polyploid genomes. Linkage based QTL mapping strategies, on the contrary, can be subjected to ascertainment bias as they often use SNP arrays or previously designed markers for genotyping.

The choice of the best strategy depends then on the plant material used in the mapping study. For diversity panels, the AgRenSeq and HAWK strategies, both relying on a *k*-mer based association mapping approach, are suitable. AgRenSeq was shown to be very efficient to identify and clone four stem rust *R* genes in *Aegilops tauschii* (83). HAWK has not yet been applied to plants but was successful in identifying several disease-associated sequences in human (85). In bi-parental populations, CoSSA and RenSeq showed to be both efficient to map *R* genes. CoSSA seems however faster than RenSeq as there is no need to design a bait library, the DNA library preparation is cheaper, and the bioinformatics workflow is faster as there are no read mapping

Plant material	Biased for specific gene families		Mapping / Cloning experiment		Unbiased strategies	
	Diversity panel	Bi-parental population	Diversity panel	Bi-parental population	Bi-parental population	
Approach
Time	Long to build the panel Rest very fast	Fast	Long to build the panel Rest very fast	Very fast	Very slow	
Cost	\$	\$\$\$	\$	\$(if assembly available)	\$	
Examples	<i>Sr33^a</i> , <i>Sr45^a</i> , <i>Sr46^a</i> , <i>SrTA1662^a</i>	<i>Rpi-amt3^b</i> , <i>Ry^c_{STO}</i>	Human diseases ^d , ampicillin resistance in <i>E. coli</i> ^f	<i>Sen1^e</i> , <i>Sen3^e</i> , <i>Sen4^e</i> , <i>Sen5^e</i> , several wheat disease resistance QTLs ^e	<i>Gro1-4ⁱ</i> , <i>Rpi-blb2^g</i> , <i>R3b^h</i> , <i>R8</i>	

Figure 3. Overview of different strategies for the mapping and cloning of *R* genes

To choose the optimal approach for a *R* gene mapping or cloning experiment, ones should not consider the described approaches as fixed and should consider combining or inter-changing the steps from different approaches. Although the AgRenSeq and SMRT-RenSeq strategies were formerly developed to identify NLR genes, the bait library can be designed to target any type of gene. RB: Resistant Bulk, SB: Susceptible bulk, HAWK: Hitting Associations With *K*-mers, CoSSA: Comparative Subsequence Sets Analysis, BAC: Bacterial Artificial Chromosome. ^a (83), ^b (84), ^c (52), ^d (85), ^e this thesis, ^f (32), ^g (86), ^h (87), ⁱ (88).

and variant calling steps. For complex traits, the BSA (Bulked Segregant Analysis) approach used with RenSeq and CoSSA should be replaced by linkage-based QTL mapping. As discussed in **chapter 3**, we can imagine broader applications of CoSSA and expand it to an association mapping type of approach. Instead of building bulks composed of descendants from a bi-parental population, bulks composed of a broader panel of genotypes with opposite phenotypes could be used. This approach would however require the causal gene(s) to be identical-by-descent. Recently, a similar approach was designed by (89). The authors sequenced the H1 and the R2 bulks, composed each of 15 varieties holding the *H1* nematode and the *R2* late blight *R* genes respectively. Three negative bulks composed each of 15 varieties which do not hold *H1* nor *R2* were also sequenced. SNPs with skewed frequencies between the bulks were selected to design KASP markers diagnostic for the two *R* genes.

If one's aim is to clone the targeted *R* gene and not only to design diagnostic markers, the chosen mapping strategy should be followed by the generation of a *de novo* assembly of the targeted region. To obtain a correct assembly of the targeted *R* gene region, the classical approach was to generate BAC (Bacterial Artificial Chromosome) libraries and engage a long, laborious, but efficient process of BAC walking or BAC landing. In the study from (85), HAWK was used to identify *k*-mers linked with the trait and they directly assembled these *k*-mers. This strategy has not been applied to *R* gene cloning in plants yet but it would obviously be unsuccessful to assemble correctly *R* gene clusters because repetitiveness cannot be resolved with short *k*-mers. In this study, my first attempt to assemble the *Sen1* haplotype, which was not included in this thesis, was to extract the reads from the resistant parents of the AxDK population that contain the *Sen1* specific *k*-mers, and to assemble them *de novo*, as we did in **chapter 3** in the reference-free CoSSA workflow. This assembly of short reads could be used to design new markers to continue the fine-mapping but it was too scattered to identify the candidate genes. Likewise, in the study from (84), short reads obtained with RenSeq were assembled and the authors showed that the assembly of *R* gene clusters using short reads can lead to chimeric genes. My second attempt, also not included in

this thesis, was to make a hybrid assembly combining the short reads containing the *Sen1* specific *k*-mers and PacBio reads from another variety containing *Sen1*. The use of long reads is an efficient strategy to solve paralogous and repetitive sequences such as *R* genes rich regions, which cannot be solved with short reads. The sequencing depth of the PacBio reads was however insufficient to assemble correctly the *Sen1* *R* gene cluster. The assembly was indeed very scattered and none of the scaffolds contained more than one *R* gene. With their innovative SMRT-RenSeq approach, (84) were more successful in assembling correctly the *R* gene cluster containing *Rpi-amr3* with the use of PacBio reads combined with RenSeq. Finally, we used the phased *de novo* assembly of a tetraploid variety containing our gene of interest *Sen1* to pursue the fine-mapping and perform the candidate gene search. This strategy was successful in identifying *Sen1* candidate genes in the fine-mapped region. It is difficult to compare the efforts put in generating the assemblies from this study and the study of (84) as we used a *de novo* assembly produced by NRGene (NRGene, Nes Ziona, Israel), and we only had to correct manually the assembly of one candidate gene using Nanopore data to validate that the gene model was correct. The PacBio assembly errors from (84) had to be corrected manually using RenSeq short reads which seems a laborious step. Even if Illumina paired end short reads and the 10X Genomics Chromium system used in the NRGene assembly are less expensive than the PacBio technology, the phased NRGene assembly is probably more expensive than the SMRT-RenSeq assembly because the full genome was assembled and phased, and the assembly was outsourced. The PacBio technology is still quite expensive, but with the RenSeq genome complexity reduction, the costs can be reduced. The full phased *de novo* assembly that we used has however the advantage to be useful for many more purposes than the SMRT-RenSeq assembly.

To conclude, the *R* gene fine-mapping and cloning strategy I would recommend depends on the data available. I recommend the use of CoSSA for the mapping and fine-mapping of the gene as it is not biased toward a specific gene family, it can be applied to crops for which a reference genome is not available, it is suitable for polyploids and it is quick and cheap. For crops with a big genome, I would recommend to combine CoSSA with a genome complexity reduction approach, such as RenSeq if the targeted gene family is known, or other approaches which are not biased for specific gene families. If an assembly such as the phased NRGene *de novo* assembly is available for a variety containing the gene of interest, it is the preferred strategy to pursue the candidate gene search as no additional costs are necessary. If none is available, I would recommend to produce long reads for a genotype containing the

gene of interest and to extract the reads to assemble using the resistance specific *k*-mers identified with CoSSA. To reduce the costs of the long reads sequencing, a complexity reduction approach can also be applied at this step.

'Shared *k*-mers': a powerful multi-purpose tool

CoSSA is based on the use of *k*-mers and set algebra to compare DNA sequences between samples. Comparing directly the sequenced reads instead of *k*-mers would be difficult because of their size and sequence errors. Cutting the reads into smaller *k*-mers of unvarying size, counting them, storing them and comparing them are easy tasks to perform by computers and unambiguous as exact matches are considered. The advantage of using *k*-mers is therefore the computational simplicity and speed. Another advantage that we highlighted in **chapter 3** is that *k*-mers allow to work without a reference genome, which is a huge advantage if no reference genome is available or if the reference genotype is distant from the study material. During my PhD, I became, however, quickly aware that the main advantage of working with *k*-mers is the simple concept of 'shared *k*-mers', which is a very powerful tool that can be used to answer a broad range of biological questions.

In **chapters 3 and 5**, we used the 'shared *k*-mers' concept, and even the 'non-shared *k*-mers' concept to increase the haplotype specificity of the designed markers. We indeed included susceptible varieties in the CoSSA workflow and removed from the resistant haplotype specific *k*-mers, the *k*-mers shared with the susceptible varieties. This simple step allows to quickly design diagnostic markers for use in MAS (Marker Assisted Selection). As far as I know, this step was novel and not included in other gene mapping strategies when we published CoSSA. Recently, the strategy developed by (89) included a similar step as three pools composed each of 15 varieties negative for the two targeted *R* genes were sequenced and used to increase the haplotype specificity of the designed KASP markers. In **chapters 3 and 5**, we also used the 'shared *k*-mers' tool to validate the presence of a specific haplotype in pedigree clones and panels of varieties and wild species. These experiments allowed us to trace back the resistance source to ancestral donors but also to quickly check if a variety hold the targeted *R* gene or not. We can imagine many other applications to the 'shared *k*-mers' tool such as for example studying the relatedness between species or varieties, or more specific tasks such as finding unique 21-mers in candidate *R* genes to design specific RNAi constructs for functional analyses. As more and more sequencing data are publicly available, I expect the *k*-mer databases to grow, for research as well as

commercial purposes, and the screening of molecular markers in the lab to be gradually replaced by *k*-mers as *in silico* tags.

Final remarks

Breeding for resistant varieties and their clever deployment, by taking into account the pathogen population composition, is the best weapon we have for disease management in general. For efficient breeding programs, potato breeders need knowledge about the different resistance sources that they can use, knowledge about how to use these resistances, and diagnostic molecular markers to trace the resistance in their material. In this thesis, several major and minor effect wart disease resistance loci were identified. We designed haplotype specific flanking markers and validated their diagnostic power in diversity panels representative of the potato breeding material. Moreover, we identified the different resistance sources already present in the potato breeding gene pool which is a valuable achievement as it will facilitate the combination of different resistant sources and avoid redundancy. The importance of combining major and minor effect loci in resistant varieties was highlighted. It is indeed necessary to bring full resistance to the higher pathotypes and avoid *S. endobioticum* population shifts. We showed that TNL genes are involved in resistance to *S. endobioticum* which brings new insights to better understand the host-pathogen interactions of this pathosystem. The work presented in this thesis also contributed to improve and simplify gene mapping strategies. The method we developed can be used for any trait and for any crop as it does not require a reference genome and it is suitable for polyploid genomes. Finally, this work brought new perspectives such as the improvement of the pathotyping system, which currently does not cover the potato *R* genes and the *S. endobioticum* *Avr* genes diversity, and the identification of unknown resistance sources in the breeding gene pool which will help to select suitable parents for future mapping and breeding strategies.

References

1. Hehl R, Faurie E, Hesselbach J, Salamini F, Whitham S, Baker B, *et al.* TMV resistance gene *N* homologues are linked to *Synchytrium endobioticum* resistance in potato. *Theor Appl Genet.* 1999;98(3):379–386.
2. Brugmans B, Hutten RGB, Rookmaker ANO, Visser RGF, van Eck HJ. Exploitation of a marker dense linkage map of potato for positional cloning of a wart disease resistance gene. *Theor Appl Genet.* 2006 Jan;112(2):269–77.
3. Plich J, Przetakiewicz J, Śliwka J, Flis B, Wasilewicz-Flis I, Zimnoch-Guzowska E. Novel gene *Sen2* conferring broad-spectrum resistance to *Synchytrium endobioticum* mapped to potato chromosome XI. *Theor Appl Genet.* 2018 Nov;131(11):2321–31.
4. Prodhomme C, Esselink D, Borm T, Visser RGF, van Eck HJ, Vossen JH. 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* [Internet]. 2019 Dec [cited 2019 May 30];15(1). Available from: <https://plantmethods.biomedcentral.com/articles/10.1186/s13007-019-0445-5>
5. Vos PG, Uitdewilligen JGAML, Voorrips RE, Visser RGF, van Eck HJ. Development and analysis of a 20K SNP array for potato (*Solanum tuberosum*): an insight into the breeding history. *Theor Appl Genet.* 2015 Dec;128(12):2387–401.
6. van Berloo R, Hutten RCB, van Eck HJ, Visser RGF. An Online Potato Pedigree Database Resource. *Potato Res.* 2007 Jan;50(1):45–57.
7. Bukasov SM. Les Bases de la Sélection de la Pomme de terre (suite et fin). *Rev Bot Appliquée Agric Colon.* 1940;20(223):179–89.
8. Lunden AP, Jørstad I. Investigations on the inheritance of immunity to wart disease (*Synchytrium Endobioticum* [Schilb.] Perc.) in the potato. *J Genet.* 1934;29(3):375–385.
9. EPPO. *Synchytrium endobioticum* distribution map. Consulted the 2019.10.18. [Internet]. 2019. Available from: <https://gd.eppo.int/taxon/SYNCEN/distribution>
10. Bukasov SM, Kameraz AJA. Bases of potato breeding. Gosudarstvennoe Izdatel'stvo Sel'skhozjajstvennoj Literatury, Moskva, Leningrad. 1959;528.
11. Ross H. Potato breeding - Problems and perspectives. Suppl. to *J. Plant Breeding*, Paul Parey, Berlin and Hamburg. 1986.
12. Rothacker D. Arbeiten zur Zuchtung krebseresistenter Kartoffeln. *Zücht.* 1957;
13. Rothacker D, Muller WA. Arbeiten zur Zuchtung krebseresistenter Kartoffeln. *Zücht.* 1960;
14. Khiutti AV, Antonova OYu, Mironenko NV, Gavrilenko TA, Afanasenko OS. Potato resistance to quarantine diseases. *Russ J Genet Appl Res.* 2017 Dec;7(8):833–44.
15. van Soest LJM, Seidewitz L. Evaluation data on tuber-bearing *Solanum* species. Inst. für Pflanzenbau und Pflanzenzüchtung der FAL-Stichting voor Plantenveredeling (SVP). 1981. 165 p.
16. Hardigan MA, Laimbeer FPE, Newton L, Crisovan E, Hamilton JP, Vaillancourt B, *et al.* Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication in the cultivated potato. *Proc Natl Acad Sci.* 2017 Nov 14;114(46):E9999–10008.
17. Hawkes JG. *The Potato, Evolution, Biodiversity and Genetic Resources.* Belhaven Press, London; 1990.
18. Maris B. Races of the potato wart causing fungus *Synchytrium endobioticum* (Schilb.) Perc. and some data on the inheritance of resistance to race 6. *Euphytica.* 1961;10(3):269–276.
19. Groth J, Song Y, Kellermann A, Schwarzfischer A. Molecular characterisation of resistance against potato wart races 1, 2, 6 and 18 in a tetraploid population of potato (*Solanum tuberosum* subsp. *tuberosum*). *J Appl Genet.* 2013 May;54(2):169–78.

20. Ballvora A, Flath K, Lübeck J, Strahwald J, Tacke E, Hofferbert H-R, *et al.* Multiple alleles for resistance and susceptibility modulate the defense response in the interaction of tetraploid potato (*Solanum tuberosum*) with *Synchytrium endobioticum* pathotypes 1, 2, 6 and 18. *Theor Appl Genet.* 2011 Dec;123(8):1281–92.
21. Obidiegwu JE, Sanetomo R, Flath K, Tacke E, Hofferbert H-R, Hofmann A, *et al.* Genomic architecture of potato resistance to *Synchytrium endobioticum* disentangled using SSR markers and the 8.3k SolCAP SNP genotyping array. *BMC Genet* [Internet]. 2015 Dec [cited 2017 Aug 3];16(1). Available from: <http://www.biomedcentral.com/1471-2156/16/38>
22. Flor HH. Current Status of the Gene-For-Gene Concept. *Annu Rev Phytopathol.* 1971 Sep;9(1):275–96.
23. Dodds PN, Lawrence GJ, Catanzariti A-M, Ayliffe MA, Ellis JG. The *Melampsora lini* AvrL567 Avirulence Genes Are Expressed in Haustoria and Their Products Are Recognized inside Plant Cells. *Plant Cell.* 2004 Mar;16(3):755–68.
24. Catanzariti A-M, Dodds PN, Lawrence GJ, Ayliffe MA, Ellis JG. Haustorially Expressed Secreted Proteins from Flax Rust Are Highly Enriched for Avirulence Elicitors. *Plant Cell.* 2006 Jan;18(1):243–56.
25. van de Vossen BTLH, Prodhomme C, van Arkel G, van Gent-Pelzer MPE, Bergervoe M, Brankovics B, *et al.* The *Synchytrium endobioticum* AvrSen1 Triggers a Hypersensitive Response in *Sen1* Potatoes While Natural Variants Evade Detection. *Mol Plant Microbe Interact.* 2019 Sep 17;MPMI-05-19-0138.
26. Jupe F, Witek K, Verweij W, Śliwka J, Pritchard L, Etherington GJ, *et al.* Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. *Plant J.* 2013 Nov;76(3):530–44.
27. Potato Genome Sequencing Consortium, Pan S, Cheng S, Zhang B, Mu D, Ni P, *et al.* Genome sequence and analysis of the tuber crop potato. *Nature.* 2011 Jul 10;475(7355):189–95.
28. Sharma SK, Bolser D, de Boer J, Sønderkjaer M, Amoros W, Carboni MF, *et al.* Construction of reference chromosome-scale pseudomolecules for potato: integrating the potato genome with genetic and physical maps. *G3 Genes Genomes Genet.* 2013;3(11):2031–2047.
29. Bartkiewicz A, Chilla F, Terefe-Ayana D, Lübeck J, Strahwald J, Tacke E, *et al.* Improved genetic resolution for linkage mapping of resistance to potato wart in monoparental dihaploids with potential diagnostic value in tetraploid potato varieties. *Theor Appl Genet.* 2018 Dec;131(12):2555–66.
30. Vidal S, Cabrera H, Andersson RA, Fredriksson A, Valkonen JPT. Potato Gene *Y-1* is an *N* Gene Homolog That Confers Cell Death Upon Infection with *potato virus Y*. *Mol Plant Microbe Interact.* 2002 Jul;15(7):717–27.
31. Brigneti G, Garcia-Mas J, Baulcombe DC. Molecular mapping of the potato virus Y resistance gene *Rysto* in potato: *Theor Appl Genet.* 1997 Feb;94(2):198–203.
32. Paal J, Henselewski H, Muth J, Meksem K, Menéndez CM, *et al.* Molecular cloning of the potato *Gro1-4* gene conferring resistance to pathotype Ro1 of the root cyst nematode *Globodera rostochiensis*, based on a candidate gene approach. *Plant J.* 2004 Apr;38(2):285–97.
33. Jacobs MMJ, Vosman B, Vleeshouwers VGAA, Visser RGF, Henken B, van den Berg RG. A novel approach to locate *Phytophthora infestans* resistance genes on the potato genetic map. *Theor Appl Genet.* 2010 Feb;120(4):785–96.
34. Verzaux E, van Arkel G, Vleeshouwers VGAA, van der Vossen EAG, Niks RE, Jacobsen E, *et al.* High-Resolution Mapping of Two Broad-Spectrum Late Blight Resistance Genes from Two Wild Species of the *Solanum circaefolium* Group. *Potato Res.* 2012 Jun;55(2):109–23.
35. Tan AMY, Park T-H, Alles R, Hutten RCB, Visser RGF, van Eck HJ. *GpaXI tar I* originating from *Solanum tarijense* is a major resistance locus to *Globodera pallida* and is localised on chromosome 11 of potato. *Theor Appl Genet.* 2009 Nov;119(8):1477–87.

36. Lawrence GJ, Dodds PN, Ellis JG. TECHNICAL ADVANCE: Transformation of the flax rust fungus, *Melampsora lini*: selection via silencing of an avirulence gene: Transformation of flax rust. *Plant J*. 2009 Nov 13;61(2):364–9.
37. Cullis CA. Disease Resistance Genes in Flax. In: Cullis CA, editor. *Genetics and Genomics of Linum* [Internet]. Cham: Springer International Publishing; 2019 [cited 2019 Sep 8]. p. 215–25. Available from: http://link.springer.com/10.1007/978-3-030-23964-0_13
38. Sekhwal M, Li P, Lam I, Wang X, Cloutier S, You F. Disease Resistance Gene Analogs (RGAs) in Plants. *Int J Mol Sci*. 2015 Aug 14;16(8):19248–90.
39. Aarts N, Metz M, Holub E, Staskawicz BJ, Daniels MJ, Parker JE. Different requirements for EDS1 and NDR1 by disease resistance genes define at least two *R* gene-mediated signaling pathways in Arabidopsis. *Proc Natl Acad Sci*. 1998 Aug 18;95(17):10306–11.
40. Century KS, Holub EB, Staskawicz BJ. NDR1, a locus of *Arabidopsis thaliana* that is required for disease resistance to both a bacterial and a fungal pathogen. *Proc Natl Acad Sci USA*. 1995;5.
41. Wiermer M, Feys BJ, Parker JE. Plant immunity: the EDS1 regulatory node. *Curr Opin Plant Biol*. 2005 Aug;8(4):383–9.
42. Adachi H, Derevnina L, Kamoun S. NLR singletons, pairs, and networks: evolution, assembly, and regulation of the intracellular immunoreceptor circuitry of plants. *Curr Opin Plant Biol*. 2019 Aug;50:121–31.
43. Eitas TK, Dangl JL. NB-LRR proteins: pairs, pieces, perception, partners, and pathways. *Curr Opin Plant Biol*. 2010 Aug 1;13(4):472–7.
44. Wu C-H, Belhaj K, Bozkurt TO, Birk MS, Kamoun S. Helper NLR proteins NRC2a/b and NRC3 but not NRC1 are required for Pto-mediated cell death and resistance in *Nicotiana benthamiana*. *New Phytol*. 2016 Mar;209(4):1344–52.
45. Wu Z, Li M, Dong OX, Xia S, Liang W, Bao Y, *et al*. Differential regulation of TNL-mediated immune signaling by redundant helper CNLs. *New Phytol*. 2019 Apr;222(2):938–53.
46. Gabriëls SHEJ, Vossen JH, Ekengren SK, Ooijen G van, Abd-El-Halim AM, Berg GCM van den, *et al*. An NB-LRR protein required for HR signalling mediated by both extra- and intracellular resistance proteins: NRC1 is required for HR signaling. *Plant J*. 2007 Mar 5;50(1):14–28.
47. Wu C-H, Abd-El-Halim A, Bozkurt TO, Belhaj K, Terauchi R, Vossen JH, *et al*. NLR network mediates immunity to diverse plant pathogens. *Proc Natl Acad Sci*. 2017 Jul 11;201702041.
48. Castel B, Ngou P, Cevik V, Redkar A, Kim D, Yang Y, *et al*. Diverse NLR immune receptors activate defence via the RPW 8- NLR NRG 1. *New Phytol*. 2019 Apr;222(2):966–80.
49. Qi T, Seong K, Thomazella DPT, Kim JR, Pham J, Seo E, *et al*. NRG1 functions downstream of EDS1 to regulate TIR-NLR-mediated plant immunity in *Nicotiana benthamiana*. *Proc Natl Acad Sci*. 2018 Nov 13;115(46):E10979–87.
50. Bonardi V, Tang S, Stallmann A, Roberts M, Cherkis K, Dangl JL. Expanded functions for a family of plant intracellular immune receptors beyond specific recognition of pathogen effectors. *Proc Natl Acad Sci*. 2011 Sep 27;108(39):16463–8.
51. Collier SM, Hamel L-P, Moffett P. Cell Death Mediated by the N-Terminal Domains of a Unique and Highly Conserved Class of NB-LRR Protein. *Mol Plant Microbe Interact*. 2011 Aug;24(8):918–31.
52. Grech-Baran M, Witek K, Szajko K, Witek AI, Morgiewicz K, Wasilewicz-Flis I, *et al*. Extreme resistance to *Potato virus Y* in potato carrying the *R_{Ysto}* gene is mediated by a TIR - NLR immune receptor. *Plant Biotechnol J* [Internet]. 2019 Sep 4 [cited 2019 Oct 2]; Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/pbi.13230>
53. Wang J, Hu M, Wang J, Qi J, Han Z, Wang G, *et al*. Reconstitution and structure of a plant NLR resistosome conferring immunity. *Science*. 2019 Apr 5;364(6435):eaav5870.

54. Qi D, Innes RW. Recent Advances in Plant NLR Structure, Function, Localization, and Signaling. *Front Immunol* [Internet]. 2013 [cited 2019 Oct 9];4. Available from: <http://journal.frontiersin.org/article/10.3389/fimmu.2013.00348/abstract>
55. Tameling WIL, Nooijen C, Ludwig N, Boter M, Slootweg E, Goverse A, *et al.* RanGAP2 Mediates Nucleocytoplasmic Partitioning of the NB-LRR Immune Receptor Rx in the Solanaceae, Thereby Dictating Rx Function. *Plant Cell*. 2010 Dec;22(12):4176–94.
56. Slootweg E, Roosien J, Spiridon LN, Petrescu A-J, Tameling W, Joosten M, *et al.* Nucleocytoplasmic Distribution Is Required for Activation of Resistance by the Potato NB-LRR Receptor Rx1 and Is Balanced by Its Functional Domains. *Plant Cell*. 2010 Dec;22(12):4195–215.
57. Takemoto D, Rafiqi M, Hurley U, Lawrence GJ, Bernoux M, Hardham AR, *et al.* N-Terminal Motifs in Some Plant Disease Resistance Proteins Function in Membrane Attachment and Contribute to Disease Resistance. *Mol Plant Microbe Interact*. 2012 Mar;25(3):379–92.
58. Larkan NJ, Lydiate DJ, Parkin IAP, Nelson MN, Epp DJ, Cowling WA, *et al.* The *Brassica napus* blackleg resistance gene *LepR3* encodes a receptor-like protein triggered by the *Leptosphaeria maculans* effector AVRML1. *New Phytol*. 2013 Jan;197(2):595–605.
59. Parlange F, Daverdin G, Fudal I, Kuhn M-L, Balesdent M-H, Blaise F, *et al.* *Leptosphaeria maculans* avirulence gene *AvrLm4-7* confers a dual recognition specificity by the *Rlm4* and *Rlm7* resistance genes of oilseed rape, and circumvents *Rlm4* -mediated recognition through a single amino acid change. *Mol Microbiol*. 2009 Feb;71(4):851–63.
60. Delourme R, Pilet-Nayel ML, Archipiano M, Horvais R, Tanguy X, Rouxel T, *et al.* A Cluster of Major Specific Resistance Genes to *Leptosphaeria maculans* in *Brassica napus*. *Phytopathology*. 2004 Jun;94(6):578–83.
61. Aguilera-Galvez C, Champouret N, Rietman H, Lin X, Wouters D, Chu Z, *et al.* Two different *R* gene loci co-evolved with *Avr2* of *Phytophthora infestans* and confer distinct resistance specificities in potato. *Stud Mycol*. 2018 Mar;89:105–15.
62. Grant M, Godiard L, Straube E, Ashfield T, Lewald J, Sattler A, *et al.* Structure of the Arabidopsis *RPM1* gene enabling dual specificity disease resistance. *Science*. 1995 Aug 11;269(5225):843–6.
63. Mackey D, Holt BF, Wiig A, Dangl JL. RIN4 Interacts with *Pseudomonas syringae* Type III Effector Molecules and Is Required for RPM1-Mediated Resistance in Arabidopsis. *Cell*. 2002 Mar;108(6):743–54.
64. Petit-Houdenot Y, Fudal I. Complex Interactions between Fungal Avirulence Genes and Their Corresponding Plant Resistance Genes and Consequences for Disease Resistance Management. *Front Plant Sci* [Internet]. 2017 Jun 16 [cited 2019 Oct 10];8. Available from: <http://journal.frontiersin.org/article/10.3389/fpls.2017.01072/full>
65. Van Der Biezen EA, Jones JDG. Plant disease-resistance proteins and the gene-for-gene concept. *Trends Biochem Sci*. 1998 Dec;23(12):454–6.
66. Van der Hoorn R. Balancing selection favors guarding resistance proteins. *Trends Plant Sci*. 2002 Feb 1;7(2):67–71.
67. Spieckermann A, Kothoff P. Testing potatoes for wart resistance. *Deutsche Landwirtschaftliche Presse* 51:114–115. 1924;
68. Glynn MD. Infection experiments with wart disease of potatoes. *Synchytrium endobioticum* (Schilb.) Perc. *Ann Appl Biol*. 1925;12(1):34–60.
69. Lemmerz J. A new simplified infection procedure for testing potato cultivars for wart resistance. *Zücher*. 1930;2(288):97.
70. Nazarian-Firouzabadi F, Joshi S, Xue H, Kushalappa AC. Genome-wide *in silico* identification of LysM-RLK genes in potato (*Solanum tuberosum* L.). *Mol Biol Rep*. 2019;46(5):5005–17.
71. Niks RE, Qi X, Marcel TC. Quantitative Resistance to Biotrophic Filamentous Plant Pathogens: Concepts, Misconceptions, and Mechanisms. *Annu Rev Phytopathol*. 2015 4;53(1):445–70.

72. Poland JA, Balint-Kurti PJ, Wisser RJ, Pratt RC, Nelson RJ. Shades of gray: the world of quantitative disease resistance. *Trends Plant Sci.* 2009 Jan;14(1):21–9.
73. van de Vossen BTLH, Brankovics B, Nguyen HDT, van Gent-Pelzer MPE, Smith D, Dadej K, *et al.* The linear mitochondrial genome of the quarantine chytrid *Synchytrium endobioticum*; insights into the evolution and recent history of an obligate biotrophic plant pathogen. *BMC Evol Biol* [Internet]. 2018 Dec [cited 2019 Mar 5];18(1). Available from: <https://bmcevolbiol.biomedcentral.com/articles/10.1186/s12862-018-1246-6>
74. Baayen RP, Cochius G, Hendriks H, Meffert JP, Bakker J, Bekker M, *et al.* History of potato wart disease in Europe – a proposal for harmonisation in defining pathotypes. *Eur J Plant Pathol.* 2006 Sep;116(1):21–31.
75. EPPO. *Synchytrium endobioticum*: Diagnostic protocols. EPPO Bull. 2004 Aug;34(2):213–8.
76. EPPO. PM 7/28 (2) *Synchytrium endobioticum*. EPPO Bull. 2017 Dec;47(3):420–40.
77. Zhu S, Vossen JH, Bergervoet M, Nijenhuis M, Kodde L, Kessel GJT, *et al.* An updated conventional- and a novel GM potato late blight *R* gene differential set for virulence monitoring of *Phytophthora infestans*. *Euphytica.* 2015 Mar;202(2):219–34.
78. Kamphuis LG, Klingler JP, Jacques S, Gao L-L, Edwards OR, Singh KB. Additive and epistatic interactions between *AKR* and *AIN* loci conferring bluegreen aphid resistance and hypersensitivity in *Medicago truncatula*. *J Exp Bot.* 2019 Sep 24;70(18):4887–902.
79. MacIntosh GC. Gene pyramids and the balancing act of keeping pests at bay. *J Exp Bot.* 2019 Sep 24;70(18):4591–3.
80. Rouppe van der Voort J, van der Vossen E, Bakker E, Overmars H, van Zandvoort P, Hutten R, *et al.* Two additive QTLs conferring broad-spectrum resistance in potato to *Globodera pallida* are localized on resistance gene clusters: *Theor Appl Genet.* 2000;101(7):1122–30.
81. Uitdewilligen JGAML, Wolters A-MA, D’hoop BB, Borm TJA, Visser RGF, van Eck HJ. A Next-Generation Sequencing Method for Genotyping-by-Sequencing of Highly Heterozygous Autotetraploid Potato. Lukens L, editor. *PLoS ONE.* 2013 May 8;8(5):e62355.
82. Hamilton JP, Hansey CN, Whitty BR, Stoffel K, Massa AN, Van Deynze A, *et al.* Single nucleotide polymorphism discovery in elite North American potato germplasm. *BMC Genomics.* 2011;12(1):302.
83. Arora S, Steuernagel B, Gaurav K, Chandramohan S, Long Y, Matny O, *et al.* Resistance gene cloning from a wild crop relative by sequence capture and association genetics. *Nat Biotechnol.* 2019 Feb;37(2):139–43.
84. Witek K, Jupe F, Witek AI, Baker D, Clark MD, Jones JDG. Accelerated cloning of a potato late blight-resistance gene using RenSeq and SMRT sequencing. *Nat Biotechnol.* 2016 Apr 25;34(6):656–60.
85. Rahman A, Hallgrímsdóttir I, Eisen M, Pachter L. Association mapping from sequencing reads using *k*-mers. *eLife.* 2018;7:1–31.
86. Vossen EAG, Gros J, Sikkema A, Muskens M, Wouters D, Wolters P, *et al.* The *Rpi-blb2* gene from *Solanum bulbocastanum* is an *Mi-1* gene homolog conferring broad-spectrum late blight resistance in potato: Isolation of the late blight resistance gene *Rpi-blb2*. *Plant J.* 2005 Sep 27;44(2):208–22.
87. Li G, Huang S, Guo X, Li Y, Yang Y, Guo Z, *et al.* Cloning and Characterization of *R3b*; Members of the *R3* Superfamily of Late Blight Resistance Genes Show Sequence and Functional Divergence. *Mol Plant Microbe Interact.* 2011 Oct;24(10):1132–42.
88. Vossen JH, van Arkel G, Bergervoet M, Jo K-R, Jacobsen E, Visser RGF. The *Solanum demissum* *R8* late blight resistance gene is an *Sw-5* homologue that has been deployed worldwide in late blight resistant varieties. *Theor Appl Genet.* 2016 Sep;129(9):1785–96.
89. Meade F, Byrne S, Griffin D, Kennedy C, Mesiti F, Milbourne D. Rapid Development of KASP Markers for Disease Resistance Genes Using Pooled Whole-Genome Resequencing. *Potato Res* [Internet]. 2019 Jul 6 [cited 2019 Nov 18]; Available from: <http://link.springer.com/10.1007/s11540-019-09428-x>

Appendices

Summary
Acknowledgements
About the author
Education statement

Summary

Potato wart disease, caused by the obligate biotrophic Chytrid fungus *Synchytrium endobioticum*, is one of the most important quarantine diseases of potato. This disease was named after the symptoms caused by the pathogen, which are the proliferation of meristematic tissues leading to the formation of warts, mainly on the below-ground sprouts of potato plants. The quarantine status of *S. endobioticum* is due to the production of spores that can remain viable in the soil for more than 40 years, the lack of chemical control and the severe yield losses. In Europe, more than 40 different pathotypes of *S. endobioticum* have been recorded and only resistance to pathotype 1 is commonly deployed in the breeding germplasm. The breeding and cultivation of potato varieties resistant to a wider spectrum of pathotypes is crucial for quarantine practice to reduce the propagation of the pathogen. Therefore, the identification of genes bringing resistance to the most frequent pathotypes of the pathogen and the development of diagnostic markers for marker assisted selection (MAS) is urgently needed.

In my PhD thesis, genes involved in resistance to pathotypes 1, 2, 6 and 18 of *S. endobioticum* were identified to make an inventory of the different resistance sources at hand for potato breeders. In the first part of my thesis, we investigated the distribution of the pathotype 1 resistance in a variety panel representative of the potato breeding material (**Chapter 2**). Breeding programs of the 20th century were very successful in producing varieties resistant to pathotype 1 as 77% of the panel varieties were found to be resistant. To identify markers linked with pathotype 1 resistance, we used previously produced genotypic and phenotypic data to perform a Genome-Wide Association Study (GWAS). The GWAS resulted in the identification of markers associated with pathotype 1 resistance on the north arm of chromosome 11. In this region, the major effect gene *Sen1* was previously identified. *Sen1* is the main source of pathotype 1 resistance in the variety panel and no common ancestral donor could be identified due to the inability to define identity-by-descent (IBD). As we faced limitations to design markers fully diagnostic for pathotype 1 resistance using the GWAS approach, we aimed to develop new tools to identify haplotype specific SNPs. In **Chapter 3**, we developed a new set of workflows, called Comparative Subsequence Sets Analysis (CoSSA), for the genetic analysis of traits of interest and the identification of haplotype specific SNPs. CoSSA can be used for any crop as it is suitable for polyploids and can be used with or without a reference genome. We applied CoSSA to identify *Sen3*, a dominant gene conferring resistance to all tested pathotypes. *Sen3*

was fine-mapped to the resistance gene cluster C76 on the north arm of chromosome 11. Furthermore, we used CoSSA for the fine-mapping of *Sen1* (**Chapter 4**). *Sen1* was mapped to the same *R* gene cluster as *Sen3*. We performed a candidate gene analysis and showed that *Sen1* encodes a nucleotide-binding domain, leucine rich containing (NLR) protein from the TNL group. The two identified candidate genes were cloned and tested in complementation assays with AvrSen1, the *S. endobioticum* effector protein which triggers Hypersensitive Responses (HR) in *Sen1* plants. These findings will serve as novel tools to study the interactions between potato and *S. endobioticum*.

In the last part of my thesis, we made an, as complete as possible at this moment, inventory of the dominant potato wart disease resistance (*Sen*) genes and QTLs present in the potato breeding germplasm. We combined the GWAS and CoSSA strategies to identify two new major genes, *Sen4* and *Sen5*, respectively located on the south arm of chromosome 12 and the south arm of chromosome 5, which are involved in resistance to pathotypes 2, 6 and 18 (**Chapter 5**). We also identified several wart disease resistance QTLs which, in combination with the dominant genes, can contribute to improve resistance to the higher pathotypes. To avoid any confusion between the previously and newly identified QTLs, we introduced a new naming system which allows to differentiate each resistant haplotype identified. Finally, we screened a broad panel of potato varieties and wild *Solanum* species for the genes *Sen1*, *Sen2*, *Sen3*, *Sen4* and *Sen5*. As was already observed in **Chapter 2**, *Sen1* was highly frequent in the variety panel and it could not be traced back to a common ancestral donor. This high frequency of *Sen1* in the breeding genepool might be explained by the fact that it was present in accessions from the Chilotanum group and from *S. microdontum*, which were intensively used to adapt cultivated potato to long days conditions. The main source of resistance to the higher pathotypes in the Dutch breeding material is *Sen4*, which is IBD in every Dutch variety and can be traced back to AM78-3704. The introgression of *Sen4* in the breeding material has not been solved yet, although we provide some hypotheses. The main source of resistance to the higher pathotypes in the German and Polish material is *Sen3* which is IBD in every positive variety and can be traced back to the breeding clone BRA9089. *Sen5*, which can be traced back to the breeding clone VE71-105, is rather scarce in breeding material and originates from *S. vernei*. *Sen2*, which is also scarce in the breeding material, probably originates from *S. acaule*, *S. demissum* or *S. verrucosum*.

To put it in a nutshell, a complete picture of the major potato wart disease resistance sources present in the breeding germplasm is given in my thesis. Haplotype specific

markers have been designed for all the major genes and QTLs mapped, which will facilitate the breeding of resistant varieties. Finally, the development of CoSSA will facilitate the mapping of traits of interest and the design of haplotype specific markers for any crop.

Acknowledgements

What an adventure! Let me tell you that writing a PhD thesis is no bed of roses, and I would not have achieved this one without the help and support of many people. During the long hours of writing (when I was feeling like falling for the dark side), I often dreamt of writing the acknowledgements, but this is not as easy as it seemed. I hope that I will succeed in expressing on paper my gratitude to all of you who helped me with this achievement. If your name is not on the list, be assured that you also have my special thanks, but my memory is currently saturated with potatoes, warts, and their very intimate relationship.

First, I would like to express my special thanks to my two supervisors, **Jack** and **Herman**. During these four years, I got to know you both and realized how lucky I was to be supervised by two complementary personalities. Jack, although it took me some adaptation time at the beginning of my PhD, I really enjoyed your supervision style that left me a lot of freedom to pursue the leads I wanted, the way I wanted. Our weekly meetings were a source of inspiration for me and I learned from you to try to think always a few steps ahead. Also, thank you for encouraging me to share my work in international conferences, which is always a good opportunity to collect new ideas (and to taste new delicacies, such as the durian!). Herman, thank you for sharing with me your passion for puzzling with markers data and for digging in potato pedigrees. I was always very impressed by your knowledge about the potato's breeding history and genetics. I tried to absorb as much of it as I could during these four years, and I often wished I could use the Force to copy/paste your knowledge in my own brain.

Richard, we first met during my job interview, almost 5 years ago, and I was pleasantly surprised to see how approachable you were. You directly put me at ease and it helped me to reduce my stress! Thank you for attending our monthly project meetings, your questions and suggestions were always valuable. I also very much appreciated your contributions during my writing phase, always on time!

Gert, thank you so much for your strong support in the lab and in the greenhouse, where I always felt like being an alien landing from another planet. I learned a lot from you and I was always impressed by your efficiency (I tried very hard but could never keep up with your rhythm during our agroinfiltration experiments!). **Marjan**, you did a great job with the maintenance of our *in vitro* collection. Thank you also for your valuable help with the agroinfiltration experiments. This project would not have been so successful without you both.

A big “merci” goes to **Danny** who introduced me to the world of bioinformatics. Danny, thank you so much for your availability and for answering my questions always on the spot. I am not sure my computer would have survived my PhD in one piece if you had not been there to find the mistakes in my scripts!

The main asset of this project was the close collaboration between the Plant Breeding and the Biointeractions teams. **Theo**, I really enjoyed our discussions every time we met in the corridor, and I am still smiling! **Bart**, I am really amazed about your skills in time management. I sometimes thought you were an army of clones to be able to deal with your PhD, your job at the NPPO, your family and your numerous hobbies. This team would not be complete without **Marga** who did an amazing job on the pathogen’s side of the project. **Dong**, it was a pleasure to welcome you in the wart disease team. I wish you success with the future developments of the project and with your PhD.

A few other potato people deserve to be acknowledged. **Ronald**, thank you for gathering tubers from the genotypes I needed every year. That was not an easy job as I was often interested in very old clones! **Christel**, it was a pleasure to travel with you to the EAPR meetings. **Dirk Jan**, thank you for being so friendly with Maxime and I from the very first day of my PhD.

Joao, your expertise with GWAS and Genstat was essential for the first chapter of my thesis. **Richard F.**, **Theo B.** and **Martijn**, thank you for helping me with some aspects of bioinformatics and with domesticating the server.

During these four years, I got the chance to supervise quite a diverse bunch of students. **Jasper**, **Johan**, **Moffat**, **Xizheng**, **Michiel** and **Joop**, thank you all for your great contribution to this project. I learned a lot from each one of you and wish you all success in your future career.

One of the factors that made me apply for this PhD position was the close cooperation with the breeding companies. I would like to thank **Averis seeds**, **BNA**, **Danespo**, **HLB**, **HZPC Holland**, **C Meijer**, **SaKa**, **Teagasc** and their staff for their precious assistance and for making this project feasible. Thank you all for your contribution and suggestions during our Consortium meetings, for always answering my questions and for digging in your tubers stocks to find the genotypes I needed.

Dianka, it was a great pleasure for me to share my office with you during these four years. Thank you for being so welcoming and friendly from the very beginning, you helped me feel at home in PBR. **Jos**, you joined our office almost half way through my PhD and I am glad that you did it! I enjoyed our books chats and enjoyed when you

shared your holiday pictures with us. You were both great office mates and I already miss office E1.149.

Every PhD student would be lost without the PBR team of technicians: thank you all for helping us and answering our (too many) questions. I wish especially to acknowledge **Linda, Doret, Bernadette, Marian, Fien, Annelies, Irma** and **Danny**. Our great secretaries **Nicole, Letty** and **Daniëlle**, also deserve special thanks for their excellent work and for always being so cheerful.

This adventure would not have been possible without my two wonderful paranymphs and precious friends, **Mas** and **Pauline**. Thank you both for guiding me the past few months and for helping me with the defense. These four years would not have been the same without you Mas, and it is impossible to describe in a few words all the happy moments we shared together: squash, shopping, cakes, zoo, Dutch classes, Valencia, sushi, Paris, gossip, board games and so much more. Pauline, I am so happy that we started and finished our PhD together; your mindset helped me to put things into perspective during stressful times. I already miss your company and your facepalm in reaction to someone's jokes!

Jasper, I miss sharing my lunches with you, thank you for your support during the last months of my PhD, I send you all my positive energy to finish yours. **Jarst**, you'll be happy to know that I can't hear a joke anymore without imagining your satisfied joke face! **Johan**, the corridors feel terribly empty in my new job without your recognizable pace. I will miss your positive vibes and your numerous stories (especially the Sinterklaas one...).

Mathilde, we became quite close friends during our PhD, supporting each other during our difficult moments. You are a strong and hard-working person, and I hope that you will find a job that will allow you to blossom and express your creativity. **Mengjing**, you managed to deliver both a great thesis and a fantastic baby within four years! You are a very positive and friendly person; I am glad we worked on our PhD together.

Vivi, Behzad and **Marine**, I will keep very good memories of our long weekend in Paris. That was a lot of fun in very good company. Let's try to organize other trips together in the future!

Great thanks should go also to all my fellow PBR colleagues with whom I shared coffees, drinks, food, WE days, Christmas dinners, escape games, concerts, lunches, music, TV series, hugs, conferences, corridor chats, board games, car rides, tuk-tuk rides, lightsaber fights, carrots (Mildiou thanks you too), lab benches and so much

more: **Anne, Aina, Andres, Arwa, Carolina, Cynara, Daniel, Ehsan, Eleni, Eric, Ernst, Geert, Gurnoor, Henk, Jan-Kees, Jeroen, Johan W., Jordi, Katharina, Kim, Michiel, Miguel, Peter B., René, Robert, Sara, Sri...**

To all the French people I met in Wageningen, **Marine, Mathilde, Aurélie, Fanny, Gaby, Corentin, Thibault** and **Séverine** : c'était toujours un plaisir de pouvoir parler Français autour d'un café ou d'un verre de rosé, dans un couloir ou sur un poney !

To the person who secretly slipped a potato-shaped brush in my bag on Sinterklaas day, thank you. I still don't know who you are, but you should know that I love this brush and the mystery behind it.

Mirjan and **Theo**, you were the first Dutch friends we made when we moved to the Netherlands. Thank you so much for being so welcoming and friendly when we were new kids on the block. You helped us to feel at home in the Netherlands. We had a lot of fun going to the football game NL vs FR with you. We are looking forward to your visit in Bretagne!

And at last, but not least, I would like to address my warm thanks to my family for their strong support during these four years. Merci à tous ceux qui sont venus nous rendre visite (**Anne-Sophie, Tonton Seb, Jérôme, Pascale, Jérémy, Gladys, Jules, Lisa, Anaïs, Caroline**), quel plaisir de vous faire découvrir notre quotidien ! Un énorme merci en plus pour Tonton Seb qui a peint la sublime peinture de ma couverture de thèse. **Papa, Maman**, merci d'avoir fait tant de fois l'aller-retour Ouistreham-Wageningen pour venir nous voir, surtout pendant les premières années qui furent les plus difficiles. Merci pour les apéros Skype pendant les longs mois d'hiver et pour nous avoir chouchoutés (merci également à **Papy** et **Mamie** pour ça) à chaque fois que nous rentrions. Papa, merci pour ce magnifique texte que tu as écrit.

Maxime, merci de m'avoir suivie dans cette folle aventure. T'avoir à mes côtés m'a donné le courage de me dépasser et de donner le meilleur de moi-même pendant ces quatre ans. Tu m'as suivie sans savoir où tu allais, et je suis fière du chemin que tu as parcouru, et que nous avons parcouru. Merci pour ton soutien infailible.

Merci à tous,

Charlotte

About the author

'Twas a long time ago (but definitely NOT in a galaxy far far away...)

Longer now than it seems

In a place that perhaps you've seen in your dreams

For the story you're about to be told

Took place in the scientific world of old

Now you've probably wondered where Charlotte comes from

If you haven't, I'd say it's time you begun.....

Charlotte (what a well-chosen name for someone destined to study potatoes...!) could have been born on a Thursday, on Thursday September 12th 1991 to be exact, had her parents not decided to wait until the following day because they had this whimsical vision of life at the time and thought: "Hey, why not have our first child on Friday 13th? Wouldn't that be great? And wouldn't it be even greater to have a daughter?" So be it:

Charlotte Prodhomme was born on Friday September 13th 1991 in the beautiful Norman town of Caen, Calvados, home of the world-renowned 'Tripes à-la-mode-de-Caen' – don't believe those jealous people who'll tell you that it's just a stew made of cows' guts and carrots – and of William the Conqueror who more than 10 centuries before Brexit had already figured out that those British Isles were nothing but trouble...!

She was a lovely baby, really, and she passed from infancy to childhood in an atmosphere of bedtime stories, Walt Disney movies, horse riding lessons, piano lessons, vacation in Spain with our best friends, more bedtime stories, more Walt Disney movies, vacation in Arcachon with our best friends, more and more horse riding lessons and even more Walt Disney movies...! Needless to say, she was a bright and hardworking student who was already convinced that school was important because she was nursing the dream of becoming a vet specialized in horses (how surprising...!)

And then came high-school... and adolescence...!! Charlotte went to "VH" High in Caen (not Van Halen, folks... not yet at least... but Victor Hugo, the all-French literary genius who wrote Notre Dame de Paris... more Walt Disney movies all over again...!) where she majored in science: math (because she had to), physics (because she couldn't do otherwise) and biology (now you're talking!). Unfortunately (well, that's a matter of opinion...), she had a bad experience with a teacher who told her that being a vet was too difficult for her. Nevertheless, she graduated brilliantly in 2009 with a "Baccalauréat Série S avec Mention Bien" (the French Sesame to go to college) with which she decided to spend 2 months in Medical School, hardly enough time to become a doctor, but plenty enough time to realize that "that's not the way I want to study...! Those people are nuts..!"

College became the ideal place of great accomplishments for young-adult Charlotte: Bachelor Degree in Biology and Ecology from 2010 to 2013 in Caen with honors, Master Degree in Plant Breeding from 2013 to 2015 in Rennes again with honors during which she became more and more interested in potatoes, an internship in Limagrain in Clermont-Ferrand for her second year of Master Degree and last but not least, her PhD project in Wageningen which started in October 2015. The writing of her thesis which represents almost 5 years of her life is her greatest achievement so far. As a rigorous and demanding student, I know that she's too humble to brag so let me do it for her: she did a hell of a great job..!

From Normandy to Brittany, from Auvergne to The Netherlands, and today back in Brittany as an engineer in potato genetics for FN3PT/RD3PT since January 2020, Charlotte has always been this joyful and dedicated person who threw herself into her studies with great enthusiasm and yet remained true to herself. Today, she's an accomplished scientist who can look back on her college background with pride. Her supportive family is extremely proud of her and I am convinced that it's only the beginning of a great adventure.

*You may say I'm a dreamer
But I'm not the only one
I know one day you'll join her
And the world will be a better one*

To my beloved daughter

Dad

Education Statement of the Graduate School Experimental Plant Sciences

Issued to: Charlotte Prodhomme
Date: 17 April 2020
Group: Plant Breeding
University: Wageningen University & Research

1) Start-Up Phase	<u>date</u>
<p>► First presentation of your project</p> <p>Monday Seminar - An integrated genomics and effectomics impulse for potato wart disease resistance management and breeding</p>	22 Feb 2016
<p>► Writing or rewriting a project proposal</p> <p>An integrated genomics and effectomics impulse for potato wart disease resistance management and breeding</p>	18 Dec 2015
<i>Subtotal Start-Up Phase</i>	2.5
2) Scientific Exposure	<u>date</u>
<p>► EPS PhD student days</p> <p>Get2Gether event of 2016</p> <p>Get2Gether event of 2017</p>	<p>28-29 Jan 2016</p> <p>09-10 Feb 2017</p>
<p>► EPS theme symposia</p> <p>EPS Theme 4 Symposium 'Genome Biology', Amsterdam</p> <p>EPS Theme 4 Symposium 'Genome Biology', Wageningen</p> <p>EPS Theme 2 & Willie Commelin Scholten Symposium 'Interactions between plants and biotic agents', Wageningen</p>	<p>15 Dec 2015</p> <p>16 Dec 2016</p> <p>23 Jan 2017</p>
<p>► Lunteren Days and other national platforms</p> <p>Annual meeting Experimental Plant Sciences, Lunteren</p> <p>Annual meeting Experimental Plant Sciences, Lunteren</p>	<p>11-12 Apr 2016</p> <p>10-11 Apr 2017</p>
<p>► Seminars (series), workshops and symposia</p> <p>Seminar: Prof.dr. Jane Parker. Resistance pathway dynamics in plant immunity</p> <p>Seminar: Professor Caitlyn Allen. How Ralstonia solanacearum succeeds in plant xylem vessels</p> <p>Seminar: Professor Wenbo Ma. Effectors as molecular probes to understand pathogenesis</p> <p>Seminar: Dr. Jeffrey Endelman. Automated tetraploid genotype calling and its application to pedigree reconstruction in potato</p> <p>Seminar: Timothy Bilton. Statistical methods for analysing low-coverage GBS data</p> <p>Seminar: NRGene GenoMAGIC™: applied genomics for breeding and research</p> <p>Consortium meeting: TKI project</p> <p>Consortium meeting: TKI project</p> <p>Consortium meeting: TKI project</p> <p>Consortium meeting: TKI project</p>	<p>21 Jan 2016</p> <p>29 Apr 2016</p> <p>10 Jun 2016</p> <p>16 Nov 2016</p> <p>04 Jul 2018</p> <p>13 Sep 2018</p> <p>12 Feb 2016</p> <p>21 Jun 2016</p> <p>08 Nov 2016</p> <p>04 Apr 2017</p>

Consortium meeting: TKI project	07 Nov 2017
Consortium meeting: TKI project	14 Mar 2018
Consortium meeting: TKI project	03 Dec 2018
Consortium meeting: TKI project	28 Nov 2019
Symposium: WURomics Symposium 2016	15 Dec 2016
Symposium: Mini-symposium applied phytopathology, from the lab to the field	01 Mar 2017
Symposium: Mini-symposium EPS Polyploid genetics and breeding	14 Jun 2018
Symposium: Minisymposium on genetics and genomics of polyploids	06 Nov 2019
Public lecture: "Rewriting our genes" with Jennifer Doudna and Edze Westra	30 Sep 2016
► International symposia and congresses	
EAPR Eucarpia - 18th Joint Meeting. Vico Equense Italy	15-18 Nov 2015
Solcuc2017 3rd Cucurbitaceae Joint Conference. Valencia Spain	03-06 Sep 2017
Paris Plant Breeding Symposium 2017. AgroParisTech, Paris, France	26 Sep 2017
ICBAA2017 - International conference on big data application in agriculture. University Putra Malaysia, Malaysia	05-06 Dec 2017
Sol2018 The 15th Solanaceae Conference. Chiang Mai, Thailand	30 Sep - 04 Oct 2018
EAPR Eucarpia - 19th Joint Meeting. Rostock-Warnemünde, Germany	03-06 Dec 2018
Synchytrium endobioticum workshop. Wageningen	26-28 Jun 2019
► Presentations	
Poster: An integrated genomics and effectormics approach for potato wart disease resistance breeding - EAPR Eucarpia	15-18 Nov 2015
Poster: Pitfalls and solutions to develop haplotype specific markers for Wart Disease resistance in potato - Solcuc2017 3rd Cucurbitaceae Joint Conference	03-06 Sep 2017
Poster: Bulk Segregant K-mer Analysis (BSKA): a novel and versatile tool for genetic research - Sol2018	30 Sep - 4 Oct 2018
Talk: Wart disease resistance project, Results and Future Work in WP2 - 1st Consortium meeting - EAPR Eucarpia	18 Nov 2015
Talk: Mapping of a potato wart disease R gene using WGS data - ICBAA2017	05-06 Dec 2017
Talk: Mapping of a potato wart disease R gene using WGS data - EAPR Eucarpia 19th Joint Meeting	03-06 Dec 2018
Talk: Stacking of resistance loci to give full resistance to higher pathotypes of <i>S. endobioticum</i> - Synchytrium endobioticum workshop	27 Jun 2019
Talk: Wart disease resistance project - Work Package 2: Work progress - Last Consortium meeting - TKI project	28 Nov 2019
► Excursions	
Solynta demonstration day	23 Aug 2017
EPS company visit Keygene	12 Oct 2017

Subtotal Scientific Exposure

21.2

3) In-Depth Studies	<u>date</u>
► Advanced scientific courses & workshops	
Big Data in R. Formation given for Plant Sciences employees	09-10 May 2016
MGC Next Generation Sequencing data analysis	29-31 Aug 2016

Polyploid software workshop	12-13 Dec 2016
The Power of RNA-seq	11-13 Jun 2018
► Journal club	
Quantitative genetics journal club with other PhD students of Plant Breeding group	12 Nov 2015 - Jan 2019
► Individual research training	
Roche formation for the use of bait libraries	03, 04, 07 Dec 2015

Subtotal In-Depth Studies

7.3

4) Personal Development	<u>date</u>
► General skill training courses	
Competence Assessment	02 Feb 2016
Brain training	23 Mar 2016
WGS PhD Workshop Carousel 2016	08 Apr 2016
Efficient Writing Strategies	04 Apr - 24 May 2016
EPS introduction course	29 Sep 2016
Data Management Planning	06 Feb 2017
WGS PhD Workshop Carousel 2017	07 Apr 2017
Teaching and Supervising Thesis Students	15-16 May 2017
eLab Journal Workshop	19 Sep 2017
Scientific writing	13 Mar - 30 Apr 2018
WGS PhD Workshop Carousel 2018	25 May 2018
Nederlandse taal: van A1 naar B1	Jan 2016 - Dec 2017

Subtotal Personal Development

8.1

TOTAL NUMBER OF CREDIT POINTS*

39.1

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

This work was performed in a project called “An integrated genomics and effectoromics impulse for potato wart resistance management and breeding” which was financially supported by the Dutch Topsector Horticulture & Starting Materials, Averis seeds BV, Böhm-Nordkartoffel Agrarproduktion GmbH & Co. OHG, Danespo, HLB BV, HZPC Holland BV, C Meijer BV, SaKa Pflanzenzucht GmbH & Co. KG. Within the Topsector, private industry, knowledge institutes and the government are working together on innovations for sustainable production of safe and healthy food and the development of a healthy green environment.

Charlotte Prodhomme was supported by a Walsh Scholarship provided by Teagasc.

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

Cover painting: Sébastien Hardouin

Cover and layout design: Dennis Hendriks, ProefschriftMaken

Printing: ProefschriftMaken