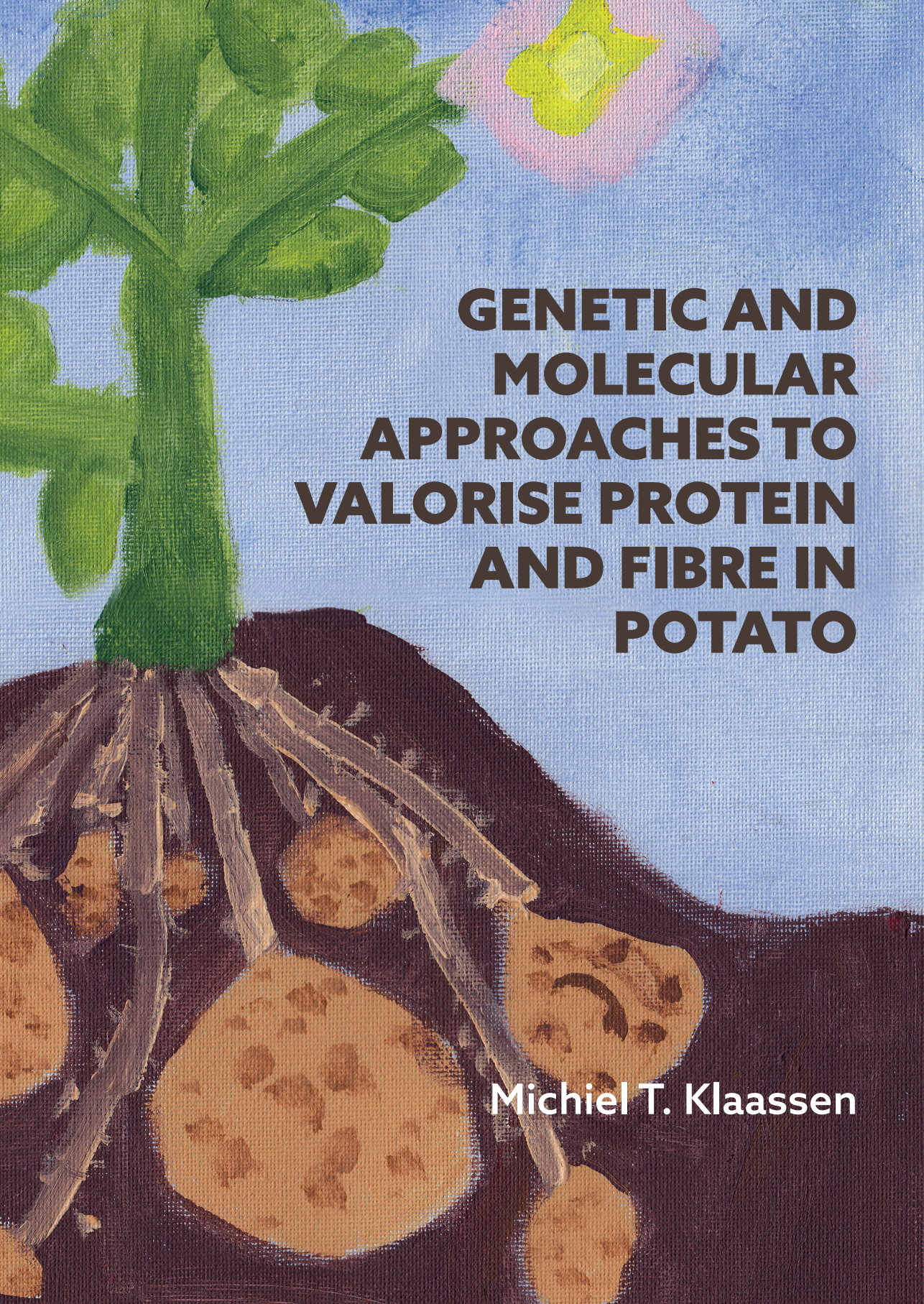

GENETIC AND MOLECULAR APPROACHES TO VALORISE PROTEIN AND FIBRE IN POTATO

Michiel T. Klaassen

Genetic and molecular approaches to valorise protein and fibre in potato

Michiel T. Klaassen

Thesis committee

Promotors
Prof. Dr L.M. Trindade
Personal chair, Bioresources Breeding and Genetics
Wageningen University & Research

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

Co-promotor
Dr C.A. Maliepaard
Associate professor, Plant Breeding
Wageningen University & Research

Other members
Prof. Dr J.-P. Vincken, Wageningen University & Research
Prof. Dr B.J. Zwaan, Wageningen University & Research
Em. Prof. Dr J.P.M. Sanders, Wageningen University & Research
Dr D.V. Bustos-Korts, Wageningen University & Research

This research was conducted under the auspices of the Graduate School of Experimental Plant Sciences (EPS)

Genetic and molecular approaches to valorise protein and fibre in potato

Michiel T. Klaassen

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 31 January 2020
at 4:00 p.m. in the Aula.

Michiel T. Klaassen

Genetic and molecular approaches to valorise protein and fibre in potato

166 pages

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

With references, with summary in English

ISBN: 978-94-6395-233-0

DOI: 10.18174/508133

Table of contents

Chapter 1

General introduction

Research goal

The goal of the research described in this dissertation was to gain insight into the genetic and molecular architectures of tuber protein content and functionalities of cell wall polysaccharides in potato.

History and use of the potato crop

Potato (*Solanum tuberosum* L.) is of ancient South American descent, where the botanical *Andigenum* group was domesticated and grown in the Andean highlands in Southern Peru and Northern Bolivia as early as 5000 B.C. (Ugent 1970). Although potato was brought to Europe in the 16th century, its agronomic advantages, that include its high-yielding ability and relatively low water needs (Hoekstra and Chapagain 2011) were only recognized by 1800 A.D. (Nunn and Qian 2011). Later it developed into a major European food crop (Sabine 1823) and was distributed from Europe to most other parts of the world. In terms of production, potato ranks as the fourth most important food crop worldwide (FAO 2018). For potato, three general market segments can be distinguished: i) fresh consumption, ii) processing and iii) starch. These market segments demand varieties with a specific properties (*i.e.* traits). Therefore, potato varieties may vary in terms of size, shape, flesh colour, flavour and starch content. For example, the potato starch industry typically makes use of bulky varieties with a high starch content, whereas the market segment for fresh consumption generally requires types that are appealing in terms of shape, size and taste.

Valorisation of by-products in the potato starch industry

At the end of the 19th century, the potato starch industry in The Netherlands consisted of many small-scale processing factories (Grommers and Van der Krogt 2009). These factories extracted starch as the main resource from potatoes. Later, these small-scale factories were consolidated to form more efficient large-scale processing factories. As a result, these larger factories processed more potatoes, thus producing more starch and releasing large volumes of by-products containing protein and cell walls. At that time, the value of protein and cell walls (*i.e.* fibre) from potatoes was limited. Therefore, by-products from the potato starch industry were released into the

environment that often resulted in pollution. In many countries, legislation was installed to prohibit the potato starch industry from discarding its by-products into the environment. These developments drove the potato starch industry to redefine their policies on how to deal with their by-products. Nowadays, innovative firms in the potato starch industry are eager to reduce their environmental impact and to increase their competitive-edge by exploiting protein and fibre in high-value applications that include healthcare, food and feed products. The Dutch potato starch industry has demonstrated that valorisation of (functional) potato protein renders ample added-value (*e.g.* Solanic® potato protein, www.avebe.com/producten). In the near future, the global market size for potato protein is expected to grow (Mordor Intelligence 2018). Hence, valorisation of by-products as high-value resources provides the industry with new opportunities for innovation, especially in the light of circular agriculture in the Netherlands (Min. LNV 2018). In circular agriculture, by-products and other resources are used effectively and efficiently to create value and reduce environmental impact.

Potato protein

Potato is a well-known starch crop but also serves as an important source of protein and fibre. As opposed to the large amount of research that has been devoted to study tuber starch content in potatoes, few studies have focused on tuber protein content from a genetic point of view. In potato tubers, soluble protein content is known to vary between 0.3 to 1.5% (w/w, fresh weight basis) (Ortiz-Medina 2006; Werij 2011). Potato protein can be categorized into three distinct groups in terms of their molecular weight: i) patatins (~43 kDa), ii) protease inhibitors (5-25 kDa) and iii) other high molecular weight proteins (Pots et al. 1999). Patatin is a major protein group in potato tubers, where it accounts for 40% of the total soluble protein content (Paiva et al. 1983; Park et al. 1983; Racusen and Foote 1980). In potato, 10 to 18 patatin genes have been reported to belong to two classes (Pikkard et al. 1987; Twell and Ooms 1988). Patatin class I genes are specifically expressed in tubers, whilst class II patatins are also expressed in roots (Prat et al. 1990). Patatin genes are located on different potato chromosomes, but an apparent cluster of seven patatin genes are positioned at the start of chromosome *8* (PGSC 2011). Protease inhibitor proteins can be categorized into three categories in terms of

their molecular weight and two main gene clusters are found at the start of chromosomes *3* and *6* in potato (PGSC 2011). Numerous studies have been carried out in an attempt to describe the biological functions of patatin and protease inhibitor proteins. In literature, these proteins are generally considered to function as storage proteins that provide a source of amino acids for tuber sprouting and juvenile plant growth (Bauw et al. 2006). Moreover, it has been suggested that these proteins may be involved in the defence against insects and pathogens (Nissen et al. 2009; Ryan 1990; Strickland et al. 1995). Besides these potential biological functions, the physico-chemical properties of functional (native) potato protein make them particularly suitable for food and pharma applications. These properties include their foaming, gelling, emulsifying, moisturizing, satiety-inducing, food preserving and skin care attributes (Bártová and Bárta 2009; Creusot et al. 2011; Giuseppin et al. 2009; Kudo et al. 2009). To further exploit the potential of protein in potato, the potato starch industry is keen to use varieties that are protein-rich. Therefore, protein content has emerged as a new breeding goal for starch potato breeders. Breeding for this trait is however not straightforward due its complex (*i.e.* polygenic) nature that is poorly understood. One putative regulator of patatin gene expression (*i.e.* DNA binding protein "Storekeeper") has been reported in literature (Zourelidou et al. 2002). It is unclear if other economically relevant potato traits are complementary or non-complementary to tuber protein content. Studies in diploid potato research populations have shown that tuber protein content is moderately heritable (Lu et al. 2012; Werij 2011). Hence, one may hypothesize that breeding for tuber protein content is attainable although this has not been studied in cultivated potato.

Potato fibre

Potato tubers consist of approximately 1% fibre (w/w, fresh weight) which mainly consists of primary cell walls from parenchyma cells (Grommers and Van der Krogt 2009; Lisinska and Leszczynski 1989; McDougall et al. 1996). Potato cell walls (PCW) are composed of pectin (56%), cellulose (30%) and hemicellulose (11% xyloglucan and 3% mannan) (Vincken et al. 2000). PCW are highly hydrophilic as they hold up to 90% (w/w) water (Mayer and Hillebrandt 1997). This property hampers the use of this colloid-like by-product in high-value applications. Therefore, PCW is frequently used in low-

value applications, *e.g.* animal feed and soil amendments. Improving the quality of PCW, by reducing its hydrophilic property is therefore relevant. To modify this property, it is necessary to understand how properties of the cell wall affect this trait (*e.g.* composition and architecture). By identifying critical factors and characterizing their effects, new strategies for improvement may be formulated. In literature it has been suggested that pectic side-chains may play a role in hydration of plant cell walls (Belton 1997; Funami et al. 2011; Larsen et al. 2011; Ramasamy et al. 2015; Ramaswamy et al. 2013). Hence, these pectic structures may be targeted to evaluate their effects on cell wall hydration properties and to assess their potential pleiotropic effects on plant growth performance.

Genetics of the potato crop

Potato displays diversity at both the species and the genome level. Contemporary taxonomic conspectus has shown that potato germplasm includes cultivated landraces and wild species (Ovchinnikova et al. 2011; Spooner et al. 2014; Spooner et al. 2007). In the germplasm, the ploidy level ranges from diploid ($2n = 2x = 24$) with two sets of homologous chromosomes to hexaploid ($2n = 6x = 72$) with six sets. In the 21st century, geneticists realized that cultivated potato varieties in Europe were tetraploids ($2n = 4x = 48$) (Cadman 1942; Lunden 1937) and that these belonged to the species *Solanum tuberosum* L. group *tuberosum*. These varieties were later characterized as autotetraploids that exhibit polysomic inheritance, *i.e.* random pairing of the homologs during meiosis with both bivalent and multivalent formation as evidenced from the occurrence of double reduction. Polyploidy, *i.e.* having more than two sets of (homologous) chromosomes, is well tolerated in potato and may have contributed to the crop's adaptability by benefiting genetic and physiological buffering capacities (Crow and Wagner 2005). Self-incompatibility and inbreeding depression are factors that maintain a high-level of heterozygosity in cultivated potato (Krantz 1924; Krantz and Hutchins 1929). Ample genetic resources from both wild potato and landraces are available to improve potato traits (Ross 1986). For instance, wild species have been utilized as a source to incorporate resistances in varieties against diseases that include late blight, viruses and cyst nematodes such as *Globodera pallida* (Bradshaw 2009; Bradshaw and Ramsay 2005).

Potato breeding

In current potato breeding, new varieties are almost exclusively selected from tetraploid F_1 progenies that are derived from single bi-parental crosses (a single meiosis event per parental gamete). The goal of this selection process is to identify individuals that complement or outperform their parents for multiple traits. To do so, very large progeny sets and ample time (10-12 years) are needed to develop elite varieties (Bradshaw 2009; Hamilton et al. 2011; Lindhout et al. 2011). This is caused by several factors. Firstly, the genetic bases underlying many agronomic potato traits are still poorly understood (Hamilton et al. 2011). Therefore, it difficult to fixate favourable alleles (in a homozygous state) and to remove deleterious ones to improve polygenic traits that are regulated by multiple genes. As suggested by Vos et al. (2015), deleterious variants are abundant in the tetraploid potato genepool and induce negative effects on traits. This especially holds true for polygenic traits, as the relatively high occurrence of deleterious variants may be compensated by their counterparts (favourable alleles). As favourable alleles are not (easily) fixed, the probability to select superior individuals is very low. Secondly, because the selection for specific traits (*e.g.* pathogen resistances) is eminent for breeders, it is challenging to simultaneously improve other complex traits, especially without a basic understanding of the underlying mechanisms (*e.g.* sink strength or uptake of nutrients) and their genetics (*e.g.* alleles in coupling or repulsion phase).

Modern tools for the dissection and improvement of potato traits

Identifying (causal) relationships between genetic variants and phenotypic variation is of fundamental importance for biologists and breeders. In the 19th century, the Augustinian monk *Gregor Mendel* studied the heredity of pea traits (Mendel 1866). This ground-breaking work pointed to the existence of biological elements called genes. Mendel's work laid the foundation for the principles of genetics and analytical procedures to study the inheritance of traits in both model species and crops. From the year 1950 to present, spectacular progress has been made in the field of molecular biology (Friedberg 2007), *e.g.* the discovery of the double-helix structure of DNA (deoxyribonucleic acid) and CRISPR/Cas-9-mediated gene editing (Doudna

and Charpentier 2014). In this so-called 'golden age' of molecular biology, tremendous progress has been made in the understanding of basic biological structures and physiological processes that influence traits. This progress has allowed the development of new tools to dissect traits and to enhance breeding. Over the years, quantitative trait locus (QTL) mapping in populations (Brigneti et al. 1997; Bryan et al. 2002) and genome-wide association studies (GWAS) in variety panels (Bradshaw et al. 2008; Kloosterman et al. 2013; Vos 2016) have been shown to function as suitable tools to perform genetic analysis in potato. Since the year 1980, genetic markers were developed and applied to construct chromosomal linkage maps in diploid potato (Bonierbale et al. 1988; Van Os et al. 2006). Later, the availability of the first potato reference genome of *Solanum tuberosum* L. group *phureja* DM1-3 (PGSC 2011), single-nucleotide polymorphism (SNP) marker arrays (Felcher et al. 2012; Hamilton et al. 2011; Uitdewilligen et al. 2013) and statistical methods (Bourke et al. 2018; Preedy and Hackett 2016), have enabled the construction of high-density SNP-based chromosome linkage maps in tetraploid potato (*e.g.* Bourke et al. 2016; Hackett et al. 2014; Hackett et al. 2013) and the subsequent reconstruction of multi-locus haplotypes (Zheng et al. 2016). Publicly available software packages are currently used for routine genetic analyses in tetraploid potato (Bourke et al. 2018; Rosyara et al. 2016). The availability of new tools, that include an optimized reference genome of the self-compatible diploid potato line *Solyntus*™ (SGSC 2019), are expected to further enhance the genetic analysis of traits in potato.

In general, GWAS and QTL mapping techniques are effective for dissecting qualitative traits (*i.e.* simple inheritance) where individual genetic variants (alleles) provide strong and detectable effects. In potato, genetic studies have also been devoted to dissect the inheritance of complex (*i.e.* polygenic and moderately heritable) traits that are influenced by multiple genes and environmental conditions. In potato, complex traits include yield, starch content and glycoalkaloid content (Massa et al. 2015; Schäfer-Pregl et al. 1998; Vos 2016). Forward genetic approaches facilitate the characterization of the genetic architecture of a trait (*i.e.* the number of loci, their effects and origin) and subsequent identification of candidate genes. To characterize or

validate the molecular functions of candidate genes, functional molecular studies (*e.g.* gene overexpression) may be carried out as a reverse genetic approach.

Objectives and outline of this dissertation

A better understanding of the genetics and biological processes or structures that underlie protein content and fibre quality, will enable the improvement and valorisation of these resources in potato. Therefore, the objectives of this study are:

1. To dissect the genetic architecture of tuber protein content;
2. To identify candidate genes and putative physiological processes that affect tuber protein content;
3. To uncover structure-function relationships of pectin and hydration of potato cell walls (fibre).

Genetic analyses were carried out with regard to tuber protein content, whereas molecular approaches were used to study the effects of genes affecting tuber protein content and fibre quality. To this end, this dissertation contains the following chapters.

In **Chapter *2*** we studied the genetics of tuber protein content in a panel of tetraploid potato varieties. We estimated the trait heritability, identified QTLs, haplotypes (alleles) underlying QTLs and putative candidate genes. Haplotype inference showed that alleles of *StCDF1* were associated with a QTL for tuber protein content. The results acquired from Chapter *2* are relevant for Chapter *3* and Chapter *4*.

Chapter *3* describes the genetic analysis of tuber protein content in a bi-parental population of cultivated potato. A dense integrated chromosomal genetic map was constructed to identify naive QTLs for tuber protein content. Cofactor QTL analysis uncovered additional loci that were not identified in the naive analysis. We also report on trait heritability. The results acquired from Chapter *3* are relevant for Chapter *2* and Chapter *4*.

In **Chapter *4*** we overexpressed a putative nitrate transporter gene (*StNPF1.11*) to study its effect on tuber protein content in potato. *StNPF1.11* was selected as a candidate gene underlying QTLs for tuber protein content (Chapter *2* and Chapter *3*). Overexpression of *StNPF1.11* induced effects on tuber protein content, leaf chlorophyll content and plant height. Correlations between *StNPF.11* expression and protein content, as well as pleiotropic effects are presented.

In **Chapter *5*** we studied the role of pectic rhamnogalacturonan (RG-I) galactan side-chains on the water-binding capacity (WBC) of potato cell walls (PCW). This was done to shed light on this structure-function relationship. Both *in-vivo* and *in-vitro* truncation of RG-I β-(1→4)-D-galactan side-chains altered the WBC of PCW. Our results reinforce the view that RG-I galactan side-chains play a role in modulating the WBC of PCW.

In **Chapter *6*** the results from Chapters *2*, *3*, *4* and *5* are evaluated and discussed in a broader context, together with implications of the results and prospects for future research.

References

Bártová V, Bárta J (2009) Chemical composition and nutritional value of protein concentrates isolated from potato (Solanum tuberosum L.) fruit juice by precipitation with ethanol or ferric chloride. Journal of Agricultural and Food Chemistry 57 (19):9028-9034

Bauw G, Nielsen HV, Emmersen J, Nielsen KL, Jørgensen M, Welinder KG (2006) Patatins, Kunitz protease inhibitors and other major proteins in tuber of potato cv. Kuras. FEBS Journal 273 (15):3569-3584

Belton P (1997) NMR and the mobility of water in polysaccharide gels. International Journal of Biological Macromolecules 21 (1):81-88

Bonierbale MW, Plaisted RL, Tanksley SD (1988) RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. Genetics 120 (4):1095-1103

Bourke PM, van Geest G, Voorrips RE, Jansen J, Kranenburg T, Shahin A, Visser RGF, Arens P, Smulders MJM, Maliepaard C (2018) polymapR—linkage analysis and genetic map construction from F1 populations of outcrossing polyploids. Bioinformatics 34 (20):3496-3502

Bourke PM, Voorrips RE, Kranenburg T, Jansen J, Visser RG, Maliepaard C (2016) Integrating haplotype-specific linkage maps in tetraploid species using SNP markers. Theoretical and Applied Genetics 129 (11):2211-2226

Bradshaw JE (2009) Potato breeding at the Scottish plant breeding station and the Scottish Crop Research Institute: 1920–2008. Potato Research 52 (2):141-172

Bradshaw JE, Hackett CA, Pande B, Waugh R, Bryan GJ (2008) QTL mapping of yield, agronomic and quality traits in tetraploid potato (Solanum tuberosum subsp. tuberosum). Theoretical and Applied Genetics 116 (2):193-211

Bradshaw JE, Ramsay G (2005) Utilisation of the Commonwealth Potato Collection in potato breeding. Euphytica 146 (1-2):9-19

Brigneti G, Garcia-Mas J, Baulcombe D (1997) Molecular mapping of the potato virus Y resistance gene Rysto in potato. Theoretical and Applied Genetics 94 (2):198-203

Bryan G, McLean K, Bradshaw J, De Jong W, Phillips M, Castelli L, Waugh R (2002) Mapping QTLs for resistance to the cyst nematode Globodera pallida derived from the wild potato species Solanum vernei. Theoretical and Applied Genetics 105 (1):68-77

Cadman C (1942) Autotetraploid-inheritance in the potato: some new evidence. Journal of Genetics 44 (1):33

Creusot N, Wierenga PA, Laus MC, Giuseppin ML, Gruppen H (2011) Rheological properties of patatin gels compared with β-lactoglobulin, ovalbumin, and glycinin. Journal of the Science of Food and Agriculture 91 (2):253-261

Crow KD, Wagner GP (2005) What Is the Role of Genome Duplication in the Evolution of Complexity and Diversity? Molecular Biology and Evolution 23 (5):887-892

Doudna JA, Charpentier E (2014) The new frontier of genome engineering with CRISPR-Cas9. Science 346 (6213):1258096

FAO (2018) Aavailable at http://faostat.fao.org (Accessed 3 December 2018).

Felcher KJ, Coombs JJ, Massa AN, Hansey CN, Hamilton JP, Veilleux RE, Buell CR, Douches DS (2012) Integration of two diploid potato linkage maps with the potato genome sequence. PloS One 7 (4):e36347

Friedberg EC (2007) Adventures in the Golden Age of Molecular Biology. Adler Museum of Medicine:33

Funami T, Nakauma M, Ishihara S, Tanaka R, Inoue T, Phillips GO (2011) Structural modifications of sugar beet pectin and the relationship of structure to functionality. Food Hydrocolloids 25 (2):221-229

Giuseppin MLF, Van der Sluis C, Laus MC (2009) Native potato protein isolates. Patent WO 2008/069650

Grommers HE, Van der Krogt DA (2009) Chapter 11 - Potato Starch: Production, Modifications and Uses, in" Starch: Chemistry and Technology", J. BeMiller and R. Whistler, Editors. Academic Press, San Diego, CA, USA

Hackett CA, Bradshaw JE, Bryan GJ (2014) QTL mapping in autotetraploids using SNP dosage information. Theoretical and Applied Genetics 127 (9):1885-1904

Hackett CA, McLean K, Bryan GJ (2013) Linkage analysis and QTL mapping using SNP dosage data in a tetraploid potato mapping population. PloS One 8 (5):e63939

Hamilton JP, Hansey CN, Whitty BR, Stoffel K, Massa AN, Van Deynze A, De Jong WS, Douches DS, Buell CR (2011) Single nucleotide polymorphism discovery in elite North American potato germplasm. BMC Genomics 12 (1):302

Hoekstra AY, Chapagain AK (2011) Globalization of water: Sharing the planet's freshwater resources. John Wiley & Sons, Hoboken, NJ, USA

Kloosterman B, Abelenda JA, Gomez MdMC, Oortwijn M, de Boer JM, Kowitwanich K, Horvath BM, van Eck HJ, Smaczniak C, Prat S (2013) Naturally occurring allele diversity allows potato cultivation in northern latitudes. Nature 495 (7440):246

Krantz FA (1924) Potato breeding methods. Minnesota Agric. Exp. Stn. Tech. Bull. 25.32

Krantz FA, Hutchins AE (1929) Potato breeding methods II: selection in inbred lines. Minnesota Agric. Exp. Stn. Tech. Bull. 58.23

Kudo K, Onodera S, Takeda Y, Benkeblia N, Shiomi N (2009) Antioxidative activities of some peptides isolated from hydrolyzed potato protein extract. Journal of Functional Foods 1 (2):170-176

Larsen FH, Byg I, Damager I, Diaz J, Engelsen SB, Ulvskov P (2011) Residue specific hydration of primary cell wall potato pectin identified by solid-state 13C single-pulse MAS and CP/MAS NMR spectroscopy. Biomacromolecules 12 (5):1844-1850

Lindhout P, Meijer D, Schotte T, Hutten RCB, Visser RGF, van Eck HJ (2011) Towards F1 Hybrid Seed Potato Breeding. Potato Research 54 (4):301-312

Lisinska G, Leszczynski W (1989) Potato starch processing. In: Lisinska, G., Leszczynski, W. (Eds.). Potato science and technology (pp. 281-346). Essex, UK: Elsevier

Lu W, Yu M, Bai Y, Li W, Xu X (2012) Crude Protein Content in Diploid Hybrid Potato Clones of Solanum phureja–S. stenotomum. Potato Research 55 (3):315-322

Lunden AP (1937) Arvelighetsundersokelser i potet. Inheritance studies in the potato) Meld Norg Landbr 17:1-156

Massa AN, Manrique-Carpintero NC, Coombs JJ, Zarka DG, Boone AE, Kirk WW, Hackett CA, Bryan GJ, Douches DS (2015) Genetic linkage mapping of economically important traits in cultivated tetraploid potato (Solanum tuberosum L.). G3: Genes | Genomes | Genetics 5 (11):2357-2364

Mayer F, Hillebrandt J-O (1997) Potato pulp: microbiological characterization, physical modification, and application of this agricultural waste product. Applied Microbiology and Biotechnology 48 (4):435-440

McDougall GJ, Morrison IM, Stewart D, Hillman JR (1996) Plant cell walls as dietary fibre: range, structure, processing and function. Journal of the Science of Food and Agriculture 70 (2):133-150

Mendel G (1866) Versuche über Pflanzenhybriden Verhandlungen des naturforschenden Vereines in Brünn, Bd. IV für das Jahr, 1865. Abhandlungen: 3-47

Min. LNV (2018) Agriculture, nature and food: valuable and connected. The Netherlands as a leader in circular economy. Ministry of Agriculture, Nature and Food Quality (Min. LNV). The Hague, The Netherlands

Mordor Intelligence (2018) Potato Protein Market - Growth, Trends And Forecasts (2019 - 2024). Report ID: 4771776. Hyderabad, India

Nissen MS, Kumar GM, Youn B, Knowles DB, Lam KS, Ballinger WJ, Knowles NR, Kang C (2009) Characterization of Solanum tuberosum multicystatin and its structural comparison with other cystatins. The Plant Cell Online 21 (3):861-875

Nunn N, Qian N (2011) The potato's contribution to population and urbanization: evidence from a historical experiment. The Quarterly Journal of Economics 126 (2):593-650

Ortiz-Medina E (2006) Potato tuber protein and its manipulation by chimeral disassembly using specific tissue explantation for somatic embryogenesis. PhD dissertation. McGill University. Department of Plant Science. Montreal, Quebec, Canada.

Ovchinnikova A, Krylova E, Gavrilenko T, Smekalova T, Zhuk M, Knapp S, Spooner DM (2011) Taxonomy of cultivated potatoes (Solanum section Petota: Solanaceae). Botanical Journal of the Linnean Society 165 (2):107-155

Paiva E, Lister RM, Park WD (1983) Induction and accumulation of major tuber proteins of potato in stems and petioles. Plant Physiology 71 (1):161-168

Park WD, Blackwood C, Mignery GA, Hermodson MA, Lister RM (1983) Analysis of the heterogeneity of the 40,000 molecular weight tuber glycoprotein of potatoes by immunological methods and by NH2-terminal sequence analysis. Plant Physiology 71 (1):156-160

PGSC (2011) Genome sequence and analysis of the tuber crop potato. Nature 475 (7355):189-195

Pikkard CS, Brusca JS, Hannapel DJ, Park WD (1987) The tow classes of genes for the major potato tuber protein, patatin, are differentially expressed in tubers and roots. Nucleic Acids Research 15 (5):1979-1994

Pots AM, Gruppen H, van Diepenbeek R, van der Lee JJ, van Boekel MAJS, Wijngaards G, Voragen AGJ (1999) The effect of storage of whole potatoes of three cultivars on the patatin and protease inhibitor content; a study using capillary electrophoresis and MALDI-TOF mass spectrometry. Journal of the Science of Food and Agriculture 79 (12):1557-1564

Prat S, Frommer W, Höfgen R, Keil M, Kossmann J, Köster-Töpfer M, Liu X-J, Müller B, Pẽna-Cortés H, Rocha-Sosa M (1990) Gene expression during tuber development in potato plants. FEBS Letters 268 (2):334-338

Preedy K, Hackett C (2016) A rapid marker ordering approach for high-density genetic linkage maps in experimental autotetraploid populations using multidimensional scaling. Theoretical and Applied Genetics 129 (11):2117-2132

Racusen D, Foote M (1980) A major soluble glycoprotein of potato tubers. Journal of Food Biochemistry 4 (1):43-52

Ramasamy UR, Gruppen H, Kabel MA (2015) Water-holding capacity of soluble and insoluble polysaccharides in pressed potato fibre. Industrial Crops and Products 64:242-250

Ramaswamy UR, Kabel MA, Schols HA, Gruppen H (2013) Structural features and water holding capacities of pressed potato fibre polysaccharides. Carbohydrate Polymers 93 (2):589-596

Ross H (1986) Potato breeding-problems and perspectives. Fortschritte der Pflanzenzuechtung (Germany)

Rosyara UR, De Jong WS, Douches DS, Endelman JB (2016) Software for genome-wide association studies in autopolyploids and its application to potato. The Plant Genome 9 (2)

Ryan CA (1990) Protease inhibitors in plants: genes for improving defenses against insects and pathogens. Annual Review of Phytopathology 28 (1):425-449

Sabine J (1823) On the Native Country of the Wild Potatoe, with an Account of Its Culture in the Garden of the Horticultural Society: And Observations on the Importance of Obtaining Improved Varieties of the Cultivated Plants. The Journal of Science and Arts (15):259-266

Schäfer-Pregl R, Ritter E, Hesselbach J, Lovatti L, Walkemeier B, Thelen H, Salamini F, Gebhardt C (1998) Analysis of quantitative trait loci (QTLs) and quantitative trait alleles (QTAs) for potato tuber yield and starch content. Theoretical and Applied Genetics 97 (5-6):834-846

SGSC (2019) Solyntus Genome Sequencing Consortium. Retrieved from www.plantbreeding.wur.nl/Solyntus/index.html. Accessed 14 Oct 2019

Spooner DM, Ghislain M, Simon R, Jansky SH, Gavrilenko T (2014) Systematics, diversity, genetics, and evolution of wild and cultivated potatoes. The Botanical Review 80 (4):283-383

Spooner DM, Núñez J, Trujillo G, del Rosario Herrera M, Guzmán F, Ghislain M (2007) Extensive simple sequence repeat genotyping of potato landraces supports a major reevaluation of their gene pool structure and classification. Proceedings of the National Academy of Sciences 104 (49):19398-19403

Strickland JA, Orr GL, Walsh TA (1995) Inhibition of Diabrotica larval growth by patatin, the lipid acyl hydrolase from potato tubers. Plant Physiology 109 (2):667-674

Twell D, Ooms G (1988) Structural diversity of the patatin gene family in potato cv. Desiree. Molecular and General Genetics MGG 212 (2):325-336

Ugent D (1970) The potato. Science 170 (3963):1161-1166

Uitdewilligen JG, Wolters A-MA, Bjorn B, Borm TJ, Visser RG, van Eck HJ (2013) A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. PloS One 8 (5):e62355

Van Os H, Andrzejewski S, Bakker E, Barrena I, Bryan GJ, Caromel B, Ghareeb B, Isidore E, de Jong W, Van Koert P (2006) Construction of a 10,000-marker ultradense genetic recombination map of potato: providing a framework for accelerated gene isolation and a genomewide physical map. Genetics 173 (2):1075-1087

Vincken JP, Borkhardt B, Bush M, Doesdijk-Voragen C, Dopico B, Labrador E, Visser R (2000) Remodelling pectin structure in potato. Conference Proceedings of Phytosfere '99 European Plant Biotechnology Network, 23-34

Vos P (2016) Development and application of a 20K SNP array in potato. PhD dissertation. Chair group: Plant Breeding. Wageningen University, Wageningen, The Netherlands. Retrieved from http://edepot.wur.nl/392278

Werij JS (2011) Genetic analysis of potato tuber quality traits. PhD dissertation. Chair group: Plant Breeding. Wageningen University, Wageningen, The Netherlands. Retrieved from http://edepot.wur.nl/183746

Zheng C, Voorrips RE, Jansen J, Hackett CA, Ho J, Bink MC (2016) Probabilistic Multilocus Haplotype Reconstruction in Outcrossing Tetraploids. Genetics 203 (1):119-131

Zourelidou M, De Torres-Zabala M, Smith C, Bevan MW (2002) Storekeeper defines a new class of plant-specific DNA-binding proteins and is a putative regulator of patatin expression. The Plant Journal 30 (4):489-497

Chapter 2

Genome-wide association analysis in tetraploid potato reveals four QTLs for protein content

Michiel T. Klaassen[1,2], Johan H. Willemsen[1], Peter G. Vos[1], Richard G.F. Visser[1], Herman J. van Eck[1], Chris Maliepaard[1], Luisa M. Trindade[1]

[1] Wageningen University & Research, Plant Breeding, Wageningen, The Netherlands

[2] Aeres University of Applied Sciences, Department of Applied Research, Dronten, The Netherlands

With minor modifications this chapter has been published in:

Molecular Breeding **(2019) 39 (11):151**

DOI: 10.1007/s11032-019-1070-8

Abstract

Valorisation of tuber protein is relevant for the potato starch industry to create added-value and reduce impact on the environment. Hence, protein content has emerged as a key quality trait for innovative potato breeders. In this study, we estimated trait heritability, explored the relationship between protein content and tuber under-water weight (UWW), inferred haplotypes underlying quantitative trait loci (QTLs) and pinpointed candidate genes. We used a panel of varieties (N = 277), that was genotyped using the SolSTW 20K Infinium single-nucleotide polymorphism (SNP) marker array. Protein content data were collected from multiple environments and years. Our genome-wide association study (GWAS) identified QTLs on chromosomes *3*, *5, 7* and *12.* Alleles of *StCDF1* (maturity) were associated with QTLs found on chromosome *5.* The QTLs on chromosomes *7* and *12* are presented here for the first time, whereas those on chromosomes *3* and *5* co-localized with loci reported in earlier studies. A positive correlation (r = 0.64) was observed between protein content and UWW. The candidate genes underlying the QTLs proposed here are relevant for functional studies. This study provides resources for genomics-enabled breeding for protein content in potato.

Keywords

Protein content, Potato, Tetraploid, Haplotypes, Genome-wide association analysis (GWAS), Candidate genes

Introduction

Global population growth, accompanied with increased consumer wealth, will change food consumption patterns worldwide (Tilman and Clark 2014). By the year 2050, the projected demand for protein from animal sources is expected to double from 2000 (Alexandratos 1999). This trend raises sustainability and food security concerns, as the intensive production of animal protein adds pressure on the environment – as vast amounts scarce (non-renewable) resources such as land, water and minerals are needed. On the contrary, the production of plant protein is more sustainable for the environment as less resources are needed (Sabaté and Soret 2014).

Potato (*Solanum tuberosum* L.) is a well-known starch crop. However, few realise that the potato crop also serves as an abundant source of plant protein (Jørgensen et al. 2006). Although protein content in potato tubers is relatively low (0.32-1.63%) (Bárta et al. 2012; Klaassen et al. 2019; Ortiz-Medina 2006), protein yield per hectare (ha) is eminent due to the high-yielding ability and high harvest-index of the potato crop that can reach up to 124 $ton^{-1}ha$ (in the Columbia Basin, USA) (Kunkel and Campbell 1987). The potato starch industry processes potatoes to produce starch and by-products. After starch is extracted from tubers, potato fruit juice (PFJ) is released as a major aqueous by-product that contains protein. After proteins are extracted from PFJ, functional (native) potato protein isolates may be utilized in high-end food and pharmaceutical applications that include foaming agents, anti-oxidants, emulsifiers (Creusot et al. 2011; Edens et al. 1999; Kudo et al. 2009), inhibitors of faecal proteolytic compounds that cause dermatitis (Ruseler-van Embden et al. 2004) and satiety agents (Hill et al. 1990). Therefore, valorisation of protein provides opportunities to create added-value for the potato starch industry. Consequently, innovative firms in the industry are keen to use protein-rich potato varieties and therefore high protein content has emerged as a key quality trait for breeders. However, breeding for protein content in potato is challenging due to the complex genetic basis underlying the trait (Klaassen et al. 2019). To facilitate breeding for protein content in potato, improved comprehension of the inheritance, quantitative trait loci (QTLs) and relationships with other agronomical relevant traits are useful.

Knowledge on the inheritance of protein content in potato is limited. To the best of our knowledge, three genetic studies on protein content in bi-parental populations have been published (Acharjee et al. 2018; Klaassen et al. 2019; Werij 2011). These studies estimated moderate levels of trait heritability (40-74%) and identified minor-effect QTLs on chromosomes *1*, *2*, *3*, *5* and *9* in both non-cultivated diploid and cultivated tetraploid potato germplasm. As for other crops that include soybean, maize and wheat (Balyan et al. 2013; Hwang et al. 2014; Karn et al. 2017), protein content has been described as a complex trait that is regulated by a plethora of interactions between genetic and environmental factors. Therefore, QTLs for protein content in heterozygous tetraploid ($2n = 4x = 48$) potato are likely to be affected by both epistasis and environmental factors.

Genome-wide association studies (GWAS), have been used as a method to dissect the genetic architecture of complex traits in multiple species that include potato (Rosyara et al. 2016; Sharma et al. 2018). As opposed to genetic studies performed on bi-parental populations, GWAS offers the advantage to identify QTLs within a panel of diverse individuals, and to potentially gain a high mapping resolution for identifying candidate genes.

In this study, we carried out a GWAS to dissect the genetics of protein content in a panel of tetraploid potato. We report on the relationship between protein content and tuber under-water weight (a proxy for starch content), haplotypes underlying QTLs and putative candidate genes.

Materials and methods

Germplasm collection

The panel ($N = 277$) consisted of tetraploid ($2n = 4x = 48$) individuals. The panel was composed of 189 varieties (D'hoop et al. 2008) and 88 starch potato progenitors that originated from five potato breeding companies (Agrico, Averis Seeds, C. Meijer, HZPC and KWS) (Supplementary Table 1). These included both modern and old individuals from different market segments and geographic origins. Analysis of population structure in the panel displayed three sub-populations, as reported earlier (D'hoop et al. 2008; Vos 2016).

These sub-populations, hereafter referred to as “Processing”, “Other” and “Starch”, were used for analyses.

Field trials

Raw phenotypic data were collected over years and locations (multi-location, multi-year) from unbalanced field trials that were carried out in the Netherlands. These trials were carried out in years 2008-2010 in Bant, Emmeloord, Metslawier, Rilland and Valthermond. The accessions were replicated three times or more, except for nine accessions that were replicated twice or once. A replicate (experimental unit), consisted of a four-plant plot within a row in the field. Raw phenotypic data were used to compute the BLUEs for the accessions. The trials were carried out during the conventional potato growing seasons in the Netherlands as described by D’hoop et al. (2011). Uniform seed tubers were used as planting material and were propagated at a single location one year prior to the trials. The seed potatoes were planted at 75-cm spacing between the rows and 35-cm between the hills. Guard rows were used to separate the plots in the trial. Regular husbandry practices for potato production in the Netherlands were carried out during the field trials. After harvest, the tubers were stored under cool conditions prior to use.

Quantification of phenotypes

Soluble protein content in potato fruit juice (PFJ) was determined by using the bicinchoninic acid (BCA) assay (Smith et al. 1985). Bovine serum albumin (BSA) was used as a standard. Protein content was quantified as described by Klaassen et al. (2019). Tuber under-water weight (UWW), a proxy for starch content, was quantified as described in a previous study (Bradshaw et al. 2008).

Best linear unbiased estimates (BLUEs)

To estimate the best linear unbiased estimates (BLUEs), a mixed model was used. BLUEs were computed using restricted maximum likelihood (REML) (D’hoop et al., 2011) as shown in equation *1*:

$$Response\ (Y) = \mu + Accession + Year + Location + (Accession \times Year) + (Accession \times Location) + (Year \times Location) + (Accession \times Year \times Location) + Error \qquad (1),$$

where the "μ" represented the overall mean response. Broad sense heritability estimates (H^2) were computed from variance components (see also D'hoop et al., 2011) as follows:

$$H^2 = \frac{\sigma^2{}_G}{(\sigma^2{}_G + \frac{\sigma^2{}_{G\times Y}}{y} + \frac{\sigma^2{}_{G\times L}}{l} + \frac{\sigma^2{}_{G\times Y\times L}}{y\times l} + \frac{\sigma^2{}_e}{r\times y\times l})} \qquad (2),$$

where the variance components $\sigma^2{}_G$ ("Accession"), $\sigma^2{}_{G\times Y}$ ("Accession" × "Year"), $\sigma^2{}_{G\times L}$ ("Accession" × "Location"), $\sigma^2{}_{G\times Y\times L}$ ("Accession" × "Year" × "Location") and $\sigma^2{}_e$ (residual "Error") were derived from REML. In equation *2*, the terms *"y", "l"* and *"r"* represented the number of years, number of locations and number of biological replicates respectively.

Genotyping and genotype calling

The panel was genotyped using the SolSTW 20K Infinium SNP marker array (Vos et al. 2015). Genotype calling (assignment of SNP allele dosages) were carried out by using fitTetra (Voorrips et al. 2011) and Illumina GenomeStudio software version 2010.3 (Illumina, San Diego, CA, USA), as described by Vos et al. (2015). The threshold for minor-allele frequency (MAF) was set at 1.5% (equivalent to 6% for tetraploid potato with four sets of homologous chromosomes). After filtering, 14,436 high-quality SNP markers were used for GWAS. The physical coordinates of the SNPs were based on the potato reference genome, *i.e.* pseudomolecules v4.03 (PGSC 2011).

Population structure analysis

The population structure of the panel was analysed by using STRUCTURE software package v2.3.4 (Pritchard et al. 2000). Ten runs were performed to estimate the *K* values using 2,000 randomly selected SNPs. A Markov chain Monte Carlo (MCMC) burn-in period of 10,000 was used and the number of iterations was set at 10,000. The appropriate number of sub-populations were

determined from *delta K* and optimal *K* values (Evanno et al. 2005) based on output data derived from STRUCTURE Harvester (Earl and vonHoldt 2012) (http://taylor0.biology.ucla.edu/structureHarvester). Membership probability estimates from thirty runs were averaged and used to assign each individual to cluster groups (sub-populations). The sub-populations were denoted as "Processing", "Other" and "Starch", based on prior knowledge that these three sub-populations existed in the panel (D'hoop et al. 2008; Vos 2016).

Genome-wide association study (GWAS)

A GWAS was performed using the phenotype (BLUEs) and genotype data. A naive model, a mixed model and a conditional mixed model were used to compute associations. For the naive model, associations between the BLUEs and SNP dosages were analysed. The mixed model was used to perform association analysis whilst correcting for kinship (*K*). As population structure was weak for the panel, as shown by (Vos et al. 2017), we did not correct for sub-populations (*Q*). For the mixed model, we used the same SNPs for GWAS and kinship correction. A sub-set of these SNPs was used for inference of population structure. To dissect the effect of maturity alleles (*StCDF1*) that are physically positioned at the start of chromosome *5*, a conditional mixed model was used. The conditional mixed model, as described in earlier studies (Kang et al. 2010; Segura et al. 2012), included the SNP marker "PotVar0079081" (Chromosome *5*, coordinate 4,489,481 Mbp) as a cofactor that tagged the early maturity allele (*StCDF1.1*) in a haplotype-specific manner (Willemsen 2018).

The following equations were used for the GWAS models:

$$Naive\ model: Y = X\alpha + \varepsilon \qquad (3),$$

$$Mixed\ model\ (kinship\ corrected): Y = X\alpha + K\mu + \varepsilon \qquad (4),$$

$$Conditional\ mixed\ model\ (cofactor + kinship\ corrected):$$

$$Y = A\beta + X\alpha + K\mu + \varepsilon \qquad (5).$$

In the equations 3, 4 and 5, "*Y*" represents the BLUEs, "*X*" represents the SNP markers (fixed effect), "*K*" represents the random kinship (co-ancestry) matrix

and "*A*" represents the SNP marker set as cofactor (fixed). The term "ε" represents the vector of random residual errors. The term "α" represents the estimated SNP effects, "β" represents the estimated effect of the SNP marker set as cofactor and "μ" represents the estimated kinship variance component. Analyses were performed in R software package GWASpoly (Rosyara et al. 2016). Phenotypic variance explained (R^2) by SNPs were calculated from squared correlation coefficients between the BLUEs and SNP dosage scores (allele copy number).

Significance threshold and QTL support interval

Manhattan plots were used to illustrate the genome-wide association scores of SNPs. These scores were computed from *P*-values of the SNPs as follows:

$$\text{Association score} = -\log_{10}(P) \qquad (6).$$

We used several significance thresholds to identify QTLs. To correct for multiple testing we used the 5% Bonferroni threshold ($-\log_{10}(P) = 5.3$). The Bonferroni threshold is known to inflate the probability of Type II errors (false-negative findings) in the presence of high linkage disequilibrium between markers (Gao et al. 2008; Johnson et al. 2010). Therefore, the 5% Li and Ji threshold was also computed by correlated multiple testing (Li and Ji 2005) ($-\log_{10}(P) = 3.9$). Correlated multiple testing was conducted at $\alpha = 0.05$, to adjust for the effective number of independent tests and compensate for Type II errors. For naive analyses, permutation testing was carried out with $N = 1000$ permutations at $\alpha = 0.05$ to define the threshold value ($-\log_{10}(P) = 5.0$) (Churchill and Doerge 1994) ($-\log_{10}(P) = 5.0$). The support intervals of QTLs were set at 1.5 Mbp for non-introgressed regions and 2.5 Mbp for introgressed regions as described by Vos (2016).

Haplotype inference

Determination of haplotypes underlying QTLs was performed using a contemporary haplotype inference method developed in tetraploid potato (Willemsen 2018). This method estimated the linkage phase between pairs of SNP markers, followed by joining of linked SNPs into haplotypes. Only SNPs

exceeding the Li and Ji threshold ($-\log_{10}(P) = 3.9$) were used for haplotype construction and to obtain the dosages of the haplotypes.

SNP allele frequency

The SNP allele frequency (%) in the panel of tetraploid accessions, was computed by using the SNP dosage scores and number of accessions as follows:

$$SNP\ allele\ frequency\ (\%) = \frac{Sum\ of\ SNP\ dosage\ scores}{(Number\ of\ accessions \times 4)} \times 100 \qquad (6).$$

Results

Phenotypic variation of the traits (BLUEs)

The panel displayed variation for the BLUEs of protein content, tuber under-water weight (UWW) and protein content in potato fruit juice (PFJ) (**Fig. 1**). Protein content ranged between 0.73-1.72% (w/w).

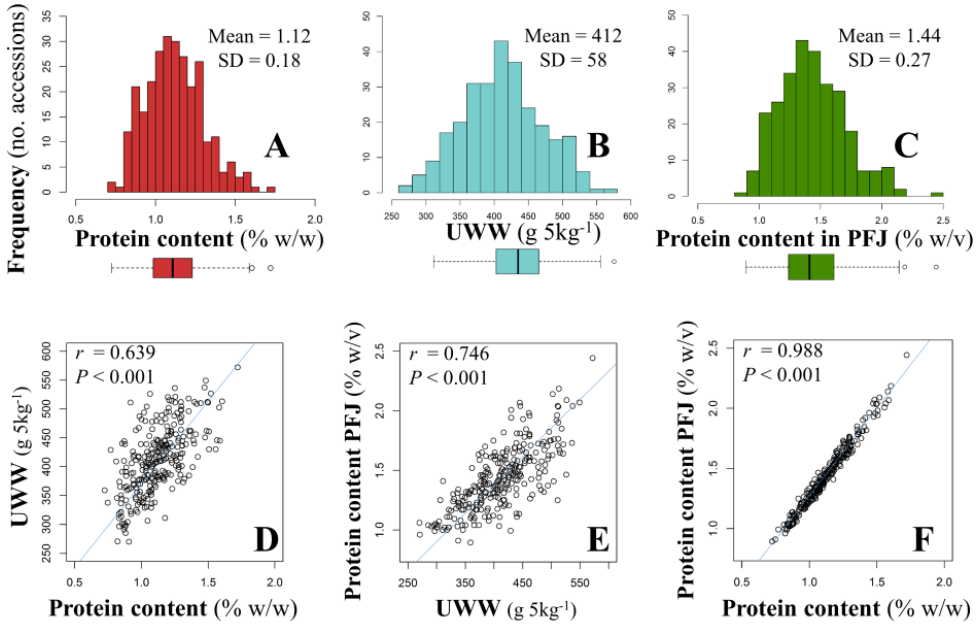

Fig. 1 Distributions and scatterplots of the phenotypic values (BLUEs). Distributions for (**A**) protein content (% w/w), (**B**) tuber under-water weight (UWW) (g 5kg^{-1}) and (**C**) protein content PFJ (potato fruit juice) (% w/v). Scatterplots for (**D**) protein content versus under-water weight (UWW), (**E**) UWW versus protein content PFJ (potato fruit juice) and (**F**) protein content versus protein content PFJ. The linear regression lines are shown in blue. SD = standard deviation. r = Pearson's correlation coefficient. P = probability value.

The broad sense heritability estimates (H^2) for protein content, protein content in PFJ and UWW were 48%, 58% and 81% respectively (**Table 1**). BLUEs for UWW and protein content in PFJ ranged between 270-572 g 5kg^{-1} and 0.89-2.44% (w/v) respectively. To evaluate correlations between the traits, scatterplots for the phenotypic values (BLUEs) were evaluated. Moderate to high correlations ($P < 0.001$) were observed for the phenotypic BLUEs (**Fig. 1**) (protein content versus UWW: $r = 0.639$; UWW versus protein content PFJ: $r = 0.746$; protein content versus protein content PFJ: $r = 0.988$).

Table 1 Variance components and heritability estimates for protein content, protein content PFJ and tuber under-water weight (UWW).

Variance components$^{\pi}$	Protein content	Protein content PFJ	UWW
G	0.029	0.069	2841
G × L	0.002	0.004	97
G × Y	0.002	0.004	61
G × L × Y	0.007	0.011	342
Residual error	0.020	0.032	172
Total	0.059	0.119	3513
H^2	0.483	0.579	0.809

π = variance components derived from REML. G = "Accession" variance; L = "Location" variance; Y = "Year" variance; G × L = "Accession" by "Location" interaction variance; G × Y = "Accession" by "Year" interaction variance; G × L × Y = "Accession" by "Location" by "Year" interaction variance; Total = sum of all variances; H^2 = broad sense heritability estimate.

Population structure

Population structure of the panel was analysed using SNP marker data from the array. Three clusters (sub-populations) were characterized (Supplementary Figure 2) and were denoted as "Processing" (N = 35), "Other" (N = 136) and "Starch" (N = 106). As the sub-populations showed unequal trait values (One-way ANOVA, $P = 2.46 \times 10^{-3}$; Supplementary Figure 1), protein content was found to be confounded with population structure. Likewise, the Q-Q plot for naive GWAS on the panel showed inflated probabilities (Supplementary Figure 3), that may have been caused by population structure. Therefore, a kinship-corrected GWAS was also performed to identify QTLs for protein content (**Fig. 2**).

Identification of QTLs

To identify QTLs for protein content, a kinship-corrected GWAS was carried out with 14,436 SNPs using the panel of 277 accessions. Three QTLs were identified above the Li and Ji threshold ($-\log_{10}(P) = 3.9$) on chromosomes *3*, *5* and *7* (**Fig. 2**), that each explained 9-12% of the phenotypic variance (R^2) (**Table 2**). The strongest association, that also exceeded the Bonferroni

threshold, was found at the start of chromosome *5* at 4.71 Mbp ($-\log_{10}(P)$ = 5.84; R^2 = 0.11). The end of chromosome *3* harboured a QTL at 60.84 Mbp ($-\log_{10}(P)$ = 4.07; R^2 = 0.11). A third QTL was positioned at the end of chromosome *7* at 50.15 Mbp ($-\log_{10}(P)$ = 3.97; R^2 = 0.12). The naive GWAS identified significant QTLs on all the twelve potato chromosomes (Supplementary Figure 3), but were expected to be false-positive associations because the trait values were confounded with population structure in the panel (Supplementary Figure 1). To dissect the potential year (season) effects, correlation analysis and GWAS were performed on the BLUEs for 2008, 2009 and 2019 (Supplementary Figures 7 and 8). As observed for the panel, GWAS on the BLUEs for 2008 showed associations at the start of chromosome *5*. The BLUEs for 2009 produced associations again at the start of chromosome *5* and at the end of chromosome *7*. For the BLUEs of 2010, associations were found at the ends of chromosomes *2* and *11*. Moderate to high correlations were observed for the BLUEs between the individual years. As the raw values for UWW and protein content in PFJ were used to correct for the values for protein content, GWAS were performed on these two traits as well (Supplementary Figure 6).

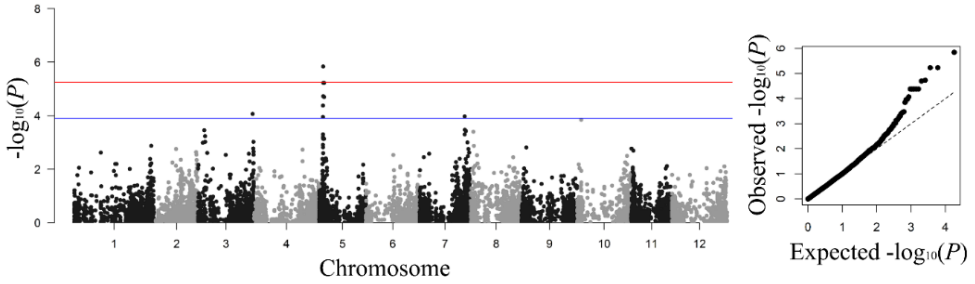

Fig. 2 Manhattan plot for the kinship-corrected GWAS for protein content on the panel (N = 277) with the Quantile-Quantile plot for the observed versus expected probabilities. Top (red) line = Bonferroni threshold at 5.3. Lower (blue) line = Li and Ji threshold at 3.9.

To pinpoint putative candidate genes from the genomic regions underlying the QTLs, linkage disequilibrium (LD)-based QTL support intervals were used as described by Vos et al. (2017). The genes underlying these intervals were

retrieved from the potato reference genome (PGSC 2011). From the longlists of genes (Supplementary Table 4), putative candidates were selected based on their annotation (gene name). As a result, the QTL interval on chromosome *3* co-localized with a nitrate transporter (60.09 Mbp). The interval on chromosome *5* harboured *StCDF1* (4.54 Mbp) and a cluster of nine nitrate transporters (6.00-7.52 Mbp). No obvious candidate genes could be proposed to be implicated with the QTL on chromosome *7*.

Haplotypes underlying QTLs

A contemporary approach by Willemsen (2018) was used to determine the haplotype-specificity of the SNP markers underlying the QTLs. Results showed that all SNPs underlying the QTL on chromosome *5* (that exceeded the Li and Ji threshold) were haplotype-specific (**Table 2**). These SNPs were haplotype-specific for a late maturity allele of *StCDF1* (Supplementary Table 2), as proposed by Willemsen (2018). Moreover, these SNPs also tagged a unique introgression segment from wild potato (*Solanum vernei* Bitter & Wittm.) as described by van Eck et al. (2017). Over the years, this introgression segment has been used by potato breeders to introduce resistance against *Globodera pallida* nematodes (the so-called *Gpa5* locus) in the genepool of cultivated potato (Rouppe van der Voort et al. 2000; Van Eck et al. 2017). Graphical genotypes of the panel, as performed by van Eck et al. (2017), illustrated that this introgression segment was mainly present in the starch varieties and starch progenitors. For these varieties and progenitors, the introgression segment was found to be present in either simplex (a single copy) or duplex (two copies) form (Supplementary Figure 4). The SNPs underlying the QTLs on chromosome *3* and *7* were not found to be haplotype-specific.

Table 2 Associated SNP markers from the kinship-corrected GWAS for protein content on the panel (N = 277)

SNP marker	Chr.	QTL peak position (bp)	Association -log$_{10}$(P) $^{\pi}$	$R^{2\,\Delta}$	Minor allele freq. of Alt SNP variant (%)	SNP variants (Ref/Alt) ‡
PotVar0020884	*3*	60 844 314	4.07	0.11	34.36 (A)	C / <u>A</u>
PotVar0078022	*5*	4 406 638	4.38	0.10	9.03 (A)	G / <u>A</u>
PotVar0078229	*5*	4 411 283	4.38	0.10	9.03 (A)	G / <u>A</u>
PotVar0078670	*5*	4 432 880	4.38	0.10	9.03 (A)	G / <u>A</u>
PotVar0078972	*5*	4 447 319	4.38	0.10	9.03 (A)	G / <u>A</u>
PotVar0079124	*5*	4 490 397	4.38	0.10	9.03 (A)	C / <u>A</u>
PotVar0080027	*5*	4 709 697	5.84	0.11	9.12 (A)	C / <u>A</u>
PotVar0080320	*5*	4 724 800	4.73	0.10	8.84 (G)	A / <u>G</u>
PotVar0129937	*5*	4 921 097	5.22	0.11	9.84 (G)	A / <u>G</u>
PotVar0117324	*5*	5 691 686	4.70	0.10	9.93 (G)	A / <u>G</u>
PotVar0117367	*5*	5 693 006	5.22	0.12	9.84 (G)	A / <u>G</u>
Solcap_snp_c2_26012	*7*	50 152 831	3.97	0.12	42.22 (G)	<u>A</u> / G

π = Li and Ji threshold at 3.9; Bonferroni threshold at 5.3. Δ = Phenotypic variance explained (*R-sq.*) by simple linear regression of SNP marker dosage scores on the BLUEs for protein content. $\ddagger$ = The favourable SNP variant for a positive association (effect) with protein content is underlined. Chr. = chromosome. QTL peak position (bp) denotes the physical coordinates (PGSC 2011). Ref = reference SNP variant. Alt = alternative SNP variant. Freq. = frequency.

Variance explained by multiple QTLs

By using multiple linear regression, we tested the cumulative effect of multiple significant SNP markers underlying QTLs together. The SNPs underlying the QTLs on chromosomes *3*, *5* and *7* together explained 22% of the variance (Supplementary Table 3). When the SNP on chromosome *5* was excluded, the QTLs on chromosomes *3* and *7* together explained 21%. The combination of SNPs on chromosomes *5* and *7* jointly explained 20%. The QTLs on chromosomes *3* and *5* jointly explained less variance (13%).

Sub-population QTLs

In an attempt to circumvent the confounding effect of population structure in the panel, GWAS was performed on the sub-populations "Starch" (N = 106) and "Other" (N = 136). The sub-population "Processing" was not included as it consisted of a relatively small number of individuals (N = 35). Kinship-corrected GWAS on the sub-population "Starch" identified one QTL above the Li and Ji threshold at the end of chromosome *3* (R^2 = 0.15) (**Fig. 3**; **Table 3**). The QTL peak caused by SNP marker PotVar0020225 was positioned 0.708 Mbp north from the QTL identified in the panel (**Table 2**). The naive GWAS on the sub-population "Starch" did not identify significant QTLs, although noticeable associations were found slightly below the thresholds at the end of chromosome *3* (Supplementary Figure 5). Kinship-corrected GWAS on the sub-population "Other" identified a QTL at the start of chromosome *5* (R^2 = 0.15) (**Fig. 3**). This sub-population QTL was positioned 0.975 Mbp (**Table 3**) south from the QTL found in the panel, that was introgressed from wild potato into the starch varieties and starch progenitors as a source of resistance against nematodes (**Table 2**). The SNPs tagging this haplotype in the sub-population "Other" were lower than the minor allele frequency (MAF) threshold of 1.5%. Therefore this haplotype remained unnoticed and could not uncover a QTL.

To verify whether or not the QTL at the start of chromosome *5* was associated with plant maturity (*StCDF1*) (Kloosterman et al. 2013), we performed conditional kinship-corrected GWAS on the sub-population "Other" by using the SNP marker "PotVar0079081" as a cofactor that tags the early maturity allele (*StCDF1.1*), as described by Willemsen (2018). This approach, reduced the significance of the original QTL at the start of chromosome *5* (from $-\log_{10}(P) = 4.46$ down to 3.19) (**Fig. 3**). This finding suggested that the maturity score of potato varieties, as largely controlled by *StCDF1.1*, indirectly influenced protein content in this sub-population. By performing the cofactor analysis, an otherwise masked QTL was uncovered at the end of chromosome *12* (Peak SNP: "PotVar0052807"; 59,294,858 bp; $-\log_{10}(P) = 4.63$). Naive GWAS on the sub-population "Other" showed inflated associations that

probably caused false-positive QTLs on chromosomes *1*, *2*, *3*, *4*, *5*, *7* and *10* (Supplementary Figure 5).

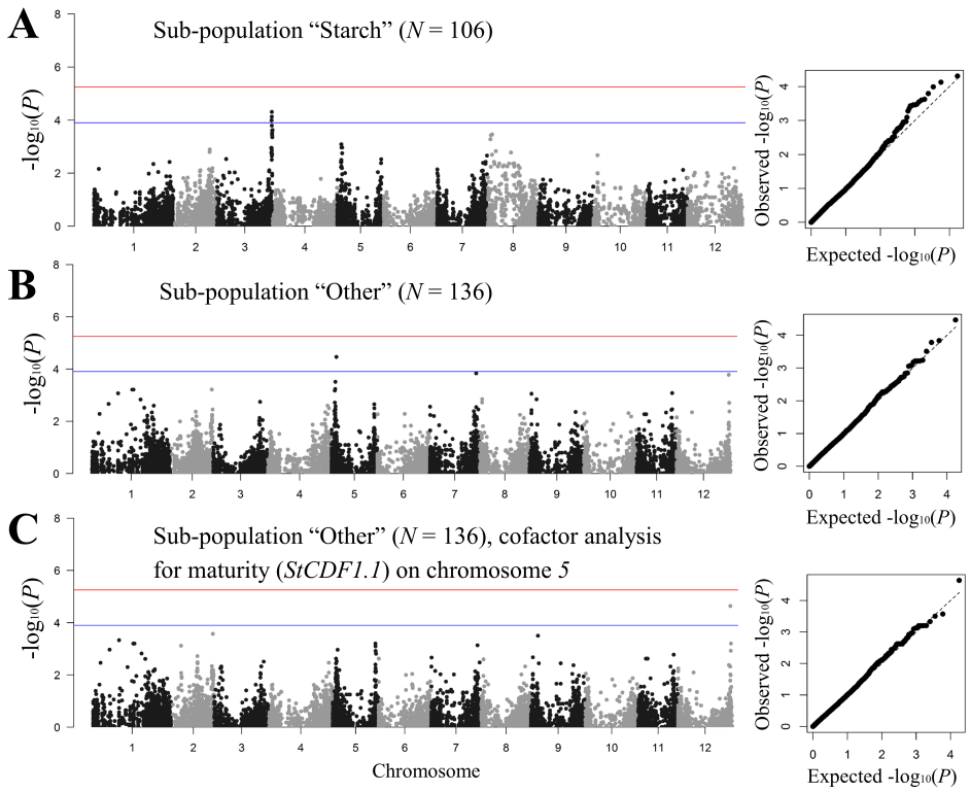

Fig. 3 Manhattan plots for the kinship-corrected GWAS for protein content on sub-populations (**A**) "Starch" ($N = 106$), (**B**) "Other" ($N = 136$) and (**C**) "Other" ($N = 136$) by including SNP marker "PotVar0079081" as a cofactor for early maturity (*StCDF1.1*). The Quantile-Quantile plots for the observed versus expected probabilities are shown at the right. Top (red) line = Bonferroni threshold at 5.3. Lower (blue) line = Li and Ji threshold at 3.9. P = probability value.

Table 3 Associated SNP markers from the kinship-corrected GWAS for protein content on sub-populations "Starch" and "Other"

SNP marker	Chr.	QTL peak position (bp)	Association score -$\log_{10}$(*P*)	$R^{2\,\Delta}$	Minor allele freq. (%)	SNP variants (Ref/Alt) ǂ
Sub-population "Starch" (*N* = 106)						
PotVar0020225	*3*	61 551 606	4.48	0.15	33.49 (A)	G / A
PotVar0020407	*3*	61 494 613	3.96	0.14	44.82 (G)	G / A
PotVar0020216	*3*	61 551 841	3.96	0.13	43.87 (G)	G / A
PotVar0020017	*3*	61 892 531	3.91	0.15	49.05 (A)	G / A
Sub-population "Other" (*N* = 136)						
PotVar0117190	*5*	5 684 749	4.46	0.15	20.77 (A)	G / A

Δ = Variance explained (*R-sq.*) by SNP markers, from simple linear regression of SNP marker dosage scores on the BLUEs for protein content. ǂ = The favourable SNP variant for a positive association (effect) with protein content is underlined. Chr. = chromosome. QTL peak position = physical chromosome position of SNP marker in the potato reference genome pseudomolecules v4.03 (PGSC 2011). Ref = reference SNP variant. Alt = alternative SNP variant. Freq. = frequency. *P* = probability value.

Discussion

GWAS as a tool to detect QTLs

We used GWAS to shed light on the complex genetic architecture of protein content in potato. We identified QTLs with minor effects on chromosomes *3*, *5*, *7* and *12* (**Fig. 2**; **Fig. 3**). The QTLs identified on chromosomes *3* and *5*, coincided with previous studies (Acharjee et al. 2018; Klaassen et al. 2019; Werij 2011). For chromosome *3*, the QTL identified in the entire panel was also observed in the sub-population "Starch". For chromosome *5*, we uncovered an introgression segment from wild potato that was associated with protein content (Supplementary Figure 4). This introgressed segment harboured a late maturity allele of *StCDF1* (Supplementary Table 2), as well as the *Gpa5* resistance allele against potato cyst nematodes (*Globodera pallida*). However, the SNPs tagging this introgression segment did not bring forth a QTL in the sub-population "Starch", even though the allele frequency

of these SNPs in this sub-population was considerable (9-10%). We also observed that the additive effect of this QTL was lower than expected when combined with the other two QTLs on chromosomes *3* and *7* (Supplementary Table 3). We showed that protein content was confounded with population structure in the panel. This result was likely caused by higher BLUEs values for protein content in the sub-population "Starch" (Supplementary Figure 1). Therefore, we propose that the QTL on chromosome *5* in the panel could be an artefact. Validation studies, for instance using bi-parental mapping populations, may confirm the relevance of SNPs underlying this QTL for use in breeding to improve protein content. If these SNPs are to be used for breeding, they will at least provide a source of resistance against cyst nematodes and contribute towards a later maturity index due to *StCDF1*. In the sub-population "Other" we also identified a QTL at the start of chromosome *5*. Conditional GWAS on this sub-population showed that this association was not caused by the introgression segment from wild potato. Instead, this QTL coincided with the early maturity allele of *StCDF1* (*StCDF1.1*). Findings from GWAS on the panel as well as the sub-populations, showed that different haplotypes at the start of chromosome *5* were associated with protein content.

To the best of our knowledge, the identified QTLs on chromosomes *7* and *12* have not been described before in literature. Bi-parental populations, that descend from crosses between protein-rich varieties, can be used to test/validate and stack multiple copies of favourable variants/alleles for multiple protein content QTLs simultaneously. For instance, the cross between the starch varieties *Kartel* × *Seresta* will allow the SNPs underlying all three QTLs identified in the panel here, to segregate in nulliplex (null), simplex (one) and duplex (two) dosages in the F_1 progeny. This cross will provide improved insight into the cumulative effects of the underlying haplotypes. Our results, as presented in Supplementary Table 3, suggest both additive and epistatic effects of the SNPs (alleles). We observed that the effects of genotype-by-environment (G × E) interactions were small to moderate for protein content (**Table 1**). On the other hand, a large proportion of variance was ascribed to the residuals (error). Hence, future genetic studies

on protein content may be improved by reducing the residual error in these experiments.

Missing heritability

Studies in soybean, wheat and maize describe protein content as a complex trait that is governed by multiple genes and environmental factors. We estimated a moderate trait heritability for protein content (H^2 = 0.48). This H^2 value ranged between 40-74%, i.e. in line with previous studies (Klaassen et al. 2019; Werij 2011). GWAS on the panel identified three QTLs, that cumulatively explained 22% of the variance. Hence, we demonstrate a clear example of missing heritability. Several factors may have contributed to this finding, that include the limited statistical power to detect loci with small effects, interactions between loci, effects or rare variants and potential banishment of true-positive QTLs due to kinship correction. Alternatively, overestimation of the broad sense heritability estimate (H^2) may also have occurred. In any case, it should be noted that our H^2 will be much larger than the narrow sense (h^2) estimate.

To optimize the detection of QTLs by GWAS, the design and methodology should be considered carefully. Using more individuals will likely increase statistical power, as shown in numerous human and crop genetic studies *e.g.* for soybean (Bandillo et al. 2015). Optimization of GWAS will likely identify loci with minor effects or those caused by rare variants with a low allele frequency. Certainly the population structure, distribution of the phenotypic values, as well as the ascertainment bias of SNPs in marker arrays should be considered beforehand as proposed by Vos (2016).

Correlation between tuber protein content and under-water weight

For other crops, a negative correlation is often observed between protein content and other major (seed) storage compounds *e.g.* oil content in soybean (Patil et al. 2017). Interestingly, while expecting a similar trade-off in potato, we found a moderate positive correlation (r = 0.64) between protein content and under-water weight (UWW: a proxy for starch content) (**Fig. 1**). Therefore, selection pressure for high UWW in the starch genepool, aimed to

increase starch content, may have coincided with unconscious selection for high protein content (Supplementary Figure 1). Kinship-corrected GWAS on UWW in the panel did not identify potential associations between UWW and maturity alleles of *StCDF1* at the start of chromosome *5* (Supplementary Figure 6). The statistical power produced by the 277 individuals here may have been insufficient to uncover significant signals due to the complex (polygenic) genetic architecture of starch content in potato. A positive correlation between protein content and UWW suggests that these traits may be (partly) interrelated due to shared biological mechanisms. It is well established that photosynthesis-derived carbon and nitrogen assimilation pathways are connected and tightly controlled in plants. Molecular studies have shown that intracellular glucose is used by plants to synthesize both protein and starch (Bihmidine et al. 2013). Reduced levels of ADP-glucose (*i.e.* glucosyl donor of glucose) by inactivated ADP-glucose pyrophosphorylase (AGPase) in barley mutants, was accompanied with the down-regulation of genes related to amino acid and storage protein biosynthesis (Faix et al. 2012). Therefore, the genes that regulate protein content in potato may affect starch content, yet this point remains to be addressed in future studies. Unravelling the positive correlation between protein and starch content in potato, will certainly be dealt with in future studies.

Putative candidate genes for protein content

To pinpoint putative candidate genes, we used LD-bound QTL support intervals to narrow down on genomic regions. This approach identified several candidates that included *StCDF1* (maturity) and nitrate transporters (Supplementary Table 4). Conditional GWAS on the sub-population "Other" showed that a late maturity allele of *StCDF1* was positively associated with protein content. Nitrate transporters are known to function in the uptake and allocation of inorganic nitrate (NO_3^-) in plants (Hsu and Tsay 2013; Léran et al. 2014). Nitrate is the predominant nitrogen-containing macronutrient in aerobic soils under temperate climatic conditions. Hence, allelic variants of nitrate transporters may differ in nitrate uptake and interaction with nitrogen-responsive genes that ultimately affect protein content, as proposed for rice (Hu et al. 2015). Future molecular studies on the above mentioned candidate

genes that include gene expression, overexpression and knock-out studies, are certainly relevant to study their biological functions and effects on protein content in potato.

Acknowledgements

The authors thank the potato breeding companies from the *Centre for Biosystems and Genomics* (CBSG) consortium for providing the raw data.

Supplementary data are available online at:

https://link.springer.com/article/10.1007/s11032-019-1070-8

References

Acharjee A, Chibon P-Y, Kloosterman B, America T, Renaut J, Maliepaard C, Visser RG (2018) Genetical genomics of quality related traits in potato tubers using proteomics. BMC Plant Biology 18 (1):20

Alexandratos N (1999) World food and agriculture: outlook for the medium and longer term. Proceedings of the National Academy of Sciences 96 (11):5908-5914

Balyan HS, Gupta PK, Kumar S, Dhariwal R, Jaiswal V, Tyagi S, Agarwal P, Gahlaut V, Kumari S (2013) Genetic improvement of grain protein content and other health-related constituents of wheat grain. Plant Breeding 132 (5):446-457

Bandillo N, Jarquin D, Song Q, Nelson R, Cregan P, Specht J, Lorenz A (2015) A population structure and genome-wide association analysis on the USDA soybean germplasm collection. The Plant Genome 8 (3)

Bárta J, Bártová V, Zdráhal Zk, Šedo O (2012) Cultivar Variability of Patatin Biochemical Characteristics: Table versus Processing Potatoes (Solanum tuberosum L.). Journal of Agricultural and Food Chemistry 60 (17):4369-4378

Bihmidine S, Hunter C, Johns C, Koch K, Braun D (2013) Regulation of assimilate import into sink organs: update on molecular drivers of sink strength. Frontiers in Plant Science 4 (177)

Bradshaw JE, Hackett CA, Pande B, Waugh R, Bryan GJ (2008) QTL mapping of yield, agronomic and quality traits in tetraploid potato (Solanum tuberosum subsp. tuberosum). Theoretical and Applied Genetics 116 (2):193-211

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics 138 (3):963-971

Creusot N, Wierenga PA, Laus MC, Giuseppin ML, Gruppen H (2011) Rheological properties of patatin gels compared with β-lactoglobulin, ovalbumin, and glycinin. Journal of the Science of Food and Agriculture 91 (2):253-261

D'hoop BB, Paulo MJ, Mank RA, Van Eck HJ, Van Eeuwijk FA (2008) Association mapping of quality traits in potato (Solanum tuberosum L.). Euphytica 161 (1-2):47-60

D'hoop BB, Paulo MJ, Visser RGF, van Eck HJ, van Eeuwijk FA (2011) Phenotypic analyses of multi-environment data for two diverse tetraploid potato collections: comparing an academic panel with an industrial panel. Potato Research 54:157-181

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4 (2):359-361

Edens L, Plijter JJ, Van DLJAB (1999) Novel food compositions. Google Patents

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14 (8):2611-2620

Faix B, Radchuk V, Nerlich A, Hümmer C, Radchuk R, Emery RN, Keller H, Götz KP, Weschke W, Geigenberger P (2012) Barley grains, deficient in cytosolic small subunit of ADP-glucose pyrophosphorylase, reveal coordinate adjustment of C: N metabolism mediated by an overlapping metabolic-hormonal control. The Plant Journal 69 (6):1077-1093

Gao X, Starmer J, Martin ER (2008) A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. Genetic Epidemiology 32 (4):361-369

Hill AJ, Peikin SR, Ryan CA, Blundell JE (1990) Oral administration of proteinase inhibitor II from potatoes reduces energy intake in man. Physiology & Behavior 48 (2):241-246

Hsu P-K, Tsay Y-F (2013) Two phloem nitrate transporters, NRT1. 11 and NRT1. 12, are important for redistributing xylem-borne nitrate to enhance plant growth. Plant Physiology 163 (2):844-856

Hu B, Wang W, Ou S, Tang J, Li H, Che R, Zhang Z, Chai X, Wang H, Wang Y (2015) Variation in NRT1. 1B contributes to nitrate-use divergence between rice subspecies. Nature Genetics 47 (7):834

Hwang E-Y, Song Q, Jia G, Specht JE, Hyten DL, Costa J, Cregan PB (2014) A genome-wide association study of seed protein and oil content in soybean. BMC Genomics 15 (1):1

Johnson RC, Nelson GW, Troyer JL, Lautenberger JA, Kessing BD, Winkler CA, O'Brien SJ (2010) Accounting for multiple comparisons in a genome-wide association study (GWAS). BMC Genomics 11 (1):724

Jørgensen M, Bauw G, Welinder KG (2006) Molecular Properties and Activities of Tuber Proteins from Starch Potato Cv. Kuras. Journal of Agricultural and Food Chemistry 54 (25):9389-9397

Kang HM, Sul JH, Service SK, Zaitlen NA, Kong S-y, Freimer NB, Sabatti C, Eskin E (2010) Variance component model to account for sample structure in genome-wide association studies. Nature Genetics 42:348

Karn A, Gillman JD, Flint-Garcia SA (2017) Genetic analysis of teosinte alleles for kernel composition traits in maize. G3: Genes | Genomes | Genetics 7 (4):1157-1164

Klaassen MT, Bourke PM, Maliepaard C, Trindade LM (2019) Multi-allelic QTL analysis of protein content in a bi-parental population of cultivated tetraploid potato. Euphytica 215 (2):14

Kloosterman B, Abelenda JA, Gomez MdMC, Oortwijn M, de Boer JM, Kowitwanich K, Horvath BM, van Eck HJ, Smaczniak C, Prat S, Visser RGF, Bachem

CWB (2013) Naturally occurring allele diversity allows potato cultivation in northern latitudes. Nature 495 (7440):246-250

Kudo K, Onodera S, Takeda Y, Benkeblia N, Shiomi N (2009) Antioxidative activities of some peptides isolated from hydrolyzed potato protein extract. Journal of Functional Foods 1 (2):170-176

Kunkel R, Campbell GS (1987) Maximum potential potato yield in the Columbia Basin, USA: Model and measured values. American Potato Journal 64 (7):355-366

Léran S, Varala K, Boyer J-C, Chiurazzi M, Crawford N, Daniel-Vedele F, David L, Dickstein R, Fernandez E, Forde B (2014) A unified nomenclature of NITRATE TRANSPORTER 1/PEPTIDE TRANSPORTER family members in plants. Trends in Plant Science 19 (1):5-9

Li J, Ji L (2005) Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity 95 (3):221

Ortiz-Medina E (2006) Potato tuber protein and its manipulation by chimeral disassembly using specific tissue explantation for somatic embryogenesis. PhD dissertation. McGill University. Department of Plant Science. Montreal, Quebec, Canada

Patil G, Mian R, Vuong T, Pantalone V, Song Q, Chen P, Shannon GJ, Carter TC, Nguyen HT (2017) Molecular mapping and genomics of soybean seed protein: a review and perspective for the future. Theoretical and Applied Genetics 130 (10):1975-1991

PGSC (2011) Genome sequence and analysis of the tuber crop potato. Nature 475 (7355):189-195

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155 (2):945-959

Rosyara UR, De Jong WS, Douches DS, Endelman JB (2016) Software for genome-wide association studies in autopolyploids and its application to potato. The Plant Genome 9 (2)

Rouppe van der Voort J, van der Vossen E, Bakker E, Overmars H, van Zandvoort P, Hutten R, Klein Lankhorst R, Bakker J (2000) Two additive QTLs conferring broad-spectrum resistance in potato to Globodera pallida are localized on resistance gene clusters. Theoretical and Applied Genetics 101 (7):1122-1130

Ruseler-van Embden JGH, Van Lieshout L, Smits S, Van Kessel I, Laman J (2004) Potato tuber proteins efficiently inhibit human faecal proteolytic activity: implications for treatment of peri-anal dermatitis. European Journal of Clinical Investigation 34 (4):303-311

Sabaté J, Soret S (2014) Sustainability of plant-based diets: back to the future. The American Journal of Clinical Nutrition 100 (Suppl 1):476S-482S

Segura V, Vilhjalmsson BJ, Platt A, Korte A, Seren U, Long Q, Nordborg M (2012) An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nature Genetics 44 (7):825

Sharma SK, MacKenzie K, McLean K, Dale F, Daniels S, Bryan GJ (2018) Linkage Disequilibrium and Evaluation of Genome-Wide Association Mapping Models in Tetraploid Potato. G3: Genes | Genomes | Genetics 8 (10):3185-3202

Smith PK, Krohn RI, Hermanson G, Mallia A, Gartner F, Provenzano M, Fujimoto E, Goeke N, Olson B, Klenk D (1985) Measurement of protein using bicinchoninic acid. Analytical Biochemistry 150 (1):76-85

Tilman D, Clark M (2014) Global diets link environmental sustainability and human health. Nature 515 (7528):518

Van Eck HJ, Vos PG, Valkonen JP, Uitdewilligen JG, Lensing H, de Vetten N, Visser RG (2017) Graphical genotyping as a method to map Ny (o, n) sto and Gpa5 using a reference panel of tetraploid potato cultivars. Theoretical and Applied Genetics 130 (3):515-528

Voorrips RE, Gort G, Vosman B (2011) Genotype calling in tetraploid species from bi-allelic marker data using mixture models. BMC Bioinformatics 12 (1):172

Vos P (2016) Development and application of a 20K SNP array in potato. PhD dissertation. Chair group: Plant Breeding. Wageningen University, Wageningen, The Netherlands. Retrieved from http://edepot.wur.nl/392278

Vos PG, Paulo MJ, Voorrips RE, Visser RG, van Eck HJ, van Eeuwijk FA (2017) Evaluation of LD decay and various LD-decay estimators in simulated and SNP-array data of tetraploid potato. Theoretical and Applied Genetics 130 (1):123-135

Vos PG, Uitdewilligen JG, Voorrips RE, Visser RG, van Eck HJ (2015) Development and analysis of a 20K SNP array for potato (Solanum tuberosum): an insight into the breeding history. Theoretical and Applied Genetics 128 (12):2387-2401

Werij JS (2011) Genetic analysis of potato tuber quality traits. PhD dissertation. Chair group: Plant Breeding. Wageningen University, Wageningen, The Netherlands. Retrieved from http://edepot.wur.nl/183746

Willemsen JH (2018) The identification of allelic variation in potato. PhD dissertation. Chair group: Plant Breeding. Wageningen University, Wageningen, The Netherlands

Chapter 3

Multi-allelic QTL analysis of protein content in a bi-parental population of cultivated tetraploid potato

Michiel T. Klaassen[1,2], Peter M. Bourke[1], Chris Maliepaard[1], Luisa M. Trindade[1]

[1] Wageningen University & Research, Plant Breeding, Wageningen, The Netherlands

[2] Aeres University of Applied Sciences, Department of Applied Research, Dronten, The Netherlands

This chapter has been published in:

***Euphytica* (2019) 215 (2):14**

DOI: 10.1007/s10681-018-2331-z

Abstract

Protein content is a key quality trait for the potato starch industry. The objective of this study was to identify allele-specific quantitative trait loci (QTLs) for tuber protein content in cultivated potato (*Solanum tuberosum* L.) at the tetraploid level. We analysed 496 full-sib F_1 clones in a three-year field trial to dissect the complex genetic architecture of soluble tuber protein content. Genotypic data from a 60K single nucleotide polymorphism (SNP) array was used for SNP dosage scoring, constructing homologue specific linkage maps and assembly of a dense integrated chromosomal linkage map. From the integrated map, probabilistic multi-locus identity-by-descent (IBD) haplotypes (alleles) were estimated and used to detect associations between the IBD haplotypes and the phenotypic trait values. Moderate levels of trait heritability were estimated between 40 and 74% that corresponded with previous studies. Our contemporary naive analysis identified potential additive QTLs on chromosomes *2*, *3*, *5* (top arm) and *9* across the years. Moreover, cofactor QTL analysis identified two masked QTLs on chromosomes *1* and *5* (lower arm). The QTLs on chromosomes *2*, *5* (lower arm) and *9* are reported here for the first time. The QTLs that we identified on chromosomes *1*, 3 and *5* (top arm) show overlap with previous studies for protein content in potato. Collectively the naive QTLs explained 12 to 17% of the phenotypic variance. The underlying alleles of the QTLs provided both positive and negative effects on the phenotype. Our work uncovers the complex genetic architecture of this trait and describes potential breeding strategies for improvement. As protein has emerged as a high-value component from industrial potato starch production, the dissection of the genetic architecture and subsequent improvement of this trait by breeding has great economic and environmental relevance.

Keywords

Protein content, Potato, Tetraploid, Haplotypes, Alleles, QTL analysis

Introduction

Potato (*Solanum tuberosum* L.) is the fourth most important food crop worldwide (FAO 2014) and is becoming increasingly important in developing countries. It is a major source of starch, protein, vitamins and minerals and therefore an important crop for both human consumption and the starch industry (Jørgensen et al. 2006). When starch is industrially extracted from potato tubers, large quantities of potato fruit juice (PFJ) are released as an aqueous by-product that contains soluble protein (Bárta et al. 2012). In the past, PFJ was not valorised by the potato starch industry and was treated as a waste-stream. Consequently, PFJ was discharged into rivers and channels which often resulted in environmental pollution. Nowadays, functional proteins are extracted from PFJ by innovative industrial processors to create added-value in the starch potato production chain. Functional potato proteins are economically valuable as food ingredients due to their techno-functional properties such as gelling behaviour (Creusot et al. 2011), anti-oxidant properties (Kudo et al. 2009) and high nutritional value (Bártová and Bárta 2009). The concentration of soluble protein in PFJ of commercial varieties is known to range from 1-1.5% (Ortiz-Medina 2006). Soluble protein in PFJ is generally classified into three main groups consisting of patatin, protease inhibitors and a group of high-molecular weight proteins (Pots et al. 1999). The processing of tubers with high levels of protein content is economically relevant for the potato starch industry as these compounds render high economic value. Therefore, improving protein content in starch potato varieties has emerged as a topic for innovation amongst starch potato breeders. Shedding light on the genetic architecture of protein content – by characterizing underlying QTLs – provides relevant insight for defining strategies on how to improve this trait by means of breeding.

The genetic factors underlying protein content in potato are poorly studied. To the best of our knowledge, only two genetic studies describing QTLs have been published. These studies involved the use of a diploid bi-parental potato mapping population of limited agronomic value (Acharjee et al. 2018; Werij 2011). In diploid populations the broad sense heritability of protein content has been estimated between 56-66% (Lu et al. 2012; Werij 2011). These findings indicate that a moderate proportion of the trait variance can be ascribed to genetic factors within a particular experimental setup. Acharjee et al. (2018) and Werij (2011) identified QTLs for protein content on chromosomes *1*, *3* and *5*, illustrating that these genetic loci affect the level of soluble protein content in potato tubers. At present, no QTL

study has been reported for protein content in cultivated tetraploid potato. The identification of QTLs in tetraploid potato provides relevant insight for breeding as most crosses are made using tetraploids that contain four sets of homologous chromosomes ($2n = 4x = 48$). Genetic studies in other crops reveal that protein content is a quantitative trait that is controlled by cumulative actions of both genetic and environmental factors. In soybean, wheat and maize it has been shown that protein content is regulated by multiple loci that are likely to be influenced by genotype-by-environment interactions (Balyan et al. 2013; Hwang et al. 2014; Karn et al. 2017). Marker-assisted breeding for elite varieties with enhanced levels of protein content is therefore challenging without a basic understanding of the genetic architecture and QTLs of the trait.

As potato breeding is almost exclusively performed by making tetraploid crosses, it is relevant to perform genetic studies at a tetraploid level. Performing these studies is challenging due to the complexities of dealing with tetrasomic inheritance, allowing for pairing between all sets of homologous chromosomes, and the computation power needed to analyse vast quantities of marker-data that originate from commonly used single-nucleotide polymorphism (SNP) arrays. In recent years however, great progress has been made in the development of novel, rapid and user-friendly tools for the construction of chromosomal linkage maps from genetic marker-based data and mapping of trait-derived QTLs in outbred tetraploid species that are highly heterozygous (*i.e.* potato). New tools allow the genetic analysis of large numbers of SNPs available from modern genotyping arrays. These include statistical methods for the construction of high-density SNP-based chromosomal linkage maps (Bourke et al. 2016; Hackett et al. 2014; Hackett et al. 2013) and the reconstruction of multi-locus probabilistic haplotypes in outcrossing tetraploids (Zheng et al. 2016). Simulations by Zheng et al. (2016) have illustrated that probabilistic haplotype reconstruction is able to quantify the presence of all possible combinations of parental alleles in the progeny and that it can deal with quadrivalent pairing of four synapsed homologous chromosomes during the first stages of meiosis. This approach is robust in handling the possible but low-frequent occurrence of double reduction products from quadrivalent pairing and in using genetic maps containing some degree of errors and missing allele dosage information in parents and offspring. The application of these methods enables the dissection of the genetic architecture of complex quantitative traits in tetraploid potato.

In this study, we investigated the genetic architecture of protein content by QTL analysis of a large bi-parental tetraploid mapping population ($2n = 4x = 48$) from a

cross between two contrasting commercial varieties. Phenotypic data from 496 F_1 individuals was collected from field trials that were carried out over three consecutive years. Genotypic data originating from a 60K SNP array were transformed into multi-locus probabilistic haplotypes for QTL analysis. This transformation step included SNP dosage scoring, construction of homologue specific linkage maps, construction of an integrated linkage map and estimation of identity-by-descent (IBD). We identified QTLs on several chromosomes and compared these with previous studies. The aim of this study was to shed light on the genetic architecture of protein content in cultivated tetraploid potato by identifying and describing the genetic factors that modulate this trait.

Materials and methods

Plant material and field experiments

The complete tetraploid population ($2n = 4x = 48$) consisting of 972 full-sib F_1 clones originated from a cross between the varieties *Altus* and *Colomba*. The female parent *Altus* is a starch potato variety (Averis Seeds, Valthermond, The Netherlands). *Altus* has a high level of tuber protein content and descended from a cross between *KA 87-2306* and *Kartel*. The male parent *Colomba* is a consumption variety (HZPC, Metslawier, The Netherlands). *Colomba* has a low level of tuber protein content and resulted from a cross between *Carrera* and *Agata*. Field experiments were conducted in 2012, 2013 and 2014 during the conventional potato growing season in the northern region of the Netherlands (April to September), in Valthe (2012), Nieuw-Weerdinge (2013) and Grolloo (2014). A randomly selected subset of this population, consisting of 496 F_1 clones, and the two parental varieties were grown from seed tubers in two randomized blocks in 2013 and 2014 and each consisted of six plant plots per experimental unit. In 2012 the population was grown in a single block. The trial was well balanced as all but two clones were grown in all three years of the trial.

Quantification of soluble protein content

Soluble protein content was quantified in potato fruit juice (PFJ) using SPRINT™ Rapid Protein Analyser (CEM Corporation, NC, U.S.A.). Purified potato tuber protein (AVEBE, Veendam, The Netherlands) was used as a standard. Each PFJ sample was measured in two technical replicates. PFJ samples were extracted from

5 kg batches of representative fresh tubers from all individual F_1 clones and both parental varieties after measuring the fresh weight and under-water weight.

The tubers were sliced, mixed and processed with a juice extractor (Rotor Lips Ltd., Uetendorf, Switzerland) and PFJ was collected directly. The tubes containing the PFJ samples were kept cold on ice. After 10 minutes of settling time, a second PFJ sample was collected from the supernatant phase of the sample and 1% (v/v) of 5% sodium metabisulfite (w/v) was added to inhibit enzymatic browning. The PFJ was then centrifuged at 15,000 × g for 5 minutes and the supernatant was collected and stored at -20^{0}C until use. Tuber dry matter was inferred from tuber under-water weight as described in previous studies (Bradshaw et al. 2008; Sverrisdóttir et al. 2017). The tuber moisture content was computed as follows:

$$Tuber\ moisture\ content = 100 - tuber\ dry\ matter\ content \quad (1),$$

where tuber moisture and tuber dry matter contents are expressed in percentages.

The phenotypic values of protein were computed by correcting protein content in PFJ by tuber moisture content as follows:

$$Protein\ content = \frac{protein\ content\ in\ PFJ \times tuber\ moisture\ content}{100} \quad (2),$$

where protein content in PFJ is expressed in milligram protein per milliliter PFJ (1% = 10 mg / ml PFJ) (w/v) and tuber moisture content in percentages.

Variance components and trait heritability

The variance components of protein content were computed from the mean squares (MS) values as output from the one-way (within year) and two-way analysis (between years) of variance (ANOVA). The variance components were computed as follows:

$$\sigma^2_E = \mathrm{MS_E} \quad (3),$$

$$\sigma^2_{G\ \text{one-way}} = \frac{MS_G - MS_E}{r} \quad (4),$$

$$\sigma^2_{G\ \text{two way}} = \frac{MS_G - MS_{G\times Y}}{r \times y} \quad (5),$$

$$\sigma^2_{G\times Y\ \text{two way}} = \frac{MS_{G\times Y} - MS_E}{r} \quad (6),$$

where σ^2_G, $\sigma^2_{G\times Y}$ and σ^2_E are the variance components of the F_1 clones, F_1 clones by year interaction and residuals respectively. The number of years and the number of biological replicates are expressed in the terms *y* and *r* respectively.

The following ANOVA models were used for analysis:

$$ANOVA\ within\ year: y_{ij} = \mu + \tau_i + \alpha_j + \varepsilon_{ij} \quad (7),$$

$$ANOVA\ between\ years: y_{ijk} = \mu + \tau_i + \beta_k + \alpha\beta_{jk} + \tau\beta_{ik} + \varepsilon_{ijk} \quad (8),$$

where y_{ij} and y_{ijk} represent the protein content values of the *i*-th clone, in the *j*-th block, in the *k*-th year, μ is the overall mean response, τ_i is the effect of the *i*-th clone, α_j is the effect of the *j*-th block, β_k is the effect of the *k*-th year, $\alpha\beta_{jk}$ is the nested effect of the *j*-th block in the *k*-th year, $\tau\beta_{ik}$ is the effect of the interaction between the *i*-th clone of the *k*-th year and ε_{ijk} is the random error term. The broad sense heritability estimates (H^2) were computed as follows:

$$H^2\ within\ year = \frac{{\sigma^2}_G}{({\sigma^2}_G + \frac{{\sigma^2}_E}{r})} \quad (9),$$

$$H^2\ between\ years = \frac{{\sigma^2}_G}{({\sigma^2}_G + \frac{{\sigma^2}_{G\times Y}}{y} + \frac{{\sigma^2}_E}{r\times y})} \quad (10).$$

The phenotypic values of protein content were collected from multi-year field trials at the three different locations. A mixed model was used to obtain best linear unbiased estimates (BLUEs) with adjusted mean values for the F_1 clones for the phenotypic values from 2012 to 2014. The estimation was done using restricted maximum likelihood (REML) to compute the response as reported (Björn et al. 2011):

$$Response = \mu + genotype + year + error \quad (11),$$

where the μ is the overall mean response, the genotype effect is fixed and year effects and errors are random terms.

Genotyping using a single-nucleotide polymorphism (SNP) array

Genotyping of the complete population of 972 individuals and the parental varieties was performed with the 60K Axiom SNP marker array. This array consists of a subset of the 20K SNPs from the SolSTW Infinium SNP array (Vos et al. 2015) and

an additional 40K SNPs that originated from RNA sequences of both parental varieties used in this study (unpublished data).

Data processing and genotype calling

Allele dosage scores were assigned to the SNP markers using the fitTetra R package (Voorrips et al. 2011) as previously described (Bourke et al. 2015). SNP allele dosage scores assigned by fitTetra were tested with the function CheckF1 (Bourke et al. 2018b) in R to identify the best-fitting segregation model for the SNPs. SNPs that did not correspond to the assumed segregation were discarded. SNPs with high skewness (using a chi-square test, $\alpha = 0.001$) or more than 5% missing values were removed. Also, F_1 clones with more than 10% missing SNPs were removed. The complete mapping population of 972 individuals was used to construct a high-density tetraploid integrated chromosomal linkage map.

Chromosomal linkage maps and identity-by-descent estimation

The complete population of 972 F_1 clones was used for constructing the chromosomal linkage maps according to methods described by Bourke et al. (2016) with minor modifications. First, simplex × nulliplex markers were assigned to 12 putative chromosomal clusters at a linkage LOD score threshold of 10, after which they were separated into putative homologue clusters at a LOD score threshold of 30. Per chromosome, pairwise recombination frequencies and LOD scores were calculated between all marker segregation types and marker alleles were phased and assigned to homologues. Next, a developmental version of the MDSmap software (Preedy and Hackett 2016) was used to order the markers. This resulted in twelve integrated chromosomal linkage groups representing the twelve potato chromosomes. These maps were produced using unconstrained weighted metric multi-dimensional scaling with the squared LOD scores for linkage as weights and using Haldane's mapping function. This was followed up by principal curve fitting in two dimensions to order the markers. Outlying markers in principal curve analysis, as judged by the eye, and those with a nearest-neighbour fit exceeding 5 were removed to select only high-quality markers as described by Preedy and Hackett (2016). Up to three rounds of MDSmap were performed until all outlying markers were removed. After the first and second round of MDSmap, in total 86 and 30 markers, respectively, were removed as possible outliers. In the third round, no further outliers were identified, resulting in stable integrated chromosomal linkage maps. Linkage groups were renumbered according to the reference potato genome

sequence (PGSC 2011) using the known assignments of SNPs on the physical map containing the DNA sequence assembly of the twelve potato chromosomes (PGSC pseudomolecules v4.03).

The identity-by-descent (IBD) probabilistic haplotypes were estimated using TetraOrigin (Zheng et al. 2016). SNPs from each possible segregation type were selected at centiMorgan (cM) map positions (rounded off to 1 decimal place), with preference given to markers with the smallest amount of missing data whenever multiple markers occupied the same position. TetraOrigin (Zheng et al. 2016) was run in Mathematica version 10 (Wolfram Research Inc., Champaign, Illinois, USA) with bivalentPhasing set to True and bivalentDecoding set to False for taking into account the occurrence of double reduction in the probabilistic haplotypes of the F_1 clones. The allele dosage error probability for both parents (*epsF*) and F_1 clones (*eps*) were set to 0 and 0.001 respectively (Bourke 2018). These setting were used due to the high quality (confidence) dosage scores that were assigned to both parents from technical replicates ($N = 12$) of the SNP array. Moreover, these setting have been shown to be effective and appropriate for use in TetraOrigin as demonstrated by Zheng et al. (2016). For the other parameters, the default settings were used (*maxStuck* = 10, *maxIteration* = 100, *minRepeatRun* = 3, *maxPhasingRun* = 20).

Naive QTL analysis

A naive single-locus QTL analysis was carried out for all chromosomes. This analysis was carried out using the IBD probabilistic haplotypes produced by TetraOrigin, after splines were fitted on a grid at cM map positions. The model used has previously been described as Kempthorne's "additive model" (Hackett et al. 2014; Hackett et al. 2013), with the difference that all possible genotypes under a quadrivalent model allowing for double reduction were included. The terms in the model correspond to the haplotype probabilities X_i from *Altus* (parent 1: $1 \leq i \leq 4$) and *Colomba* (parent 2: $5 \leq i \leq 8$); we relate the phenotypes of the F_1 clones to the haplotype probabilities in the following manner (Single-locus QTL model A):

$$Single - locus\ QTL\ model\ A\text{:}$$

$$y = \mu + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_4 X_4 + \alpha_5 X_5 + \alpha_6 X_6 + \alpha_7 X_7 + \alpha_8 X_8 + \varepsilon \quad (12).$$

Given the constraints $\sum_{i=1}^{4} X_i = 2$ and $\sum_{i=5}^{8} X_i = 2$, two terms were eliminated to avoid over-parametrization and co-linearity for regaining independence between the explanatory variables (Single-locus QTL model B), so that:

$$Single - locus\ QTL\ model\ B:$$

$$y = \mu' + \alpha'_2 X_2 + \alpha'_3 X_3 + \alpha'_4 X_4 + \alpha'_6 X_6 + \alpha'_7 X_7 + \alpha'_8 X_8 + \varepsilon \quad (13).$$

For the QTL analysis a genome-wide significance threshold was determined by permutation testing on the phenotypic values with N = 1000 cycles and α = 0.05 (Churchill and Doerge 1994). The QTL analysis was performed over a sliding window at 1-unit cM intervals. The minimum Schwarz information criterion (Schwarz 1978), also known as the Bayesian information criterion (BIC), was used to explore the most likely to be bi-allelic QTL model (of all possible combinations) and the origin of its effect – both additive and dominant – in a similar manner as described by Hackett et al. (2014).

The phenotypic difference between the average homologue effect ($\bar{h}$) on the overall mean trait value of the population ($\bar{y}$) was estimated simultaneously for all homologues of chromosomes containing a QTL as: $\bar{h} - \bar{y}$, by using the following formula:

$$\bar{h} = \frac{\sum_{i=1}^{N} \pi_i y_i}{\sum_{i=1}^{N} \pi_i} \quad (14).$$

The terms in the model correspond to the IBD probabilities (π_i), the trait values of the F_1 clones (y) and the homologues from *Altus* (parent 1: $1 \leq i \leq 4$) and *Colomba* (parent 2: $5 \leq i \leq 8$). The obtained results were plotted to visualize the effects of the homologues. The genotypic information coefficient (GIC) was computed (Bourke et al. 2018a) to provide insight into the information density to infer QTL effects across the mapped genome.

Cofactor QTL analysis

A cofactor QTL analysis that allows for multiple QTLs as cofactors (Jansen and Stam 1994) was carried out in a stepwise procedure. First a naive single-locus QTL analysis (model B) was performed on the protein content values of BLUEs 2012-2014. The positions of significant QTL peaks derived from the naive QTL analysis were defined as cofactors in a subsequent analysis. The residuals resulting from this analysis were saved. Subsequently these residuals were analysed once again using the naive single-locus QTL analysis model B. A genome-wide significance threshold was determined by permutation testing on these residuals with N = 1000 cycles and *alpha* = 0.05 (Churchill and Doerge 1994).

Computation of LOD scores

The LOD scores for all the QTL analyses were computed (Broman et al. 2003), so that:

$$LOD = \frac{N}{2}\, log_{10}(\frac{RSS_{0\ model}}{RSS_{1\ model}}) \quad (15),$$

where $RSS_{0\ model}$ is the residual sum of squares for the null model (no QTL) and $RSS_{1\ model}$ is the residual sum of squares from the fit of the full model.

Results

Integrated linkage map construction

The integrated linkage map (**Table 1**) consisted of 23,328 high quality segregating SNPs. For *Altus* the parental-specific SNPs included simplex × nulliplex and duplex × nulliplex and for *Colomba* these SNPs included nulliplex × simplex and nulliplex × duplex. Herewith *Altus* contributed 10,360 SNPs and *Colomba* 7,998 SNPs. The mapped SNPs were not distributed equally over the chromosomes and homologues of each chromosome. The number of SNP markers ranged between 1,204 and 2,916 between the chromosomes. Also the density of SNPs across homologues of the same chromosomes were variable (**Table 1**). The genetic location (cM) and physical position (Mbp) of the SNPs mapped as expected and contained only a few outliers that deviated from the expected patterns from the genetic versus physical positions of the markers (**Fig. 1**). The centromeric regions of the potato chromosomes (Sharma et al. 2013) – characterized by the absence of recombination – could be clearly identified as visible horizontal stretches in the figure.

Table 1 Linkage map summary for the parental varieties *Altus* and *Colomba*

				Altus‡				*Colomba*‡					
Chr.	*Total*[Δ]	*Altus*†	*Colomba*†	*h1*	*h2*	*h3*	*h4*	*h1*	*h2*	*h3*	*h4*	*Length of integrated chr. linkage map (cM)*	*Length of physical chr. map of PGSC v4.03 (Mb)*
1	*2780*	*981*	*982*	*384*	*519*	*213*	*232*	*344*	*285*	*179*	*397*	*118.8*	*88.7*
2	*2604*	*1410*	*705*	*331*	*650*	*556*	*100*	*237*	*325*	*237*	*231*	*88.0*	*48.6*
3	*2916*	*1706*	*775*	*389*	*788*	*254*	*239*	*181*	*216*	*118*	*331*	*103.0*	*62.3*
4	*1985*	*905*	*631*	*121*	*293*	*228*	*344*	*200*	*61*	*391*	*200*	*107.5*	*72.2*
5	*1743*	*751*	*705*	*288*	*257*	*173*	*115*	*164*	*208*	*301*	*77*	*85.8*	*52.1*
6	*1991*	*1059*	*538*	*365*	*458*	*200*	*179*	*228*	*144*	*130*	*148*	*86.6*	*59.5*
7	*1814*	*807*	*600*	*310*	*192*	*209*	*117*	*123*	*167*	*247*	*227*	*81.0*	*56.8*
8	*1675*	*616*	*723*	*231*	*100*	*222*	*214*	*190*	*103*	*337*	*207*	*88.1*	*56.9*
9	*1778*	*733*	*639*	*335*	*275*	*165*	*94*	*183*	*156*	*317*	*124*	*95.2*	*61.5*
10	*1204*	*355*	*499*	*197*	*77*	*126*	*81*	*183*	*146*	*124*	*158*	*87.3*	*59.8*
11	*1454*	*574*	*608*	*217*	*130*	*119*	*143*	*171*	*278*	*98*	*151*	*77.6*	*45.5*
12	*1384*	*463*	*593*	*115*	*80*	*194*	*209*	*187*	*117*	*185*	*103*	*84.8*	*61.2*
Total	*23328*	*10360*	*7998*	*3283*	*3819*	*2659*	*2067*	*2391*	*2206*	*2993*	*2354*	*1103.7*	*725.1*

Table S1 provides detailed information of the phased SNP markers. Chr. = chromosome. h1 to h4 = homologue 1 to homologue 4. Δ Number of mapped markers on the integrated chromosomal linkage maps. † Markers with segregating alleles from one parent only which include simplex × nulliplex and duplex × nulliplex combinations in the parent under consideration. ǂ Number of homologue-specific markers in simplex condition in the parent under consideration which includes alleles from simplex × nulliplex, simplex × simplex, simplex × duplex and simplex × triplex combinations in one parent and vice-versa for the other parent.

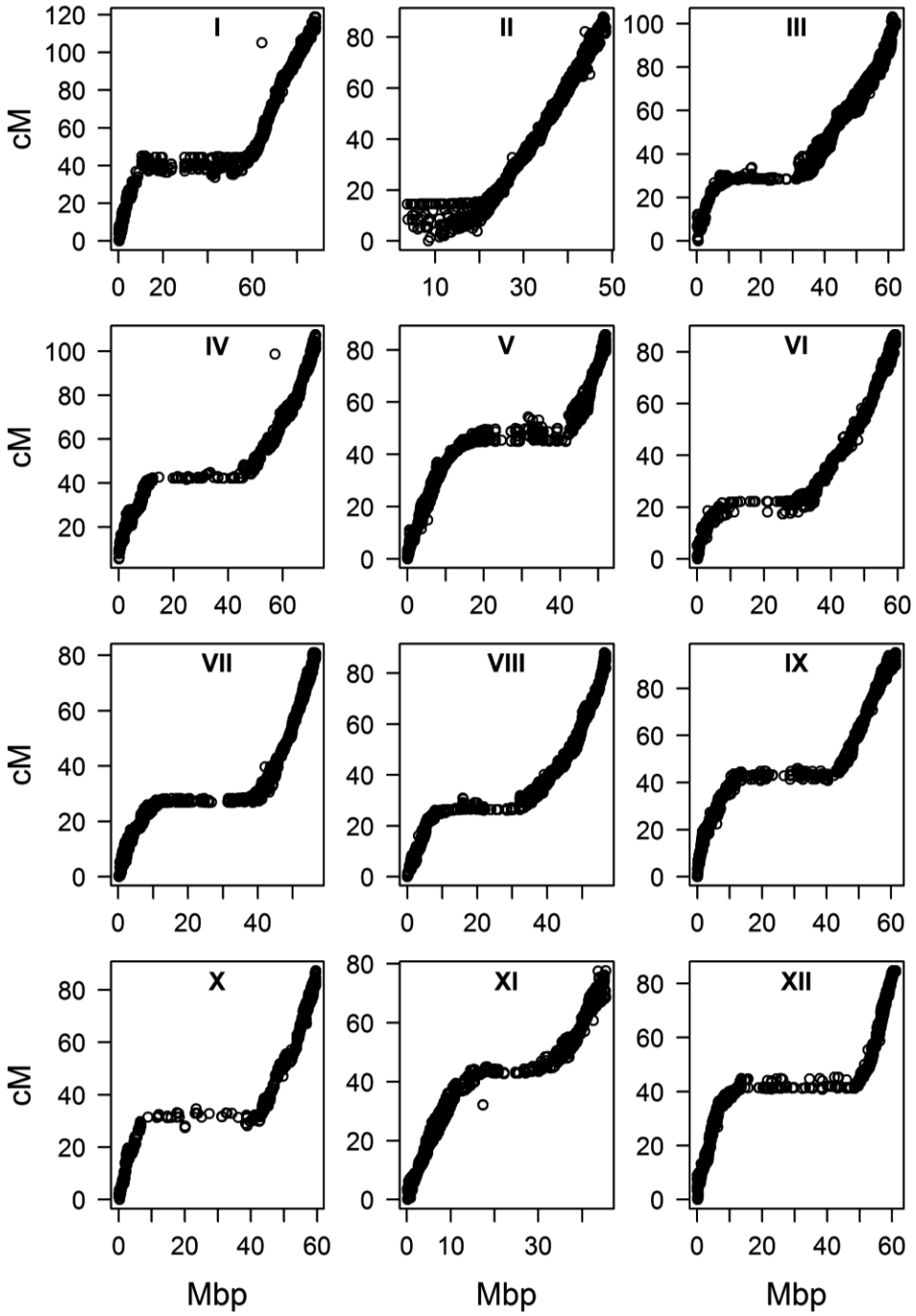

Fig. 1 Plots of the genetic location (cM) vs. physical position (Mbp) of SNPs across the chromosomes**.** The twelve chromosomes of potato are shown in the boxes. The horizontal stretches in the plotted data represent the centromeric regions on the chromosomes.

Phenotypic protein content values

We did not observe spatial trends for the phenotypic values in the field trials. The distribution of the trait valued did not show any (clear) segregation patterns and followed approximately a normal distribution (**Fig. 2**). Compared to the parental varieties, extreme high and low trait values of the F_1 clones were observed in both 2012 and 2013. In contrast, only extreme low trait values were found in 2014.

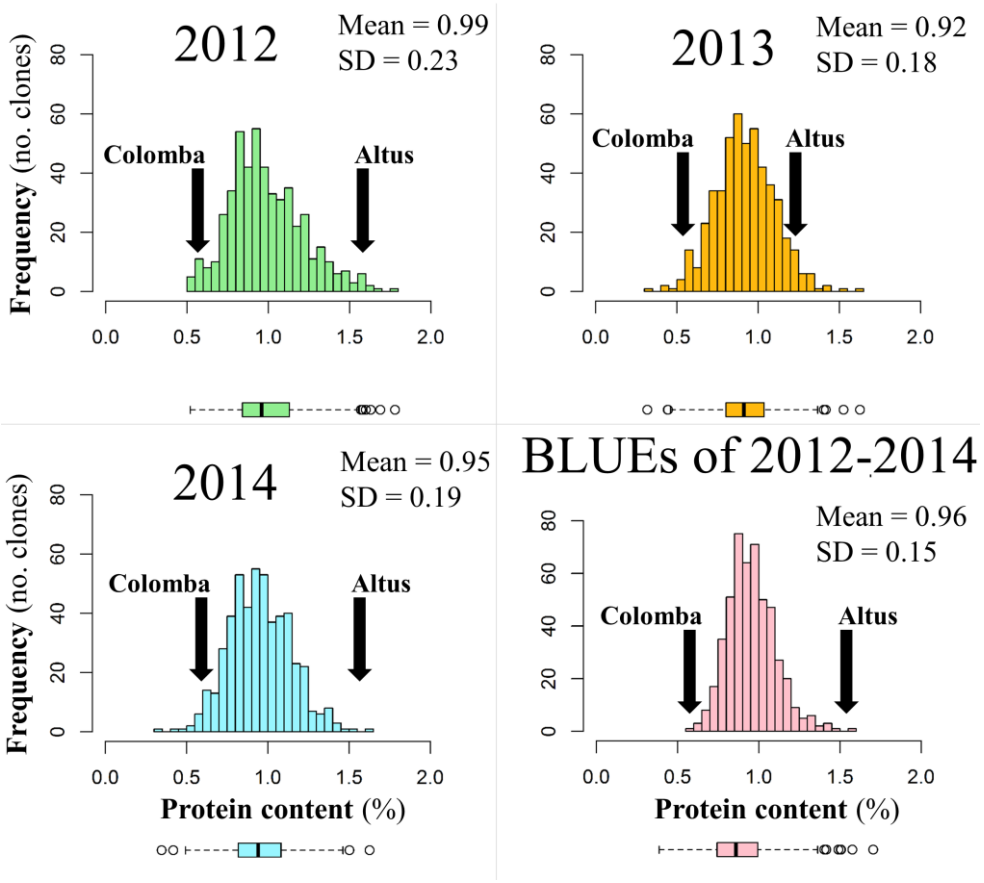

Fig. 2 Distributions of protein content across the years. The green colour stands for year 2012, the orange colour stands for year 2013, the blue colour stands for year 2014 and the pink colour stands for the BLUEs of the years 2012 to 2014. Corresponding boxplots are shown below the histogram figures. The arrows indicate the mean values for the parental varieties *Altus* (parent 1) and *Colomba* (parent 2). SD = standard deviation.

Significant effects of the clones were both within ($P = 5 \times 10^{-7}$ for 2013 and 2014) and between ($P = 5 \times 10^{-7}$ for 2013-2014) the years (**Table 2**). The broad sense heritability estimate was 40% for 2013, 55% for 2014 and 74% over 2013 and 2014. Evidence for clone-by-year interaction was not found ($P = 0.395$ for 2013-2014). The data for 2012 was not included for heritability estimation as the population lacked replication in the field for this year ($N = 1$).

Table 2 Summary statistics of protein content for the years 2013, 2014 and 2013-2014

Parameter	2013	2014	2013-2014
Min.	0.32	0.35	0.58
Mean	0.92	0.95	0.93
Max.	1.63	1.63	1.63
σ^2_G	0.013*	0.021*	0.042*
$\sigma^2_{G\times Y}$	-	-	0.001 *N.S.*
σ^2_ε	0.039	0.034	0.062
H^2	0.40	0.55	0.74

σ^2_G = Genotype/clone variance, $\sigma^2_{G\times Y}$ = Genotype/clone by Year interaction variance, σ^2_ε = Residual variance, H^2 = Broad sense heritability estimate, * statistically significant ($P < 0.001$), *N.S.* = statistically non-significant ($P > 0.05$).

Correlation analyses of phenotypic values between the years revealed low to poor R^2 values that significantly deviated from zero (2012-2013: $R^2 = 0.09$, $P = 8.9 \times 10^{-12}$; 2012-2014: $R^2 = 0.13$, $P = 4.1 \times 10^{-16}$; 2013-2014: $R^2 = 0.09$, $P = 1.8 \times 10^{-11}$) (**Fig. 3**).

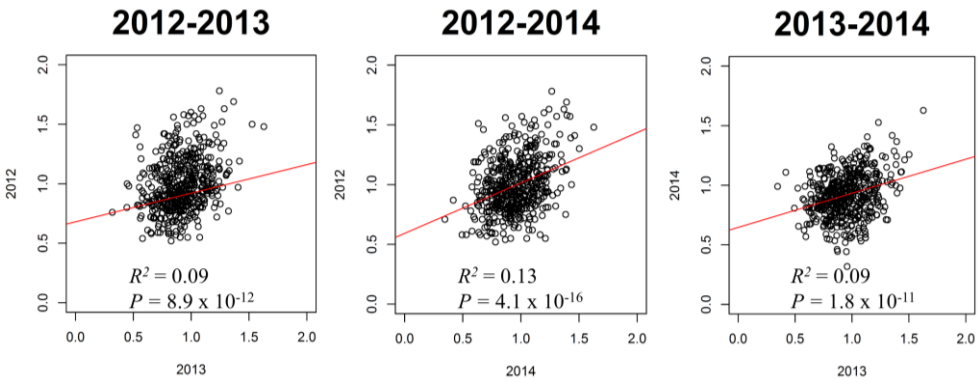

Fig. 3 Plots for the values of protein content between the years 2012-2013, 2012-2014 and 2013-2014. The red line in the figure represents the simple linear regression line.

Naive QTL analysis and variance explained by QTLs

Naive QTL analysis was performed by regression analysis (single-locus QTL model B) – by using the phenotypic values and the IBD probabilistic haplotypes – at all positions on the chromosomes. Separate analyses were conducted for protein content in 2012, 2013, 2014 and the BLUEs of 2012-2014. QTLs were detected on chromosomes *2*, *3*, *5* and *9* (**Fig. 4**) for 2013, 2014 and the BLUEs of 2012-2014 (**Table 3**). No significant QTLs were found for 2012. The largest QTL was found on chromosome *5* for 2014, however in the other years this QTL was not found.

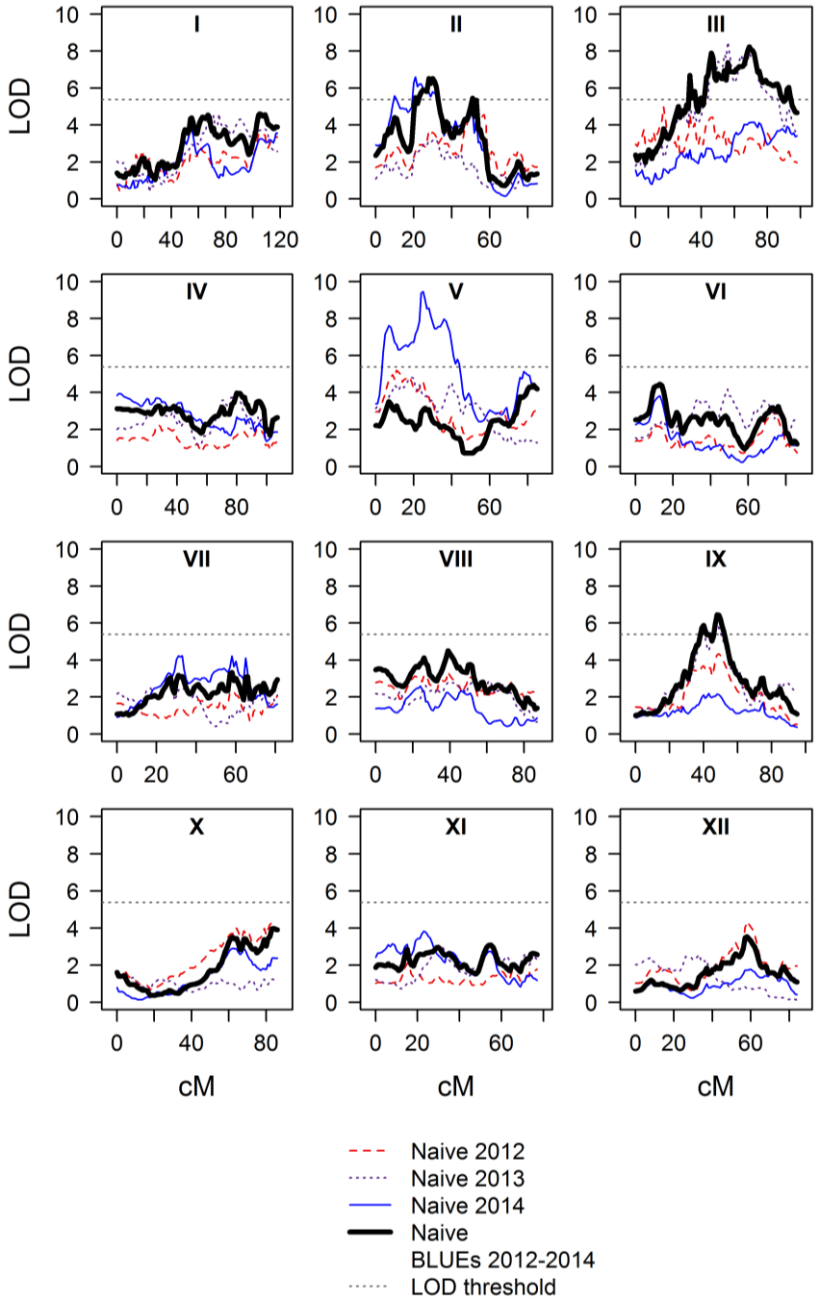

Fig. 4 LOD profiles of the naive single-locus QTL analysis for protein content. The boxes represent the twelve individual chromosomes of potato. The red dashed line represents the year 2012, the purple dotted line represents the year 2013, the blue line represents the year 2014 and the black line in bold represents the BLUEs of the years 2012-2014. The horizontal dashed line represents the permutation-based LOD threshold (LOD = 5.3).

Table 3 Statistics of QTLs for protein content from the naive single-locus QTL analysis

Year	Chr.	LOD peak position and LOD-2 interval (cM)	LOD score	R^2
2013	*3*	56: 44-75	8.4	0.076
2013	*9*	48: 47-53	5.9	0.051
2014	*2*	21: 9-33	6.6	0.059
2014	*5*	25: 22-39	9.5	0.084
BLUEs 2012-2014	*2*	30: 19-54	6.5	0.058
BLUEs 2012-2014	*3*	69: 32-82	8.2	0.074
BLUEs 2012-2014	*9*	48: 34-55	6.5	0.058

Naive = naive QTL analysis without cofactors. Chr. = chromosome. LOD = logarithm of the odds. LOD-2 interval = support interval above the QTL threshold (LOD = 5.3). BLUEs = best linear unbiased estimates.

Numerous underlying haplotypes (or alleles) of the QTLs found on chromosomes *2*, *3*, *5* and *9* provided positive or negative effects on the trait values (**Fig. 5**). For the strongest QTL (chromosome *5* from 2014), the haplotypes originating from *Colomba* (parent 2) on homologues *5* and *6* provided positive effects, whilst homologue 8 provided a negative effect.

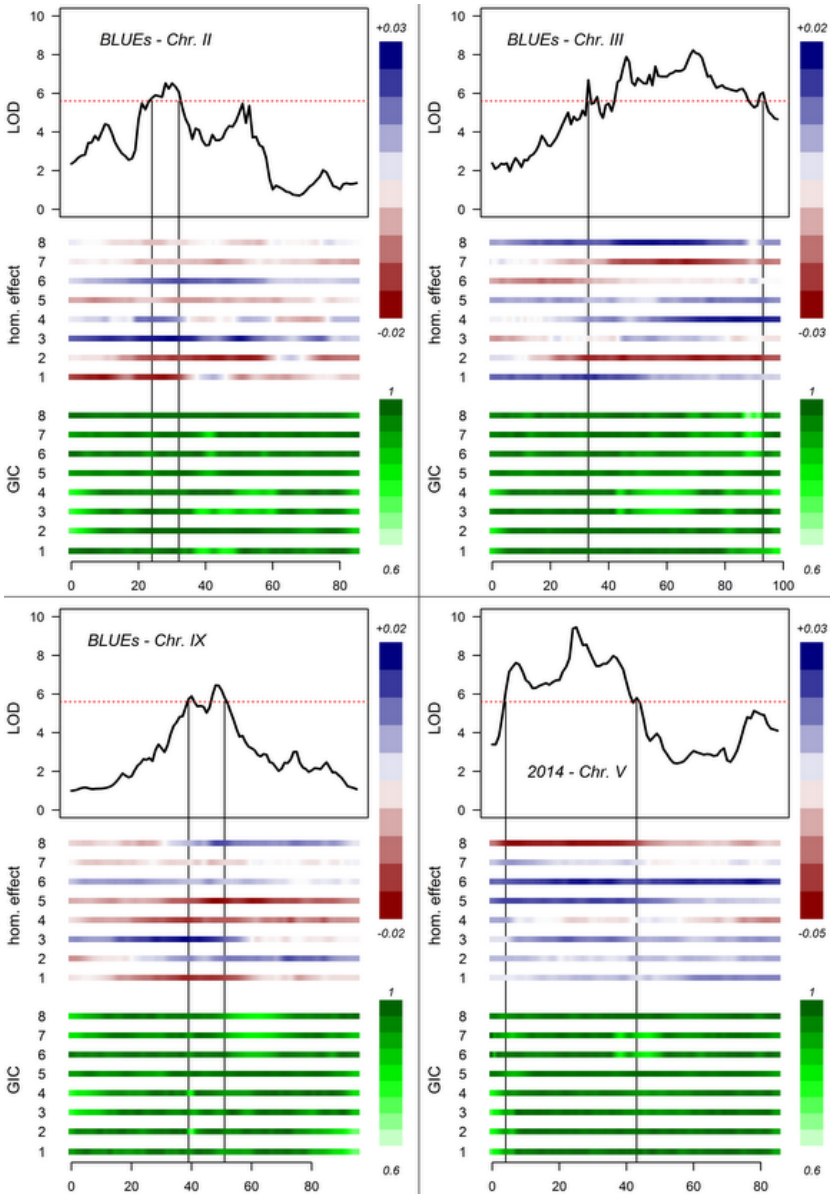

Fig. 5 Contribution of homologues on the values of protein content for the QTLs identified on chromosomes *2* (top left), *3* (top right), *5* (bottom right) and *9* (bottom left). Homologues (hom.) 1 to 4 originate from *Altus* (parent 1) and homologues 5 to 8 from *Colomba* (parent 2). The upper vertical scale in blue to red represents the contribution of the homologues on the values of protein content. The lower vertical scale in dark green to light green represents the genotypic information coefficient (GIC) of the homologues. The horizontal red dashed line in the LOD plot represents the permutation-based LOD threshold (LOD = 5.3). The horizontal scale below the homologue box represents the genetic (cM) positions.

The percentage of phenotypic variation (R^2) explained by individual QTLs ranged between 5.1 to 8.4% (**Table 3**). The largest amount of variation (8.4%) was explained by the QTL on the top arm of chromosome *5* at 25 cM in 2014. The QTLs on chromosome *3* explained 7.6% and 7.4% in 2013 (56 cM) and BLUEs of 2012-2014 (69 cM) respectively. The QTLs on chromosomes *2* and *9* explained between 5.1 and 5.9% of the phenotypic variation. The most probable QTL segregation models of the QTLs in **Table 3** were assessed by means of the minimum Schwarz information criterion (SIC) to assess the strength of evidence of the effects and origins of the parental haplotypes that contribute towards the trait. The differences in the minimum SIC of segregation models with balanced group sizes revealed values that are considered low (less than 2 – data not shown) (Neath and Cavanaugh 2012) and are therefore not elaborated on. However, one exception was found for chromosome *5* in 2014 with a segregating QTL – QQQQ × QQQq – from *Colomba* (parent 2). This model provided a minimum SIC value of 19 that was considered as a probable segregation model for this QTL. This segregation model accounted for a mean protein content value of 0.905 for 'q' versus 0.995 for 'Q' of the segregating allele on homologue four from *Colomba*. This finding was found to be consistent with the negative effect of homologue *8* for this QTL in 2014 from *Colomba* (**Fig. 5**). The variation explained by all QTLs for 2013 (chromosomes *2* and *9*), 2014 (chromosomes *2* and *5*) and BLUEs of 2012-2014 (chromosomes *2*, *3* and *9*) accounted for 11.9%, 12.7% and 17.2% of the total phenotypic trait variation respectively.

QTL cofactor analysis

To identify masked QTLs, cofactor analysis was performed on the protein content values for 2013, 2014 and the BLUEs of 2012-2014 by using the QTLs identified by the naive analysis (**Table 3**) as cofactors (Jansen and Stam 1994). No masked QTLs were revealed by cofactor analysis of the BLUEs 2012-2014 when the QTLs on chromosomes *2*, *3* and *9* were set as cofactors (**Fig. 6**). However, two masked QTLs were identified on chromosomes *1* (**Fig. S4**) and *5* (**Fig. S11**) after other combinations of QTLs were used as cofactors (**Table 4**).

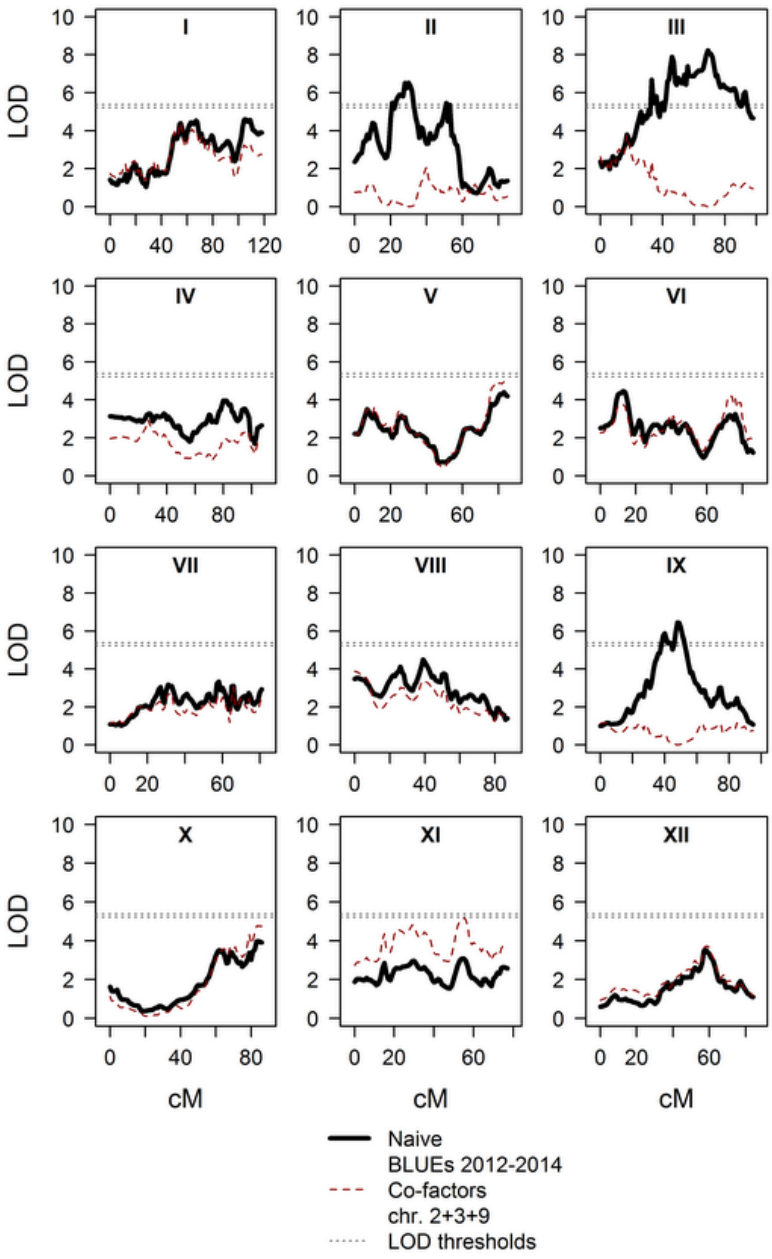

Fig. 6 LOD profiles for protein content of both the naive and cofactor QTL analysis for the BLUEs of the years 2012-2014. The black line in bold represents the LOD scores of the naive QTL analysis and the red dashed line represents the LOD scores of the cofactor QTL analysis with three peak QTLs set as cofactors from chromosomes *2*, *3* and *9* (**Table 3**). The top horizontal dashed line shows the permutation-based LOD threshold for the naive QTL analysis (LOD = 5.3) and the lower horizontal dashed line shows the permutation-based LOD threshold for the cofactor QTL analysis (LOD = 5.2).

Table 4 Statistics of QTLs for protein content from the cofactor analysis

Year	Naive QTL(s) as cofactor(s): chr. (position)[a]	LOD scores of QTLs by cofactor analysis (chr. & position)[b]	R^2 of QTLs by cofactor analysis[c]	New QTLs identified by cofactor analysis: chr. (position)[d]
2013	*3* (56)	5.7 (*9*-40) **Fig. S1**	0.051	No
2013	*9* (48)	7.8 (*3*-56) **Fig. S2**	0.07	No
2013	*3* (56) + *9* (48)	*N.S.* **Fig. S3**	*N.S.*	No
2014	*2* (21)	9.3 (*5*-24); 5.5 (*5*-78) **Fig. S4**	0.083; 0.050	*5* (78)
2014	*5* (25)	6.2 (*2*-23) **Fig. S5**	0.056	No
2014	*2* (21) + *5* (25)	*N.S.* **Fig. S6**	*N.S.*	No
BLUEs 2012-2014	*2* (30)	5.4 (*1*-67); 9.1 (*3*-69); 6.3 (*9*-40) **Fig. S7**	0.049; 0.081; 0.056	*1* (67)
	3 (69)	6.8 (*2*-30); 5.7 (*9*-40) **Fig. S8**	0.062; 0.052	No
	9 (48)	6.2 (*2*-30); 7.4 (*3*-69) **Fig. S9**	0.056; 0.067	No
	2 (30) + *3* (69)	6.1 (*9*-40) **Fig. S10**	0.059	No
	2 (30) + *9* (48)	8.2 (*3*-69); 5.9 (*5*-83) **Fig. S11**	0.074; 0.053	*5* (83)
	3 (69) + *9* (48)	6.7 (*2*-28) **Fig. S12**	0.06	No
	2 (30) + *3* (69) + *9* (48)	*N.S.* **Fig. 6**	*N.S.*	No

[a] QTL(s) identified by the naive QTL analysis (**Table 3**). [b] QTL(s) identified by the cofactor QTL analysis after one or more QTL(s) identified by the naive QTL analysis were used as cofactors. [c] Variance explained (*R-sq.*) by QTL(s) identified by the cofactor QTL analysis. [d] New QTLs identified by the cofactor QTL analysis that were not identified by the naive QTL analysis. Chr. = chromosome. Position = QTL peak position in centiMorgan (cM). LOD stands for logarithm of the odd. BLUEs = best linear unbiased estimates. *N.S.* = statistically non-significant.

Discussion

The development of elite potato varieties with high levels of protein content is an innovative topic of great economic and environmental relevance for the potato starch industry. Hence protein content in potato has emerged as a novel breeding goal. Shedding light on the genetic architecture underlying this trait and identifying QTLs are the first necessary steps for defining strategies that may enable (marker-assisted) breeding of elite varieties with high protein content. In this study, we evaluated the genetics of protein content for a large bi-parental mapping population of 496 full-sib F_1 clones. These clones originated from a cross between two genetically divergent cultivated tetraploid potato varieties with a high and low level of protein content. An integrated chromosomal linkage map was constructed and was used to compute probabilistic haplotypes across all chromosomes that were used for QTL analysis. Potential QTLs were found on five chromosomes and the formation of extreme trait values – *i.e.* transgressive segregation – was observed in all three years. The occurrence of the extreme trait values in the progeny may be caused by transgressive segregation due to complementary action of additive alleles contributed by different parental varieties. We report broad sense heritability estimates of 40% (2013) and 55% (2014) and 74% (2013-2014), indicating that a moderate proportion of the trait variance can be ascribed to genetic factors. Therefore it can be postulated that breeding for this complex quantitative trait is theoretically possible.

Chromosomal linkage map and probabilistic haplotypes

The 60K SNP marker array provided a wealth of information for the construction of the integrated chromosomal linkage map that was subsequently transformed into multi-locus probabilistic haplotypes. Only well-performing polymorphic SNPs were used for genotype calling after the removal of monomorphic SNPs and those with too many missing values. This relatively strict procedure resulted in a smaller sub-set of high-quality SNPs – 23,328 in total – that were used for linkage mapping and subsequent QTL analysis. Preferably, a smaller set of high-quality SNPs is to be used for linkage mapping instead of a larger set containing potentially erroneous markers (Preedy and Hackett 2016). This was observed in the process of marker ordering for the linkage maps. After three rounds of MDSmap, merely 0.5% (116 SNPs) from the total sub-set of 23,328 SNPs were removed as possible outliers. Our final integrated chromosomal linkage map corresponds well to the physical reference map of potato (PGSC pseudomolecules v4.03). No structural differences – such as noticeable translocations, inversions, insertions or deletions – between our map and the

reference genome were observed. The total genetic distance of the map here (1104 cM) is comparable to the length reported in genetic maps of tetraploid potato (1042 to 1088 cM) using fewer SNPs and populations of smaller sizes (Hackett et al. 2013; Massa et al. 2015; Rak et al. 2017). To the best of our knowledge, we present here the most marker-dense potato genetic linkage map and includes SNPs of all possible allele dosage types that segregate from both parents.

TetraOrigin (Zheng et al. 2016) was used to estimate multi-locus probabilistic haplotypes across all chromosomes with high levels of genotype information content across all chromosomes (data not shown). The application of probabilistic haplotypes in genetic association studies in diploids allows for higher statistical power than single-marker procedures, as has been shown in human studies (de Bakker et al. 2005). In tetraploids this application is also expected to generate an equal or higher level of true statistical power than single-marker procedures as information of all 8 homologues is used (Bourke 2014).

To avoid the identification of false-positive QTLs, a stringent genome-wide LOD threshold was estimated from 1000 permutation cycles. The application of this stringent threshold for QTL detection may result in less power to detect minor QTLs, possibly leading to a higher type II error rate (*i.e.* false negatives). However, results from our co-factor QTL analysis did not reveal any other masked QTLs when compared to the results derived from the naive QTL analysis. The high marker-density of the integrated chromosomal genetic map and final definition of the multi-locus probabilistic haplotypes, in combination with a large mapping population, provided a framework for conducting a reliable QTL analyses with great power. This presumption was found to be true after QTL peaks were mapped only 120 kb away from the well-known potato maturity locus (Cycling DNA-binding with one finger Factor: *StCDF1*) at the top of chromosome *5* (data not shown).

Size of the mapping population

Genetic studies on protein content in other crops such as soybean, wheat and maize illustrate that this trait is quantitative and that it is regulated by multiple genes that are likely to be influenced by genotype-by-environment interactions (Balyan et al. 2013; Hwang et al. 2014; Karn et al. 2017). Therefore, the use of a relatively large mapping population to study this quantitative trait has strongly contributed towards detecting potential QTLs. Moreover, QTL effects are expected to be estimated more accurately in this large population. The use of smaller populations may cause false

inflation of QTL effects (Hackett et al. 2014; Vales et al. 2005). The presence of potential epistatic interactions of the QTLs was not analysed in this study. Knowledge on epistasis is relevant to assess whether and which combinations of haplotypes contribute positively or negatively towards the trait. The availability of novel statistical frameworks and computation power are needed to answer these questions for tetraploid potato.

QTLs for protein content

The factors underlying the genetic architecture of protein content in potato are poorly understood. This study provides insight into the first QTLs for protein content in cultivated tetraploid potato. By using a large bi-parental (*Altus* × *Colomba*) mapping population consisting of 496 tetraploid F_1 clones, we detected potential naive QTLs on chromosomes *2*, *3*, *5* and *9*, each explaining between 5.1 and 8.4% of the trait variance. The variance explained by these QTLs together (11.9 to 17.2%) does not reflect the trait heritability estimates (40 to 74%) that express the amount of variance that can be ascribed to genetic factors. This gap may be caused by factors that include the lack of power to detect minor effect QTLs (by naive QTL analysis), epistatic interactions and genotype-by-environment interactions. QTL cofactor analysis in tetraploid potato may enable the identification of masked QTL for complex traits such as protein content as demonstrated here by cofactor QTL analysis that identified masked QTLs on chromosomes *1* (**Fig. S7**) and *5* on the lower arm (**Fig. S4**; **Fig. S11**). Another explanation for this gap may lie in the heritability estimates themselves. These estimates may be overestimated. Moreover, they do not discriminate between the parts of the trait variance that are heritable and those that are environmental. In the case of overestimation, thus leading toward phantom heritability, the variance explained solely by QTLs may reflect the biology of the trait more accurately.

Previous QTL studies on protein content in non-cultivated diploid potato (Acharjee et al. 2018; Werij 2011) reported QTLs on chromosomes *1*, *3* and *5* (top arm). Whether the haplotypes that account for the QTLs on chromosomes *1*, *3* and *5* in this study are identical to those found in diploid studies requires further research. The QTLs detected in this study on chromosomes *2*, *5* (lower arm) and *9* are novel as they have not been reported in literature before and are presented here for the first time.

The strongest QTL here was found in 2014, a year characterized by ample precipitation at the start of the potato growing season. This QTL was detected at the top arm of chromosome *5*, that harbours the major regulator of maturity and initiation of tuber formation (*StCDF1*), as well as a cluster of nitrate transporter genes. In potato, *StCDF1* has been shown to cause pleiotropic effects on multiple sub-traits, including foliar senescence, plant cycle length, the onset of tuberization and potato tuber yield (Hurtado et al. 2012; Kloosterman et al. 2013). In soybean, it has been demonstrated that the plant cycle length and ambient temperature affects protein content during seed development (Patil et al. 2017). In case of *StCDF1*, this QTL may have caused developmental instability that overshadowed other potential QTLs that were found in other years (*e.g.* QTLs on chromosomes *3* and *9*). Nitrate transporters may also influence the level of protein content. In rice it has been demonstrated that over-expression of a nitrate transporter increased the yield and nitrogen-use efficiency by 40% (Fan et al. 2016). The QTL on chromosome *3* overlaps with regions that harbour gene clusters of tuber proteins that include Kunitz-type protease inhibitors, potato protease inhibitor I (PI-1) and potato protease inhibitor II (PI-2). Protease inhibitors have been suggested to function as storage proteins (Pusztai 1972) as well as potential regulators of proteolysis by means of their inhibition activity against proteases (*e.g.* trypsin, α-chymotrypsin and elastase). The QTL on chromosome *9* also co-localizes with gene clusters of potato PI-1 and potato PI-2. We found no QTLs in 2012. This may be related to the physiological state of the propagated seed tubers that were used as starting material in the field trial of 2012. Seed tubers should be propagated at least one cycle to carry out more reliable field trials as the physiological state of seed tubers may have a strong effect on plant development (Asiedu et al. 2003).

In this three-year study, a limited amount of overlap was found between QTLs over the years. This phenomenon is reflected by the low to poor correlations of the trait values between the years. QTLs on chromosome *2* (2014), *3* (2013) and *9* (2013) overlapped with the QTLs of the BLUEs of 2012-2014. The QTL on chromosome *5* was identified only in 2014. The low reproducibility of these QTLs may (partly) be attributed to the experimental design. The population was grown at three different locations over three years with one or two biological replicates. When dealing with a quantitative trait with a moderate level of heritability, such as protein content, the reproducibility of QTLs may be improved by making use of more biological replicates to compensate for possible environmental effects such as heterogeneity of

soil quality and possible differences in (nitrogen) fertilizer residues in the plots of the trial fields.

Strategies for trait improvement

This study illustrates that the genetic architecture of protein content in tetraploid potato is quantitative and complex. The moderate level of trait heritability in this study indicates that a substantial proportion of the trait variance can potentially be ascribed to heritable factors (QTLs). In the *Altus* × *Colomba* population we estimated a moderate level of broad sense trait heritability between 40 and 74% and identified potential naive QTLs on chromosomes *2*, *3*, *5* and *9*.

The cumulative variance explained by the naive QTLs identified within 2013, 2014 and BLUEs of 2012-2014 accounted between 11.9 to 17.2% of the total phenotypic variation. These proportions of variance closely resemble the cumulative variance explained by the QTLs together. Thus it can be postulated that these QTLs exert additive effects on protein content. Therefore, protein content in potato can be improved by fixating alleles with positive effects that underlie these QTLs in gene pools for breeding. However, in this specific mapping population it is evident that a large part of the trait heritability is not explained by the identified genetic factors alone. Further research is needed to elucidate these factors (*e.g.* minor effects QTLs and genotype-by-environment interactions) that may contribute towards this phenomenon.

To improve protein content in potato by means of molecular breeding, a more complete and comprehensive understanding of the genetic architecture and regulation of the trait is needed. A full overview of all protein content related QTLs – and their potential pleiotropic effects – that are present in relevant gene pools (*e.g.* the starch potato genepool) and uncultivated material provide a better understanding of the trait that is necessary for conscious decision-making in breeding programs. Further genetic studies, that include genome-wide association studies (GWAS) of variety panels and additional QTL analyses – both bi-parental and di-allel – are necessary for generating further insights into this economically and environmentally relevant trait. To take potential pleiotropic effects into account, these panels and populations should preferable include a similar maturity index, a uniform onset of tuber formation, fresh tuber yield, tuber under-water weight and overall yield stability across different environmental conditions. An alternative and classical strategy for trait improvement is the long term-selection on a trait of interest. Long-

term selection programs have shown that this approach can be highly effective for increasing protein content in maize (Dudley 2007). This selection strategy allows the stacking and fixation of alleles with positive effects in the gene-pool for the trait of interest. For potato, this approach should without doubt also include the selection of other important potato traits such as pathogen resistances (*e.g.* late blight and potato cyst nematodes) and quality parameters (*e.g.* glycoalkaloid content). However, potential trade-offs should also be considered in the development of protein-rich potato varieties. This may include the use of heavier inputs of (mineral) nitrogen fertilizers that for a large part may be lost due to run-off and leaching into groundwater, thus causing environmental pollution. Therefore the question remains whether the nitrogen-use efficiency of protein-rich potato varieties will be improved, especially in the light of increasing attention for environmentally sustainable agriculture.

Acknowledgements

The authors would like to thank *Averis Seeds* for providing the phenotypic and raw genotypic data. Konrad Zych from the *University of Groningen* is acknowledged for distributing the pre-processed genotypic data. We would also like to acknowledge Katherine Preedy and Christine Hackett of *Biomathematics & Statistics Scotland* (BioSS) at *The James Hutton Institute* for making a developmental version of the MDSmap software available.

Supplementary data are available online at:

https://link.springer.com/article/10.1007/s10681-018-2331-z

References

Acharjee A, Chibon P-Y, Kloosterman B, America T, Renaut J, Maliepaard C, Visser RG (2018) Genetical genomics of quality related traits in potato tubers using proteomics. BMC Plant Biology 18:20

Asiedu S, Astatkie T, Yiridoe E (2003) The Effect of Seed-Tuber Physiological Age and Cultivar on Early Potato Production. Journal of Agronomy and Crop Science 189:176-184

Balyan HS, Gupta PK, Kumar S, Dhariwal R, Jaiswal V, Tyagi S, Agarwal P, Gahlaut V, Kumari S (2013) Genetic improvement of grain protein content and other health-related constituents of wheat grain. Plant Breeding 132:446-457

Bárta J, Bártová V, Zdráhal Zk, Šedo O (2012) Cultivar Variability of Patatin Biochemical Characteristics: Table versus Processing Potatoes (Solanum tuberosum L.). Journal of Agricultural and Food Chemistry 60:4369-4378

Bártová V, Bárta J (2009) Chemical composition and nutritional value of protein concentrates isolated from potato (Solanum tuberosum L.) fruit juice by precipitation with ethanol or ferric chloride. Journal of Agricultural and Food Chemistry 57:9028-9034

Björn B, Paulo MJ, Visser RG, van Eck HJ, van Eeuwijk FA (2011) Phenotypic analyses of multi-environment data for two diverse tetraploid potato collections: comparing an academic panel with an industrial panel. Potato Research 54:157-181

Bourke PM (2014) QTL analysis in polyploids. MSc thesis. Chair group: Plant Breeding. Wageningen University, Wageningen, The Netherlands. Retrieved from http://edepot.wur.nl/310159

Bourke PM (2018) Genetic mapping in polyploids. PhD dissertation. Chair group: Plant Breeding. Wageningen University, Wageningen, The Netherlands. Retrieved from http://edepot.wur.nl/444415

Bourke PM, Gitonga VW, Voorrips RE, Visser RG, Krens FA, Maliepaard C (2018a) Multi-environment QTL analysis of plant and flower morphological traits in tetraploid rose. Theoretical and Applied Genetics 131:2055-2069

Bourke PM, van Geest G, Voorrips RE, Jansen J, Kranenburg T, Shahin A, Visser RGF, Arens P, Smulders MJM, Maliepaard C (2018b) polymapR—linkage analysis and genetic map construction from F1 populations of outcrossing polyploids. Bioinformatics 34:3496-3502

Bourke PM, Voorrips RE, Kranenburg T, Jansen J, Visser RG, Maliepaard C (2016) Integrating haplotype-specific linkage maps in tetraploid species using SNP markers. Theoretical and Applied Genetics 129:2211-2226

Bourke PM, Voorrips RE, Visser RG, Maliepaard C (2015) The Double-Reduction Landscape in Tetraploid Potato as Revealed by a High-Density Linkage Map. Genetics 201:853-863

Bradshaw JE, Hackett CA, Pande B, Waugh R, Bryan GJ (2008) QTL mapping of yield, agronomic and quality traits in tetraploid potato (Solanum tuberosum subsp. tuberosum). Theoretical and Applied Genetics 116:193-211

Broman KW, Wu H, Sen Ś, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. Bioinformatics 19:889-890

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics 138:963-971

Creusot N, Wierenga PA, Laus MC, Giuseppin ML, Gruppen H (2011) Rheological properties of patatin gels compared with β-lactoglobulin, ovalbumin, and glycinin. Journal of the Science of Food and Agriculture 91:253-261

de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D (2005) Efficiency and power in genetic association studies. Nature Genetics 37:1217-1223

Dudley J (2007) From means to QTL: The Illinois long-term selection experiment as a case study in quantitative genetics. Crop Science 47:S-20-S-31

Fan X, Tang Z, Tan Y, Zhang Y, Luo B, Yang M, Lian X, Shen Q, Miller AJ, Xu G (2016) Overexpression of a pH-sensitive nitrate transporter in rice increases crop yields. Proceedings of the National Academy of Sciences 113:7118-7123

FAO (2014) Available at http://faostat.fao.org. Accessed 19 August 2014

Hackett CA, Bradshaw JE, Bryan GJ (2014) QTL mapping in autotetraploids using SNP dosage information. Theoretical and Applied Genetics 127:1885-1904

Hackett CA, McLean K, Bryan GJ (2013) Linkage analysis and QTL mapping using SNP dosage data in a tetraploid potato mapping population. PloS One 8:e63939

Hurtado PX, Schnabel SK, Zaban A, Veteläinen M, Virtanen E, Eilers PH, Van Eeuwijk FA, Visser RG, Maliepaard C (2012) Dynamics of senescence-related QTLs in potato. Euphytica 183:289-302

Hwang E-Y, Song Q, Jia G, Specht JE, Hyten DL, Costa J, Cregan PB (2014) A genome-wide association study of seed protein and oil content in soybean. BMC Genomics 15:1

Jansen RC, Stam P (1994) High resolution of quantitative traits into multiple loci via interval mapping. Genetics 136:1447-1455

Jørgensen M, Bauw G, Welinder KG (2006) Molecular Properties and Activities of Tuber Proteins from Starch Potato Cv. Kuras. Journal of Agricultural and Food Chemistry 54:9389-9397

Karn A, Gillman JD, Flint-Garcia SA (2017) Genetic analysis of teosinte alleles for kernel composition traits in maize. G3: Genes | Genomes | Genetics 7:1157-1164

Kloosterman B, Abelenda JA, Gomez MdMC, Oortwijn M, de Boer JM, Kowitwanich K, Horvath BM, van Eck HJ, Smaczniak C, Prat S (2013) Naturally occurring allele diversity allows potato cultivation in northern latitudes. Nature 495:246

Kudo K, Onodera S, Takeda Y, Benkeblia N, Shiomi N (2009) Antioxidative activities of some peptides isolated from hydrolyzed potato protein extract. Journal of Functional Foods 1:170-176

Lu W, Yu M, Bai Y, Li W, Xu X (2012) Crude protein content in diploid hybrid potato clones of Solanum phureja–S. stenotomum. Potato Research 55:315-322

Massa AN, Manrique-Carpintero NC, Coombs JJ, Zarka DG, Boone AE, Kirk WW, Hackett CA, Bryan GJ, Douches DS (2015) Genetic linkage mapping of economically important traits in cultivated tetraploid potato (Solanum tuberosum L.). G3: Genes | Genomes | Genetics 5:2357-2364

Neath AA, Cavanaugh JE (2012) The Bayesian information criterion: background, derivation, and applications. Wiley Interdisciplinary Reviews: Computational Statistics 4:199-203

Ortiz-Medina E (2006) Potato tuber protein and its manipulation by chimeral disassembly using specific tissue explantation for somatic embryogenesis (PhD dissertation). Department of Plant Science. McGill University, Montreal, Canada

Patil G, Mian R, Vuong T, Pantalone V, Song Q, Chen P, Shannon GJ, Carter TC, Nguyen HT (2017) Molecular mapping and genomics of soybean seed protein: a review and perspective for the future. Theoretical and Applied Genetics 130:1975-1991

PGSC (2011) Genome sequence and analysis of the tuber crop potato. Nature 475:189-195

Pots AM, Gruppen H, van Diepenbeek R, van der Lee JJ, van Boekel MAJS, Wijngaards G, Voragen AGJ (1999) The effect of storage of whole potatoes of three cultivars on the patatin and protease inhibitor content; a study using capillary electrophoresis and MALDI-TOF mass spectrometry. Journal of the Science of Food and Agriculture 79:1557-1564

Preedy K, Hackett C (2016) A rapid marker ordering approach for high-density genetic linkage maps in experimental autotetraploid populations using multidimensional scaling. Theoretical and Applied Genetics 129:2117-2132

Pusztai A (1972) Metabolism of trypsin-inhibitory proteins in the germinating seeds of kidney bean (Phaseolus vulgaris). Planta 107:121-129

Rak K, Bethke PC, Palta JP (2017) QTL mapping of potato chip color and tuber traits within an autotetraploid family. Molecular Breeding 37:15

Schwarz G (1978) Estimating the dimension of a model. The Annals of Atatistics 6:461-464

Sharma SK, Bolser D, de Boer J, Sønderkær M, Amoros W, Carboni MF, D'Ambrosio JM, de la Cruz G, Di Genova A, Douches DS (2013) Construction of reference chromosome-scale pseudomolecules for potato: integrating the potato genome with genetic and physical maps. G3: Genes| Genomes| Genetics 3:2031-2047

Sverrisdóttir E, Byrne S, Sundmark EHR, Johnsen HØ, Kirk HG, Asp T, Janss L, Nielsen KL (2017) Genomic prediction of starch content and chipping quality in tetraploid potato using genotyping-by-sequencing. Theoretical and Applied Genetics:1-18

Vales M, Schön C, Capettini F, Chen X, Corey A, Mather D, Mundt C, Richardson K, Sandoval-Islas J, Utz H (2005) Effect of population size on the estimation of QTL: a test using resistance to barley stripe rust. Theoretical and Applied Genetics 111:1260-1270

Voorrips RE, Gort G, Vosman B (2011) Genotype calling in tetraploid species from bi-allelic marker data using mixture models. BMC Bioinformatics 12:172

Vos PG, Uitdewilligen JG, Voorrips RE, Visser RG, van Eck HJ (2015) Development and analysis of a 20K SNP array for potato (Solanum tuberosum): an insight into the breeding history. Theoretical and Applied Genetics 128:2387-2401

Werij JS (2011) Genetic analysis of potato tuber quality traits. PhD dissertation. Chair Group: Plant Breeding. Wageningen University, Wageningen, The Netherlands

Zheng C, Voorrips RE, Jansen J, Hackett CA, Ho J, Bink MC (2016) Probabilistic Multilocus Haplotype Reconstruction in Outcrossing Tetraploids. Genetics 203:119-131

Chapter 4

Overexpression of a putative nitrate transporter (*StNPF1.11*) increases plant height, leaf chlorophyll content and tuber protein content of young potato plants

Michiel T. Klaassen[1,2], Dianka C.T. Dees[1], Rommel M. Garrido Jr.[1], Jorge Alemán Báez[1], Michiel Schrijen[1], Pablo G. Baldeón Mendoza[1], Luisa M. Trindade[1]

[1] Wageningen University & Research, Plant Breeding, Wageningen, The Netherlands

[2] Aeres University of Applied Sciences, Department of Applied Research, Dronten, The Netherlands

Functional Plant Biology (Accepted)

Abstract

Nitrate (NO_3^-) fertilizers are commonly used to improve the yield and quality of most non-legume crops, such as potato (*Solanum tuberosum* L.). Root cells absorb nitrate from the soil using plasma membrane-bound transporters. In this study, we overexpressed a putative nitrate transporter from potato (*StNPF1.11*) to study its effect on the level of tuber protein content in potato. At 10 weeks after planting, overexpression of *StNPF1.11* increased the mean level of protein content of all $N = 23$ transformants by 42% compared to the wild type control. The levels of chlorophyll content in leaves (from upper and lower plant parts) were also increased for several individuals at 10 weeks. Tuber yield (fresh) was not structurally impaired, however the mean tuber dry matter content of the transformants was reduced by 3-8% at 19 weeks. At 19 weeks, an overall increase in protein content was not observed. Throughout plant development, half of the transformants were taller than the control. Moderate positive correlations were observed between tissue-specific *StNPF1.11* expression and protein content. Our findings show that overexpression of *StNPF1.11* increased both the abundance of nitrogen-containing molecules in different tissues and plant height of young potato plants, seemingly at a minor expensive of tuber dry matter (starch) content. Basic understanding of the mechanisms that regulate plant nitrogen uptake, transport and utilization, enable the development of tools to improve both crop nutrition and quality that are needed to enhance the viability and sustainability of future plant production systems.

Keywords

Nitrate, Nitrate transporter 1/Peptide transporter Family (NPF), Potato, *StNPF1.11*, Tuber protein content, Leaf chlorophyll content, Plant height

Introduction

Over the past two decades, much progress has been made in understanding molecular processes related to nitrate uptake and allocation in higher plants. Nitrate is taken up from the soil by proton-coupled plasma membrane-bound transporters of which numerous variants belong to the large nitrate transporter 1/peptide transporter family (*NPF*, also known as the *NRT1* family), where co-transport of hydrogen ions (H^+) provide the driving force of the system (Miller and Smith 1996; Ruiz-Cristin and Briskin 1991). Besides nitrate, studies in the plant model species *Arabidopsis* have illustrated that *NPFs* may also transport hormones and other molecules. These include the hormones auxin (Krouk et al. 2010) and abscisic acid (Kanno et al. 2012), dipeptides (Rentsch et al. 1998) and the amino acid histidine (Frommer et al. 1994). To capture nitrate under varying concentrations in the soil, plant root cells possess different *NPFs* that display low, high or dual affinities for nitrate acquisition (Krapp et al. 2014). Depending on the plant species, low-affinity *NPFs* (LATS) show linear kinetics (*Km*) in the millimolar (mM) range, whilst high-affinity *NPFs* (HATS) show a *Km* of ~5-100 in µM concentrations (Aslam et al. 1992; Doddema and Telkamp 1979; Goyal and Huffaker 1986; Lee and Drew 1986; Meharg and Blatt 1995). In general, LATS exhibit faster transporter systems that HATS (Touraine and Glass 1997). Dual-affinity transporters possess the ability to functions as both LATS and HATS (Liu et al. 1999). Plants regulate nitrate/nitrogen transport, assimilation and allocation according to nitrogen availability and need, by means of (post)-transcriptional and (post-)translational regulation (Meyer and Stitt 2001). After uptake from the soil, nitrate is added to the cell cytosolic pool or utilized directly. Use of nitrate is orchestrated by the nitrogen satiety-status, where the availability of carbon and the cytosolic ion homeostasis to sustain growth. Nitrate may also be stored temporarily in vacuoles as a reserve (Martinoia et al. 1981). After uptake, nitrate is reduced to nitrite by nitrate reductase (*NR*), subsequently nitrite is reduced to ammonium by nitrite reductase (*NiR*). The glutamine synthetase (*GS*) or glutamate synthase (*GOGAT*) and glutamate dehydrogenase (*GDH*) pathways finally convert ammonium into the building blocks for the synthesis of amino acids, (storage) proteins, chlorophyll,

nucleotides and other nitrogen-containing compounds that include (glyco)alkaloids, phenylpropanoids and glucosinolates.

As reviewed by Wang et al. (2018b), a plethora of molecular studies involving *NPFs* have been carried out in *Arabidopsis*, rice and wheat. Natural genetic variation of rice *OsNPF6.5/NRT1.1b* has been hypothesized to have contributed to the diverged use of nitrate between rice sub-species (Hu et al. 2015). Overexpression of nitrate transporters have also been shown to induce large effects on key agronomic traits in different crops. For example, overexpression of the pH-sensitive HATS *OsNRT2.3b* gene increased grain yield and improved the nitrogen-use efficiency (NUE) and phosphorous uptake and translocation in rice (Fan et al. 2016; Feng et al. 2017). *NPFs* may also play important roles in regulating key quality traits in other crops. However functional studies involving *NPFs* in important food crops, including potato (*Solanum tuberosum* L.), are still lacking. Here, we studied the overexpression effect of *StNPF1.11* on the levels of protein content in tubers, leaf chlorophyll content, tuber yield and tuber dry matter content at 10 and 19 weeks after planting.

Materials and methods

StNPF1.11 underlies QTLs for tuber protein content in potato

The potato nitrate transporter gene (*StNPF1.11*; NCBI ID: XM_006355891) was selected from putative candidate genes underlying intervals spanning quantitative trait loci (QTLs) for protein content in potato (Klaassen et al. 2019; Klaassen et al. 2020; Werij 2011). In potato, *StNPF1.11* is located on chromosome *5* at the physical position of 7.436 Mb (PGSC Pseudomolecules v4.03). *StNPF1.11* belongs to the large nitrate transporter 1/peptide transporter family (NPF) (http://plants.ensembl.org). The DNA sequence of *StNPF1.11* (PGSC0003DMG400015591) shows two gene models / splice variants (PGSC0003DMT400040275; PGSC0003DMT400040276) in the potato reference genome (PGSC Pseudomolecules v4.03). *StNPF1.11* is orthologous to *AtNPF1.1*, *AtNPF1.2* and *AtNPF1.3* from *Arabidopsis* (Fig. S7).

Phylogenetic analysis

Phylogenetic analysis was performed using MEGA version 7.0 (Kumar et al. 2016). Phylogenetic ties of protein sequences from *Arabidopsis* and potato *NPF* genes were inferred using the neighbour-joining method (Saitou and Nei 1987). Distances were calculated using the Poisson correction method (Zuckerkandl and Pauling 1965). All sequences were collected from the NCBI Genbank (https://www.ncbi.nlm.nih.gov/).

Cloning and construct design

The construct (Fig. S1) was designed for constitutive expression of *StNPF1.11* by the CaMV 35S promoter from the cauliflower mosaic virus (Odell et al. 1985). The insert of *StNPF1.11* was amplified from cDNA that originated from the tetraploid ($2n = 4x = 48$) starch potato progenitor *KA 2005-1496* (Averis Seeds, Valthermond, The Netherlands). Progenitor *KA 2005-1496* was selected for cloning due to its high level of tuber protein content. PCR was performed using Phusion high-fidelity polymerase (NEB, Ipswich, MA, USA) according to the product protocol guideline. Gene specific PCR primers (Table S1) were designed using Lasergene software (DNAStar, Madison, WI, USA). Isolation of the encoding fragment was carried out using the Qiagen Gel Purification Kit (Qiagen, Hilden, Germany) according to the protocol of the manufacturer. Next, the encoding fragment was cloned using Gateway technology (Invitrogen, Gaithersburg, MD, USA) and sequenced for validation. The fragment was then transferred into pENTR donor vectors (Thermo Fisher Scientific) and later synthesized into Gateway pK7WG2 destination vectors (Invitrogen, Gaithersburg, MD, USA). Reactions were mediated by Gateway LR Clonase II enzyme mix (Invitrogen, Gaithersburg, MD, USA) as described by Karimi et al. (2002) to generate pK7WG2-*StNPF1.11* expression vectors. Isolated plasmids of the pK7WG2-*StNPF1.11* expression vectors were transferred into competent *A. tumerfaciens* AGL1 cells via electroporation as described (Takken et al. 2000), and used for transformation.

Transformation and regeneration of transformants

The pK7WG2-*StNPF1.11* expression vector (Fig. S1) was introduced into the tetraploid ($2n = 4x = 48$) starch potato variety *Kardal* (Averis Seeds, Valthermond, The Netherlands). *Kardal* was selected for transformation due to its moderate and relatively stable level of tuber protein content. *A. tumerfaciens*-mediated transformation was performed as described (Heilersig et al. 2006). The developed *StNPF1.11* transformants (OE-lines) were coded as OE-*x*-*y*, where *x* denoted the generated series of the transformants and *y* denoted the unique transformant line number. The wild type control (untransformed *Kardal*) and T1 transformants were cultured from *in-vitro* explants to develop biological replicates as described (Visser et al. 1991).

Quantitative real-time PCR (qRT-PCR) analysis

The levels of gene expression were analysed in different tissue types. Small leaves, large leaves, stems and small tubers were collected from the plants at 10 weeks after planting. At 19 weeks, large tubers were collected from the senesced plants. Small leaves were collected from the first to third compound leaf below the shoot apex. Larger compound leaves were collected from the sixth to ninth leaf below the shoot apex. Stem samples were collected from the internodes positioned at the centre of the plant. Collected tissues were frozen directly in liquid nitrogen and stored at -80^{o}C. Isolation of RNA, cDNA synthesis and qRT-PCR were carried out as described (Xu et al. 2016). PCR primer sequences (Table S2) were used to quantify the transcripts of *StNPF1.11*. Elongation factor-1α (EF-1α) was used as the reference gene (Nicot et al. 2005), as its expression is stable across different tissues in potato (Lopez-Pardo et al. 2013). Relative gene expression was computed using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen 2001).

Greenhouse conditions

The T1 transformants were planted in 3 litre pots as 10 cm sized *in-vitro* plants. The plants were grown in a greenhouse (Unifarm, Wageningen UR, Wageningen, The Netherlands). Biological replicates of transformants were grown in rows of eight consecutive plants. Complete rows of eight plants were randomly distributed in the greenhouse. The pots contained peat soil, fertilized

with PG MIX 15-10-20 (N-P-K) (Yara Ltd., Grimsby, UK) to EC = 0.8. Additionally, the plants were supplied with slow-release Osmocote Pro 19-9-10-2 (N-P-K-Mg) (Mertens, Horst, The Netherlands) that also contained trace elements (TE) (2 g/liter). The growth trial was carried out from September 2017 to January 2018 (16 hours light at 20ºC; 8 hours darkness at 18ºC). The *in-vitro* plants from series 1 (OE-1; 16 transformants in total), were transferred to the greenhouse one week prior to those of series 2 (OE-2; 7 transformants in total).

Leaf chlorophyll content

Levels of leaf chlorophyll content (SPAD units) were measured in fully expanded compounds leaves, using the hand-held SPAD-502 device (Minolta Camera Co., Ltd, Osaka, Japan). This device measured the transmittance of red (650 nm) and infrared (940 nm) radiation that passed through the leaves. Leaves from upper (fourth from apex) and lower (positioned ~10 cm above the soil) plant parts were measured.

Quantification of soluble tuber protein content

Total soluble tuber protein content was quantified using SPRINT^{TM} Rapid Protein Analyser (CEM Corporation, NC, USA) as described (Klaassen et al. 2019). Purified potato tuber protein (AVEBE, Veendam, The Netherlands) was used to produce the calibration curve. Internal controls were included for diagnostic validation. Potato fruit juice (PFJ) was extracted from fresh tubers (>1 cm). Tuber fresh weight (yield) and tuber dry matter (DM) content were measured using a gravimetric scale. To determine total tuber DM, samples were dried to constant weight at 70ºC. Protein content (% w/w) was expressed in milligram protein per gram tuber FW. In our study, 1% protein represented 10 mg of protein per g fresh tuber.

Statistical analysis

Data were analysed using Fisher's least significant difference (LSD) post hoc test at $\alpha = 0.05$ from one-way analysis of variance (ANOVA). Pearson's correlation coefficient (r) was used to verify associations between the variables. Statistical analyses were carried out in SPSS version 23.0 (IBM

Corp., Armonk, NY, USA). Principal components analyses were carried out using R software package Factoextra (Kassambara and Mundt 2016).

Results

Variable *StNPF1.11* expression in transformants and tissues

To measure *StNPF1.11* expression levels in the transformants and tissues, qRT-PCR was carried out. Expression of *StNPF1.11* differed between the transformants and tissues (Fig. 1). The highest expression levels were observed in large leaves (up to 21-fold). Several transformants (*e.g.* OE-1-19) showed high expression levels of *StNPF1.11* in different tissues, whilst others showed low values (*e.g.* OE-1-23 and OE-1-38).

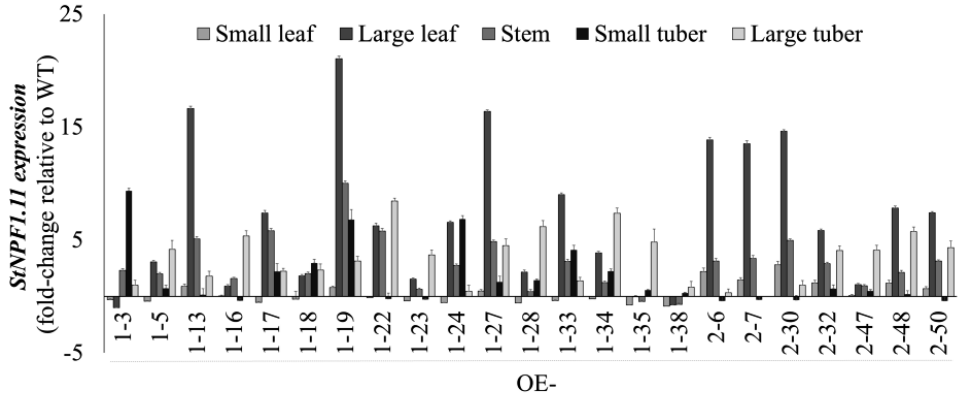

Fig. 1. Relative gene expression of *StNPF1.11* (mean ± SD) in small leaves, large leaves, stems, small tubers (10 weeks) and large tubers (19 weeks). qRT-PCR were performed in $N = 2$ technical replicates using pooled tissues that were collected from four biological replicates ($N = 4$). SD = standard deviation. OE = overexpression line. OE-1 = series 1. OE-2 = series 2. Relative expression were based on the $2^{-\Delta\Delta CT}$ method.

StNPF1.11 overexpression did not impair fresh tuber yield

Tuber properties were measured to evaluate potential pleiotropic effects induced by *StNPF1.11* overexpression. Tuber yield did not differ from the control (Fig. S4). At 10 and 19 weeks, one and four of the twenty three

transformants showed reduced values for tuber yield (fresh weight) respectively. At 10 weeks, the tuber dry matter content of the transformants did not show clear differences compared to the control (Fig. S5). However, at 19 weeks half of the transformants showed slightly reduced dry matter content (3-8%) compared to the control.

Stretched internodes and increased leaf chlorophyll content

Half of the transformants were taller than the control (Fig. 2; Fig. S2). The mean plant height of the two transformant series were increased (due to stretched internodes) by 13-21% and 15-22% at 10 and 19 weeks respectively. The height of one transformant (*i.e.* OE-1-22) was lower than the control. At 10 weeks, higher levels of chlorophyll content (SPAD units) were observed in leaves from lower and upper plant parts for several transformants. Transformant OE-2-50 showed the highest increase (39%) compared to the control (Fig. 3). The mean level of chlorophyll content of the two transformant series were increased by 3-5% in lower leaves and 5-7% in upper leaves. At 19 weeks, chlorophyll content could not be measured as the plants were senesced.

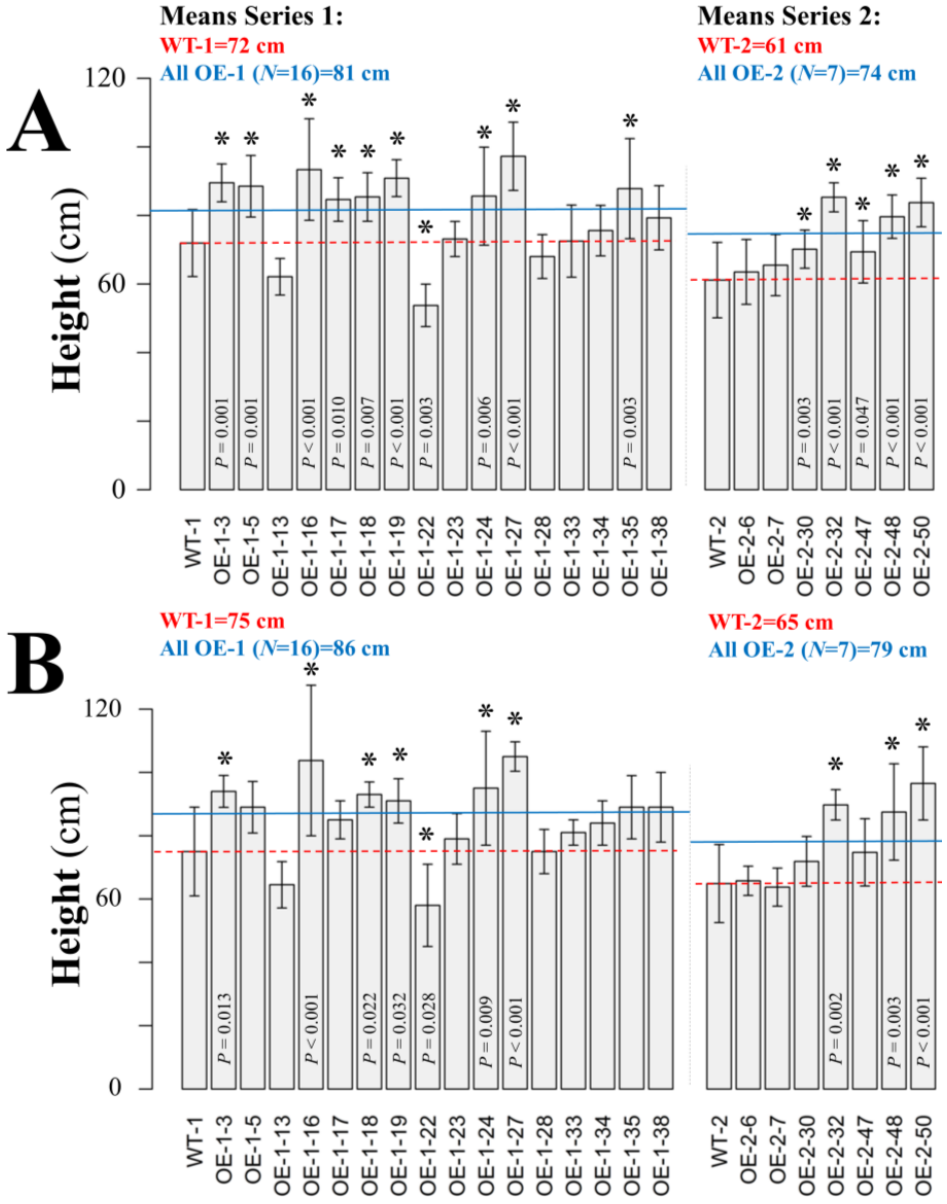

Fig. 2. Plant height (mean ± SD) of the transformants at (**A**) 10 weeks and (**B**) 19 weeks. Data were collected from N = 6-8 biological replicates at 10 weeks. At 19 weeks, four (N = 4) biological replicates were measured. The blue horizontal lines and red dashed lines represent the mean height for the transformants series and wild type control respectively. P denotes Fisher's least significant difference (LSD) probabilities from one-way ANOVA. Asterisks' denote significant differences at α = 0.05 relative to the wild type (WT) control. OE = overexpression line. OE-1 = Series 1. OE-2 = Series 2. SD = standard deviation.

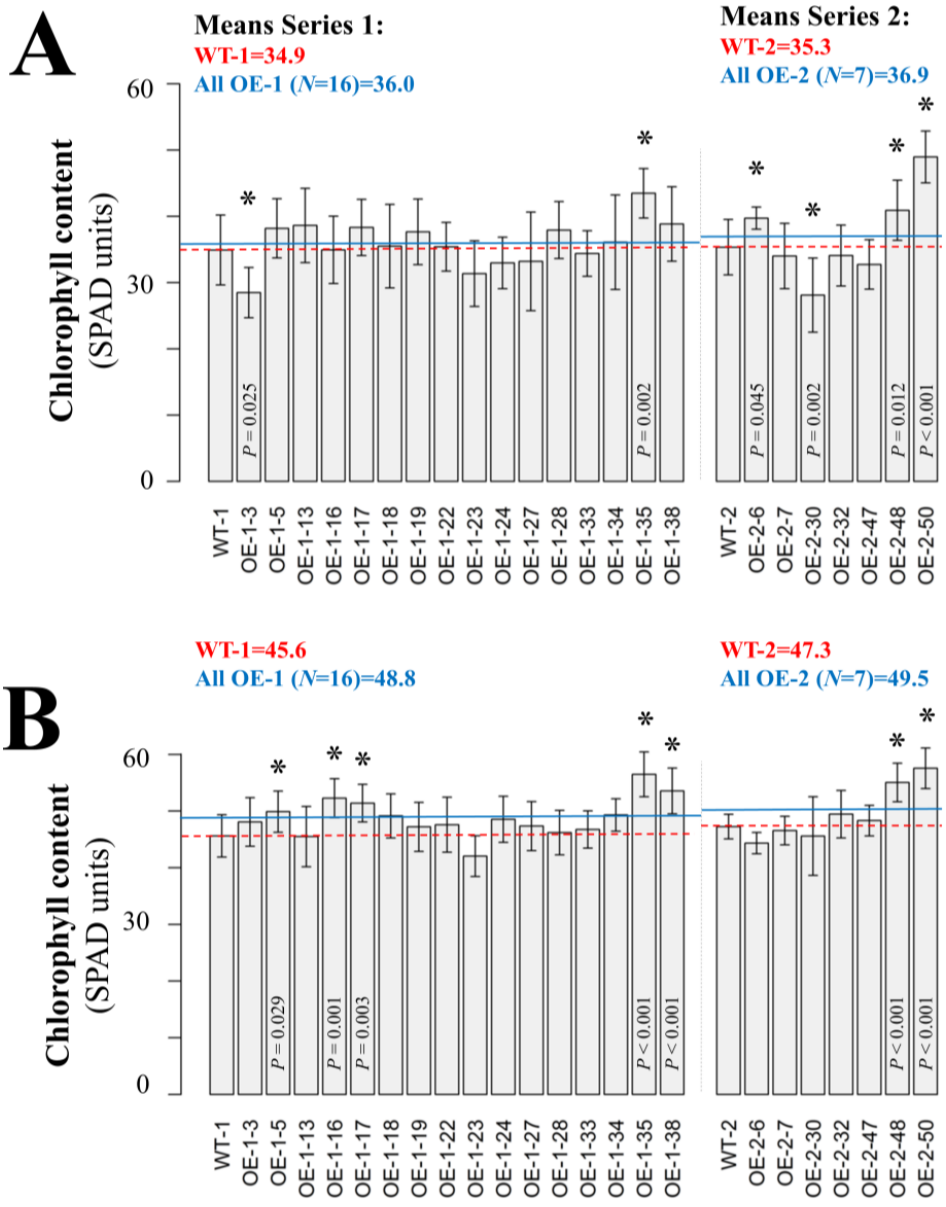

Fig. 3. Chlorophyll content in SPAD units (mean ± SD) in (**A**) lower and (**B**) upper leaves of the transformants at 10 weeks. Data were collected from six to eight biological replicates (N = 6-8), each measured in three technical replicates (N = 3). The blue horizontal lines and red dashed lines represent the average chlorophyll content for the transformants series and wild type control respectively. P denotes Fisher's least significant difference (LSD) probabilities from one-way ANOVA. Asterisks' denote significant differences at α = 0.05 relative to the wild type (WT)

control. OE = overexpression line. OE-1 = Series 1. OE-2 = Series 2. SD = standard deviation.

Overexpression of *StNPF1.11* increased tuber protein content in young tubers at 10 weeks

At 10 weeks, the levels of tuber protein content were higher for all, but one, of the twenty three transformants compared to the control (Fig. 4). At 10 weeks, the mean protein content of all transformants was 42% higher than the control. Several transformants (OE-1-19, OE-1-27, OE-2-30 and OE-2-50) showed strong increased values (73-126%). At 19 weeks, overexpression of *StNPF1.11* did not structurally increase protein content, although the two lines OE-1-19 and OE-1-27 still showed significantly higher values (28%) and seven other transformants showed higher values than the control, but not significant. These transformants were also taller at 10 weeks compared to the control.

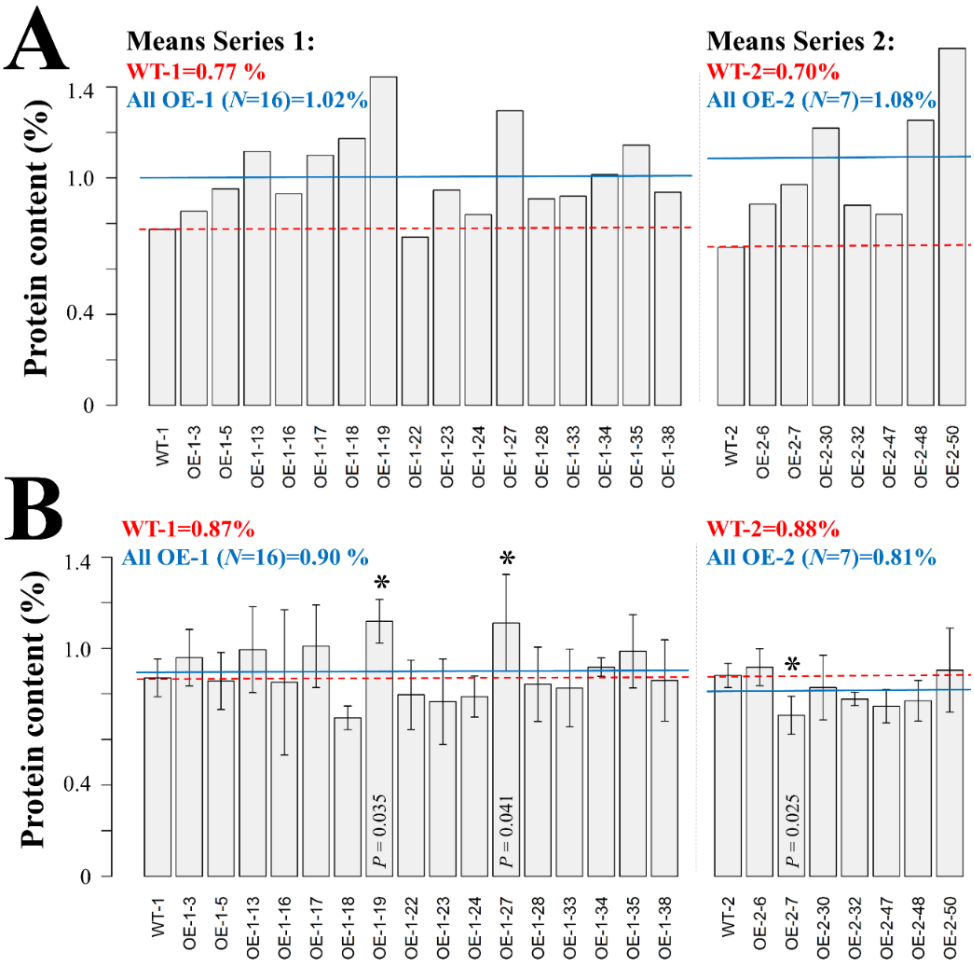

Fig. 4. Total soluble tuber protein content (mean ± SD) of the transformants at (**A**) 10 weeks and (**B**) 19 weeks after planting. At 10 weeks, tubers from four biological replicates ($N = 4$) were pooled and measured, therefore no SD bars are shown. At 19 weeks, tubers were collected from four biological replicates ($N = 4$) that were measured individually. All samples were measured in two technical replicates ($N = 2$). The blue horizontal lines and red dashed lines represent the average protein content values for the transformants series and wild type controls respectively. *P* denotes Fisher's least significant difference (LSD) probability values from one-way ANOVA. Asterisks' denote significant differences at $\alpha = 0.05$ relative to the wild type (WT) control. OE = overexpression line. OE-1 = Series 1. OE-2 = Series 2. SD = standard deviation.

Positive correlations between *StNPF1.11* expression and tuber protein content

Principal component and correlation analyses were carried out to evaluate the relationships between tissue-specific *StNPF1.11* expression, plant properties and tuber traits (Fig. 5). The first two PCA components (dimensions) accounted for 51% of the variance, where the first and second components each explained 26% and 51% respectively. Moderate positive correlations (r = 0.36-0.50) were observed between tuber protein content versus *StNPF1.11* expression in large leaves, stems and large tubers, especially at 10 weeks (see scatter bi-plots in Fig. S6). Moderate negative correlations were found between tuber dry matter content versus plant height (r = -0.31 to -0.55) and chlorophyll content in upper leaves (r = -0.34 to -0.48) (Fig. 7). At 10 weeks, protein content showed a moderate positive correlation with chlorophyll content and plant height (r = 0.39 to 0.48), whereas at 19 weeks these correlations were weaker. Between 10 and 19 weeks, protein content, tuber yield (fresh) and tuber dry matter content again showed moderate positive correlations (r = 0.39-0.49) (Fig. 6). Plant height showed a high positive correlation between 10 and 19 weeks (r = 0.96).

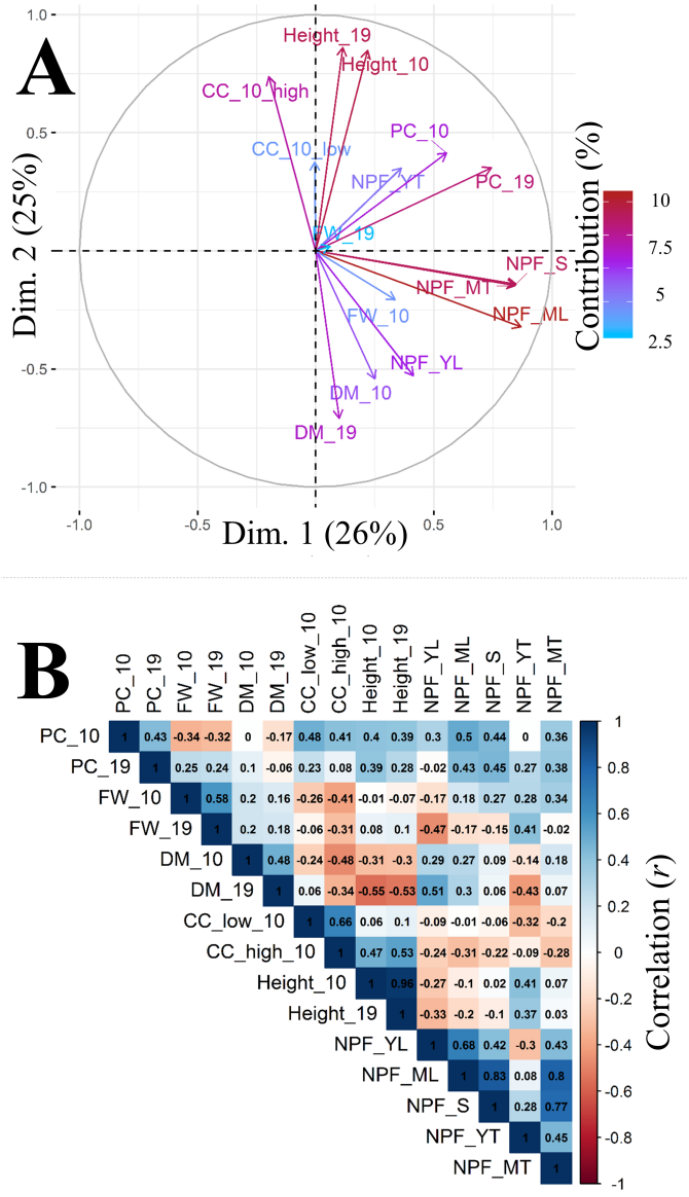

Fig. 5. (**A**) Principal components bi-plot and (**B**) correlation matrix for tissue-specific expression of *StNPF1.11*, tuber protein content, tuber yield, tuber dry matter content, leaf chlorophyll content and plant height of the transformants. Length of the arrows approximate the variance explained by the variables. The numbers in matrix show the correlation coefficient (Pearson's *r*) between the variables. PC = tuber protein content. FW = tuber yield. DM = tuber dry matter content. CC = leaf chlorophyll content. NPF = *StNPF1.11* expression. YL = small leaf. ML = large leaf. S = stem. YT = small tuber. MT = large tuber. 10 = 10 weeks. 19 = 19 weeks. Dim. = dimension.

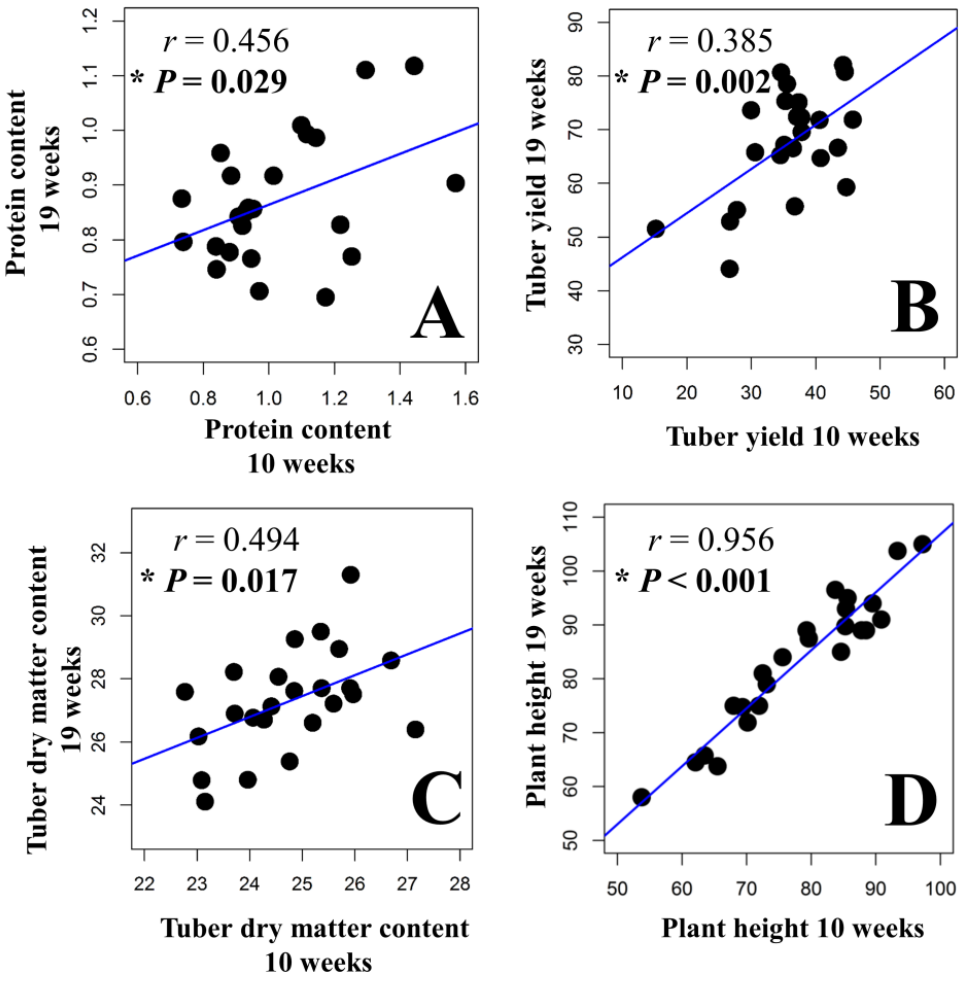

Fig. 6. Scatterplots for (**A**) tuber protein content, (**B**) tuber yield, (**C**) tuber dry matter content and (**D**) plant height of the transformants at 10 versus 19 weeks. Linear regression lines are shown in blue. *r* represent Pearson's correlation coefficients. *P* denotes the probability values. The asterisks' denote significant relationships at α = 0.05.

Discussion

This study was carried out to investigate the overexpression effect of a putative nitrate transporter (*StNPF1.11*) on the level of tuber protein content in potato tubers. At 10 weeks, the levels of protein content were clearly higher for the transformants compared to the wild type control, whereas at 19 weeks a clear overall increase for this trait was not observed. Taller plants were observed during both stages of plant development, and several transformants showed increased leaf chlorophyll content at 10 weeks. Moderate positive correlations were found between protein content and *StNPF1.11* expression levels in leaves, tubers and stems. Our results showed that overexpression of *StNPF1.11* affected tuber protein content, leaf chlorophyll content and plant height.

Overexpression of *StNPF1.11* increased leaf chlorophyll and tuber protein content at 10 weeks

The levels of chlorophyll (a molecule containing four nitrogen atoms) content were higher in the leaves of several transformants (Fig. 3). In literature it has been shown that leaf chlorophyll content (deduced from SPAD values) shows a high positive correlation with total nitrogen (N) content (Gianquinto et al. 2004). Nitrogen has been suggested to affect leaf thickness (Peng 1992). Therefore, it is noteworthy to investigate whether leaf thickness, cell size or the density of chloroplasts caused the increased levels of leaf chlorophyll content.

At 10 weeks, overexpression of *StNPF1.11* structurally increased the level of tuber protein content by up to 126% (Fig. 4). At this stage of plant development, the mean level of protein content for all transformants was 42% higher than the control. This structural effect was not clearly observed at 19 weeks. At this developmental stage merely two of the twenty three transformants showed significantly increased values (28%), although several other transformants showed higher non-significant values. Different substrates have been reported to be transported by plant *NPF* proteins, as shown in *Arabidopsis* (Chiba et al. 2015). Our data suggest that *StNPF1.11* may be involved in N transport in young potato plants, as both the levels of

tuber protein content and leaf chlorophyll content (that is associated with total leaf nitrogen) were higher than the control at 10 weeks. It remains unclear to us why we observed clear differences for protein content at 10 weeks but not at 19 weeks. Whether transcriptional or translational factors played a role remains to be elucidated. The levels of protein content showed variation between the transformants (Fig. 4). This variation may have been caused by spatial (Fig. 1) and temporal differences in *StNPF1.11* expression. Transcripts from CaMV 35S promoter-driven constructs may differ both spatially and temporally, as shown in tobacco (Williamson et al. 1989). Williamson et al. (1989) demonstrated that CaMV 35S directed expression was higher in young active tissues than in older quiescent types. Also, transcription enhancers positioned upstream of the CaMV 35S promoter (Fang et al. 1989) may also lead to differential accumulation of transcripts in different tissues. The number of construct inserts and the chromosome environments may also have played a role. To explore interactions between N-responsive genes in the nitrogen assimilation pathway and putative nitrate transporters (*NPF* genes), amino acid transporters and peptide transporters in potato, it is certainly relevant to study the overexpression effects of *StNPF1.11* as well as other *NPF* transporters from this gene family (Fig. S7). The use of tissue-specific promoters, *e.g.* from the potato tuber granule-bound starch synthase (GBSS) gene or patatin genes, may elicit more specific responses in harvestable tissues in potato. The use of different genetic backgrounds and contrasting soil N-conditions, present promising designs to study the (dynamic) biological functions of *NPF* genes in more detail. Also, dedicated molecular studies will reveal which specific substrates are transported by *StNPF1.11*.

Overexpression of *StNPF1.11* did not impair tuber yield but mildly reduced tuber dry matter content

Overexpression of *StNPF1.11* did not impair tuber yield of the transformants at 10 and 19 weeks compared to the control. Plant height of the transformants was increased by overexpressing *StNPF1.11*, both at 10 and 19 weeks. As an indirect effect, the mean tuber dry matter content of the transformants was reduced by 3-8% (Fig. S5). Very high N-fertilizer rates tend to stimulate vegetative growth in potato plants, that coincides with reduced tuber yield and reduced tuber dry matter content whilst increasing N-content in tubers

(Bélanger et al. 2002; Westermann et al. 1994; Zebarth and Rosen 2007). Nitrogen is also known to delay the onset of flowering and senescence in plants (Marín et al. 2011). Delayed flowering subsequently prolongs the period of maturation and may affect the bulking-dynamics of storage compounds (*e.g.* starch) (Scott et al. 1973; Withrow 1945). Overexpression of *StNPF1.11* may therefore have stimulated vegetative plant growth due to increased uptake of N, resulting in taller plants (Fig. S2). Increased vegetative growth may subsequently have altered the tuber sink strength, potentially leading to reduced tuber dry matter content for half (but not all) of the transformants at 19 weeks. Hence, dissecting the potential effect of *StNPF1.11* on sink strength is certainly worthwhile. Moreover, it is relevant to study the interplay between the major maturity regulator in potato (*StCDF1*) and *NPF* transporters as their patterns of transcription seem to be linked (Varala et al. 2018). These studies may also provide insight into tuber sink strength and the C/N balance.

Molecular functions of the *StNPF1.11* gene

The protein sequence of *StNPF1.11* is orthologous to three *Arabidopsis NPF* genes (Fig. S7: *AtNPF1.1*, *AtNPF1.2* and *AtNPF1.3*). *AtNPF1.1* and *AtNPF1.2* are capable of translocating nitrate, gibberellin (GA) and jasmonoyl-isoleucine (JA-Ile) (Chiba et al. 2015; Hsu and Tsay 2013). *AtNPF1.1* is also able to transport abscisic acid (ABA) (Chiba et al. 2015). At present it is still unknown which substrates are transported by *AtNPF1.3*. We observed stretched internodes that may have resulted from altered GA transport. GA is a well-known plant hormone that modulates shoot elongation by stimulating cell division and elongation (Hedden and Proebsting 1999; Kende and Zeevaart 1997). We showed that tuber protein content and leaf chlorophyll content were higher at 10 weeks. Increased nitrogen or nitrate uptake by overexpressing *StNPF1.11* may have caused these effects, but this must still be proven. Without doubt, future studies will be carried out to verify whether GA, nitrate or potentially other compounds are transported by *StNPF1.11*. It has been proposed that *NPF* proteins function as single components for low-affinity nitrate transporter (Liu and Tsay 2003). Overexpressing of *StNPF1.11* induced clear phenotypic effects, therefore we hypothesize that *StNPF1.11* is able to function as a single component that

apparently does not require a partner protein. It has been reported that several high-affinity (HATS) nitrate transporters require a partner protein to function (Zhou et al. 2000), whereas this has not been shown for low-affinity nitrate (LATS) transporters. In any case, molecular studies are needed to validate the putative LATS property of *StNPF1.11* and its substrate binding affinity.

Concluding remarks

Taken together, this study clearly demonstrated that overexpression of a putative nitrate transporter (*StNPF1.11*) from potato structurally increased tuber protein content and leaf chlorophyll content of young potato plants. Moreover, stretched internodes were observed throughout plant development. For the first time, we report that tuber protein content was increased by overexpressing a potato (*NPF* transporter) gene.

Acknowledgements

The authors kindly thank Dirk Jan Huigen and Henk Meurs for practical assistance to carry out the growth trials.

Supplementary information

Table S1 Forward (F) and reverse (R) primer sequences for PCR amplification of the putative potato nitrate transporter (*StNPF1.11*) encoding fragment.

Gene	Gene ID	Primers sequence 5'-3'
StNPF1.11	XM_006355891	F: caccATGAAACTGAAAATGGGCACAGAATCAC R: TCAAGATTCATGTAAGAGTCTATAATCACATTCTTC

Table S2 Forward (F) and reverse (R) primer sequences for qRT-PCR of the overexpressed putative potato nitrate transporter (*StNPF1.11*) and potato elongation factor (*EF1α*).

Gene	Gene ID	Primers sequence 5'-3'
StNPF1.11	XM_006355891	F: TCCTATGTGGTCTGCTGGTTT R: AGGGACAAAAACACGGTCGTA
Elongation factor (*EF1α*)*	AB061263	F: ATTGGAAACGGATATGCTCCA R: TCCTTACCTGAACGCCTGTCA

*Sequences of primers for potato elongation factor (*EF1α*) from (Nicot et al. 2005).

Fig. S1 Schematic diagram of the pK7WG2-*StNPF1.11* construct used for transformation (not to scale). The potato nitrate transporter (*StNPF1.11*) gene was cloned in the pK7WG2 destination vector (Karimi et al. 2002) and under control of the CaMV 35S promoter for constitutive expression in all plant tissues.
RB: right border; 35S-P: CaMV35S promotor; AttR1: recombination site 1; *StNPF1.11*: potato nitrate transporter gene (ID: XM_006355891; PGSC0003DMG400015591); AttR2: recombination site 2; 35S-T: CaMV35S terminator; KAN: kanamycin resistance gene; LB: left border.

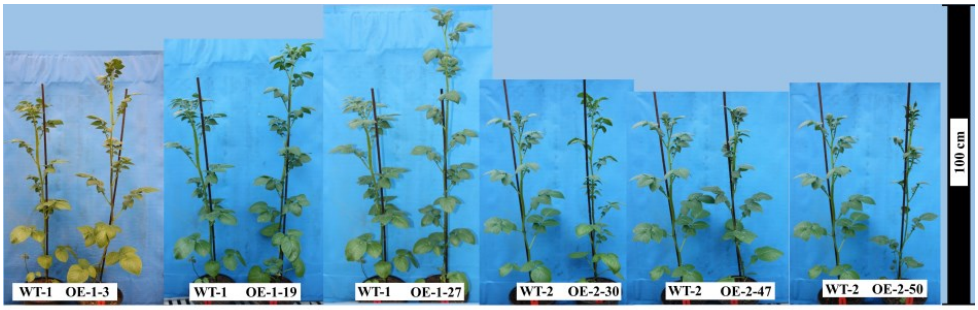

Fig. S2 Plant height of selected transformants at 10 weeks. Several transformants were taller than the wild type control. OE = overexpression line. WT = wild type. 1 = Series 1. 2 = Series 2.

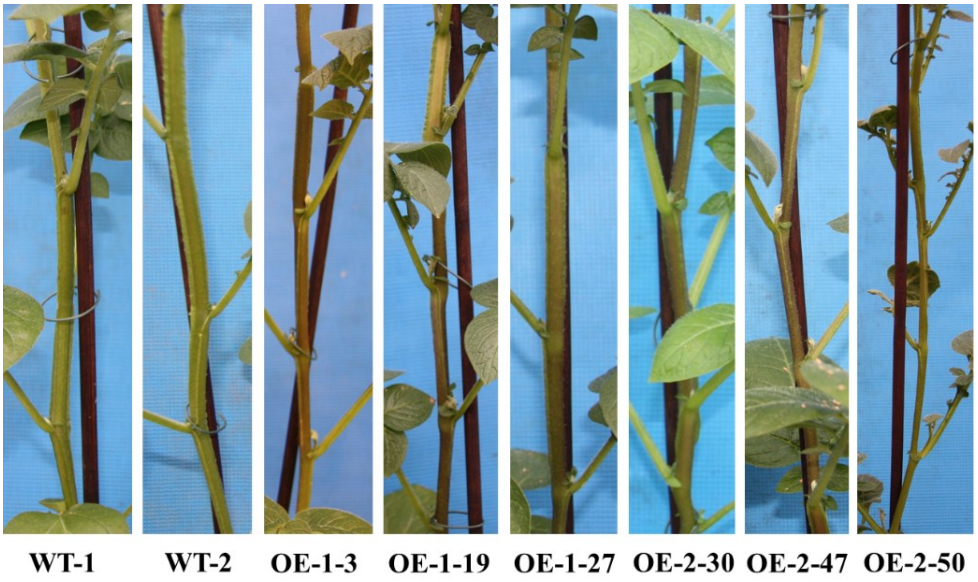

Fig. S3 Darker pigmented (brown) stems observed for selected transformants, whereas normal (light green) coloured stems were observed for the wild type (WT) controls. OE = overexpression line. 1 = Series 1. 2 = Series 2.

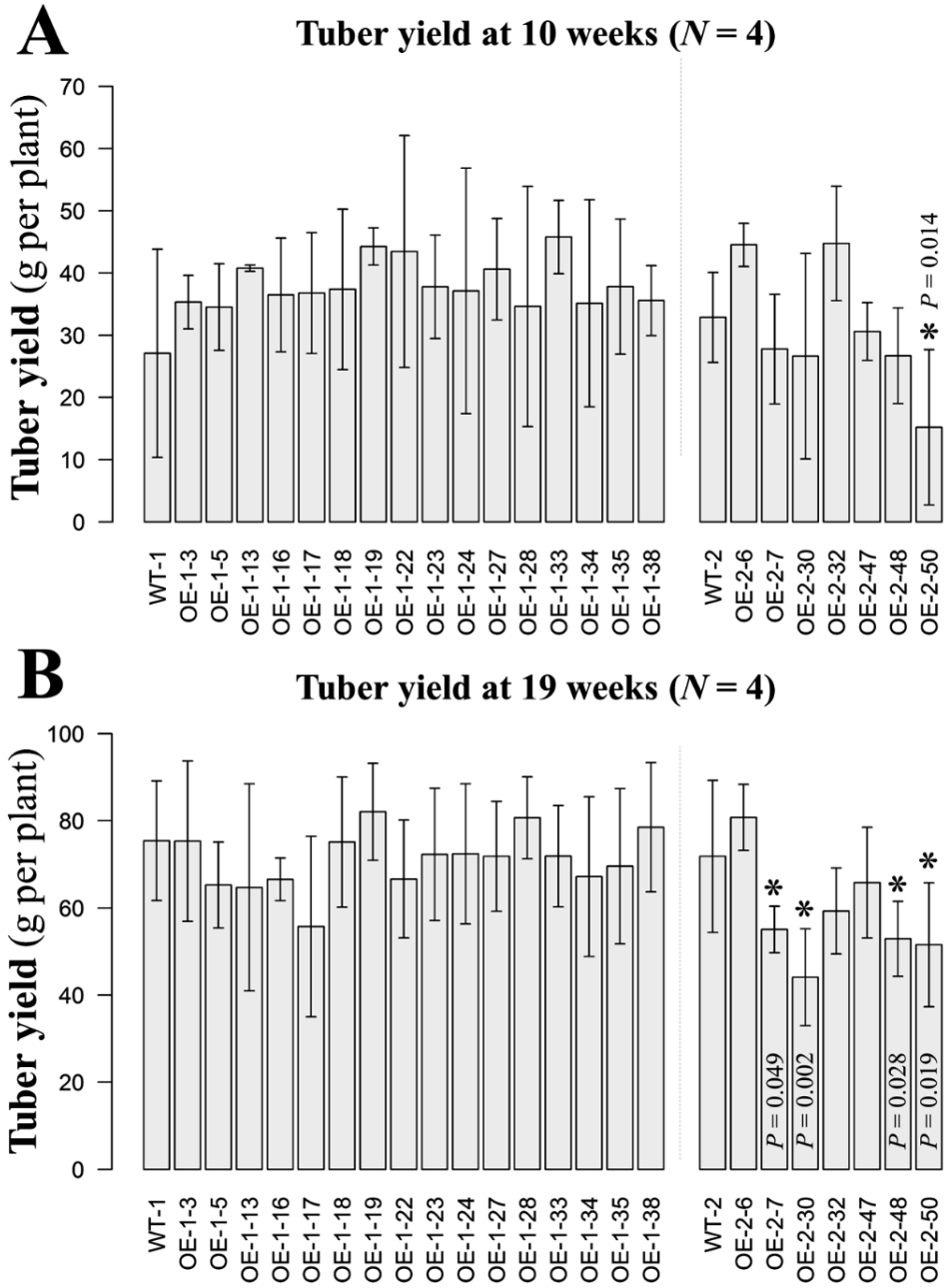

Fig. S4 Tuber yield (fresh weight) of the transformants at (**A**) 10 weeks and (**B**) 19 weeks. Data (mean ± SD) represent samples collected from four biological replicates ($N = 4$). P denotes Fisher's least significant difference (LSD) probability values from one-way ANOVA. Asterisks denote significant differences at $\alpha = 0.05$.

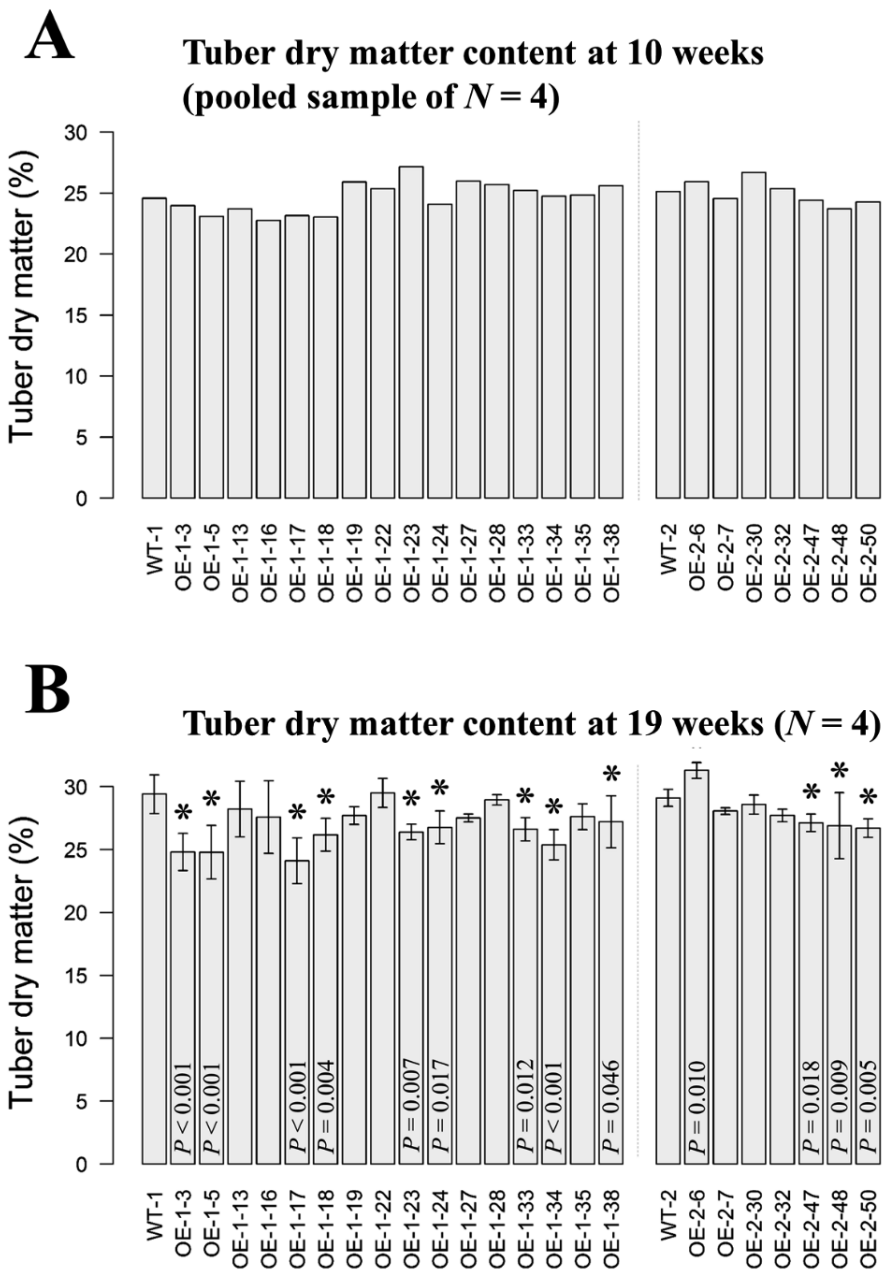

Fig. S5 Tuber dry matter content (mean ± SD) at (**A**) 10 weeks and (**B**) 19 weeks. Data represent (A) pooled samples from four biological replicates ($N = 4$). *P* represent Fisher's least significant difference (LSD) probability values from one-way ANOVA. Asterisks denote significant differences at $\alpha = 0.05$.

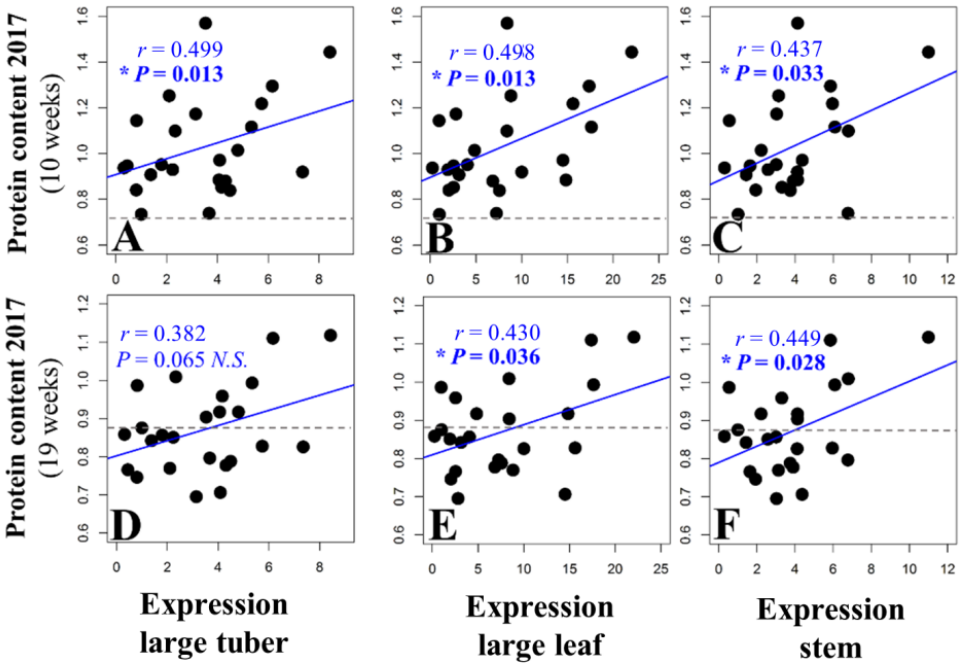

Fig. S6 Bi-variate scatter plots for *StNPF1.11* expression in large leaves (10 weeks), stems (10 weeks) and large tubers (19 weeks) of the transformants versus tuber protein content at 10 and 19 weeks. The grey dashed lines represent the values of protein content for the wild type (WT) control. Simple linear regression lines are shown in blue. *P* represents the probability values. Asterisks denote significant relationships at $\alpha = 0.05$. *N.S.* represent non-significant relationships

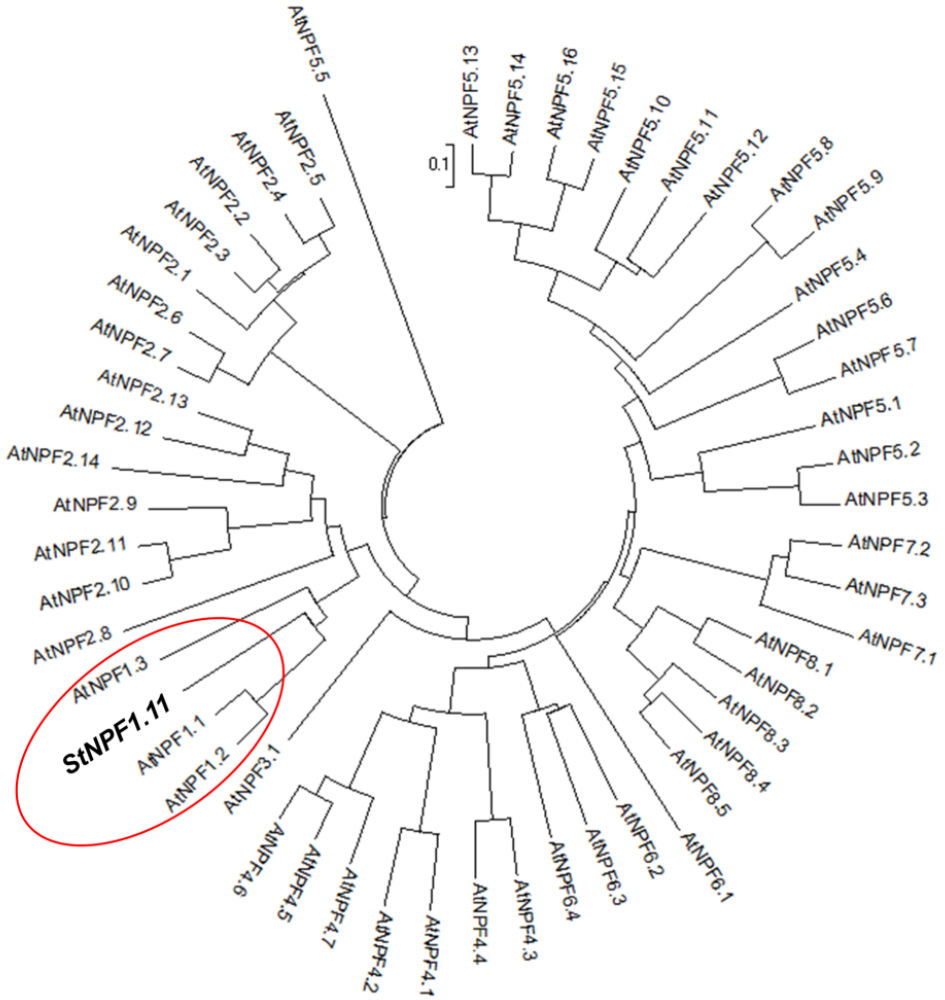

Fig. S7 Phylogenetic ties between potato *StNPF1.11* and *Arabidopsis* NPF orthologs. Protein sequences were collected from NCBI Genbank. Analyses and alignments were performed using MEGA version 7.0 (Kumar, Stecher and Tamura, 2016). The *StNPF1.11* sequence is orthologous to *AtNPF1.1*, *AtNPF1.2* and *AtNPF1.3* (shown in the red oval sphere). At = *Arabidopsis thaliana*.

References

Aslam M, Travis RL, Huffaker RC (1992) Comparative kinetics and reciprocal inhibition of nitrate and nitrite uptake in roots of uninduced and induced barley (Hordeum vulgare L.) seedlings. Plant Physiology 99 (3):1124-1133

Bélanger G, Walsh JR, Richards JE, Milburn PH, Ziadi N (2002) Nitrogen fertilization and irrigation affects tuber characteristics of two potato cultivars. American Journal of Potato Research 79 (4):269-279

Chiba Y, Shimizu T, Miyakawa S, Kanno Y, Koshiba T, Kamiya Y, Seo M (2015) Identification of Arabidopsis thaliana NRT1/PTR FAMILY (NPF) proteins capable of transporting plant hormones. Journal of Plant Research 128 (4):679-686

Doddema H, Telkamp G (1979) Uptake of nitrate by mutants of Arabidopsis thaliana, disturbed in uptake or reduction of nitrate: II. Kinetics. Physiologia Plantarum 45 (3):332-338

Fan X, Tang Z, Tan Y, Zhang Y, Luo B, Yang M, Lian X, Shen Q, Miller AJ, Xu G (2016) Overexpression of a pH-sensitive nitrate transporter in rice increases crop yields. Proceedings of the National Academy of Sciences 113 (26):7118-7123

Fang R-X, Nagy F, Sivasubramaniam S, Chua N-H (1989) Multiple cis regulatory elements for maximal expression of the cauliflower mosaic virus 35S promoter in transgenic plants. The Plant Cell 1 (1):141-150

Feng H, Li B, Zhi Y, Chen J, Li R, Xia X, Xu G, Fan X (2017) Overexpression of the nitrate transporter, OsNRT2.3b, improves rice phosphorus uptake and translocation. Plant Cell Reports 36 (8):1287-1296

Frommer WB, Hummel S, Rentsch D (1994) Cloning of an Arabidopsis histidine transporting protein related to nitrate and peptide transporters. FEBS Letters 347 (2-3):185-189

Gianquinto G, Goffart JP, Olivier M, Guarda G, Colauzzi M, Dalla Costa L, Delle Vedove G, Vos J, Mackerron DKL (2004) The use of hand-held chlorophyll meters as a tool to assess the nitrogen status and to guide nitrogen fertilization of potato crop. Potato Research 47 (1):35-80

Goyal S, Huffaker R (1986) A novel approach and a fully automated microcomputer-based system to study kinetics of NO− 3, NO− 2, and NH+ 4 transport simultaneously by intact wheat seedlings. Plant, Cell & Environment 9 (3):209-215

Hedden P, Proebsting WM (1999) Genetic analysis of gibberellin biosynthesis. Plant Physiology 119 (2):365-370

Heilersig BH, Loonen AE, Wolters A-MA, Visser RG (2006) Presence of an intron in inverted repeat constructs does not necessarily have an effect on efficiency of post-transcriptional gene silencing. Molecular Breeding 17 (4):307-316

Hsu P-K, Tsay Y-F (2013) Two phloem nitrate transporters, NRT1. 11 and NRT1. 12, are important for redistributing xylem-borne nitrate to enhance plant growth. Plant Physiology 163 (2):844-856

Hu B, Wang W, Ou S, Tang J, Li H, Che R, Zhang Z, Chai X, Wang H, Wang Y (2015) Variation in NRT1. 1B contributes to nitrate-use divergence between rice subspecies. Nature Genetics 47 (7):834

Kanno Y, Hanada A, Chiba Y, Ichikawa T, Nakazawa M, Matsui M, Koshiba T, Kamiya Y, Seo M (2012) Identification of an abscisic acid transporter by functional screening using the receptor complex as a sensor. Proceedings of the National Academy of Sciences 109 (24):9653-9658

Kassambara A, Mundt F (2016) Factoextra: extract and visualize the results of multivariate data analyses. R package version 1 (3)

Kende H, Zeevaart J (1997) The five" Classical" plant hormones. The Plant Cell 9 (7):1197

Klaassen MT, Bourke PM, Maliepaard C, Trindade LM (2019) Multi-allelic QTL analysis of protein content in a bi-parental population of cultivated tetraploid potato. Euphytica 215 (2):14

Klaassen MT, Willemsen JH, Vos PG, Visser RGF, van Eck HJ, Maliepaard C, Trindade LM (2019) Genome-wide association analysis in tetraploid potato reveals four QTLs for protein content. Molecular Breeding 39 (11):151

Krapp A, David LC, Chardin C, Girin T, Marmagne A, Leprince A-S, Chaillou S, Ferrario-Méry S, Meyer C, Daniel-Vedele F (2014) Nitrate transport and signalling in Arabidopsis. Journal of Experimental Botany 65 (3):789-798

Krouk G, Lacombe B, Bielach A, Perrine-Walker F, Malinska K, Mounier E, Hoyerova K, Tillard P, Leon S, Ljung K (2010) Nitrate-regulated auxin transport by NRT1. 1 defines a mechanism for nutrient sensing in plants. Developmental Cell 18 (6):927-937

Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular Biology and Evolution 33 (7):1870-1874

Lee R, Drew M (1986) Nitrogen-13 studies of nitrate fluxes in barley roots: II. Effect of plant N-status on the kinetic parameters of nitrate influx. Journal of Experimental Botany 37 (12):1768-1779

Liu K-H, Huang C-Y, Tsay Y-F (1999) CHL1 is a dual-affinity nitrate transporter of Arabidopsis involved in multiple phases of nitrate uptake. The Plant Cell 11 (5):865-874

Liu KH, Tsay YF (2003) Switching between the two action modes of the dual-affinity nitrate transporter CHL1 by phosphorylation. The EMBO Journal 22 (5):1005-1013

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2− ΔΔCT method. Methods 25 (4):402-408

Lopez-Pardo R, Ruiz de Galarreta JI, Ritter E (2013) Selection of housekeeping genes for qRT-PCR analysis in potato tubers under cold stress. Molecular Breeding 31 (1):39-45

Marín IC, Loef I, Bartetzko L, Searle I, Coupland G, Stitt M, Osuna D (2011) Nitrate regulates floral induction in Arabidopsis, acting independently of light, gibberellin and autonomous pathways. Planta 233 (3):539-552

Martinoia E, Heck U, Wiemken A (1981) Vacuoles as storage compartments for nitrate in barley leaves. Nature 289 (5795):292

Meharg A, Blatt M (1995) NO 3− transport across the plasma membrane of Arabidopsis thaliana root hairs: kinetic control by pH and membrane voltage. The Journal of Membrane Biology 145 (1):49-66

Meyer C, Stitt M (2001) Nitrate reduction and signalling. In: Plant Nitrogen. Springer, pp 37-59

Miller AJ, Smith SJ (1996) Nitrate transport and compartmentation in cereal root cells. Journal of Experimental Botany 47 (7):843-854

Nicot N, Hausman J-F, Hoffmann L, Evers D (2005) Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. Journal of Experimental Botany 56 (421):2907-2914

Odell JT, Nagy F, Chua N-H (1985) Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. Nature 313:810

Peng S (1992) Leaf thickness affects the estimation of leaf nitrogen concentration using a chlorophyll meter. Int Rice Res Notes 17:19-20

Rentsch D, Boorer K, Frommer W (1998) Structure and function of plasma membrane amino acid, oligopeptide and sucrose transporters from higher plants. The Journal of Membrane Biology 162 (3):177-190

Ruiz-Cristin J, Briskin DP (1991) Characterization of a H+ NO3− symport associated with plasma membrane vesicles of maize roots using 36ClO3− as a radiotracer analog. Archives of Biochemistry and Biophysics 285 (1):74-82

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4 (4):406-425

Scott R, Ogunremi E, Ivins J, Mendham N (1973) The effect of fertilizers and harvest date on growth and yield of oilseed rape sown in autumn and spring. The Journal of Agricultural Science 81 (2):287-293

Takken FL, Luderer R, Gabriëls SH, Westerink N, Lu R, De Wit PJ, Joosten MH (2000) A functional cloning strategy, based on a binary PVX-expression vector, to isolate HR-inducing cDNAs of plant pathogens. The Plant Journal 24 (2):275-283

Touraine B, Glass AD (1997) NO3-and ClO3-Fluxes in the chl1-5 Mutant of Arabidopsis thaliana (Does the CHL1-5 Gene Encode a Low-Affinity NO3-Transporter?). Plant Physiology 114 (1):137-144

Varala K, Marshall-Colón A, Cirrone J, Brooks MD, Pasquino AV, Léran S, Mittal S, Rock TM, Edwards MB, Kim GJ (2018) Temporal transcriptional logic of dynamic regulatory networks underlying nitrogen signaling and use in plants. Proceedings of the National Academy of Sciences 115 (25):6494-6499

Visser RG, Stolte A, Jacobsen E (1991) Expression of a chimaeric granule-bound starch synthase-GUS gene in transgenic potato plants. Plant Molecular Biology 17 (4):691-699

Werij JS (2011) Genetic analysis of potato tuber quality traits. PhD dissertation. Chair group: Plant Breeding. Wageningen University, Wageningen, The Netherlands. Retrieved from http://edepot.wur.nl/183746

Westermann D, Tindall T, James D, Hurst R (1994) Nitrogen and potassium fertilization of potatoes: yield and specific gravity. American Potato Journal 71 (7):417-431

Williamson JD, Hirsch-Wyncott ME, Larkins BA, Gelvin SB (1989) Differential Accumulation of a Transcript Driven by the CaMV 35S Promoter in Transgenic Tobacco. Plant Physiology 90 (4):1570-1576. doi:10.1104/pp.90.4.1570

Withrow AP (1945) The interrelationship of nitrogen supply and photoperiod on the flowering, growth and stem anatomy of certain long and short day plants. Butler University Botanical Studies 7 (1/13):40-64

Xu X, Dechesne A, Visser RG, Trindade LM (2016) Expression of an (Engineered) 4, 6-α-Glucanotransferase in Potato Results in Changes in Starch Characteristics. PloS One 11 (12):e0166981

Zebarth B, Rosen CJ (2007) Research perspective on nitrogen BMP development for potato. American Journal of Potato Research 84 (1):3-18

Zhou J-J, Fernández E, Galván A, Miller AJ (2000) A high affinity nitrate transport system from Chlamydomonas requires two gene products. FEBS Letters 466 (2-3):225-227

Zuckerkandl E, Pauling L (1965) Evolutionary divergence and convergence in proteins. In: Evolving genes and proteins. Elsevier, pp 97-166

Chapter 5

RG-I galactan side-chains are involved in the regulation of the water-binding capacity of potato cell walls

Michiel T. Klaassen[1,2], Luisa M. Trindade[1]

[1] Wageningen University & Research, Plant Breeding, Wageningen, The Netherlands

[2] Aeres University of Applied Sciences, Department of Applied Research, Dronten, The Netherlands

This chapter has been published in:

***Carbohydrate Polymers* (2020) 227:115353**

DOI: 10.1016/j.carbpol.2019.115353

Abstract

Potato cell walls (PCW) are a low value by-product from the potato starch industry. Valorisation of PCW is hindered by its high water-binding capacity (WBC). The composition of polysaccharides and interactions between these entities, play important roles in regulating the WBC in the cell wall matrix. Here, we show that *in vivo* exo-truncation of RG-I β-(1→4)-D-galactan side-chains decreased the WBC by 6-9%. In contrast, exo-truncation of these side-chains increased the WBC by 13% *in vitro*. We propose that degradation of RG-I galactan side-chains altered the WBC of PCW, due to cell wall remodelling and loosening that affected the porosity. Our findings reinforce the view that RG-I galactan side-chains play a role in modulating WBC, presumably by affecting polysaccharide architecture (spacing) and interactions in the matrix. Better understanding of structure-function relationships of pectin macromolecules is needed before cell wall by-products may be tailored to render added-value in food and biobased products.

Keywords

Water-binding capacity (WBC), Potato cell walls (PCW), Pectin, Rhamnogalacturonan I (RG-I), β-(1→4)-D-galactan, β-galactosidase

Introduction

Plant cell walls consist of a matrix of polysaccharides and minor amounts of (glyco)proteins. Besides surrounding and protecting the inner cell compartments, cell walls fulfil numerous functions in plant development. These functions include cell differentiation, organogenesis, adhesion, expansion and wall mechanical strength (Aldington and Fry 1993; Cosgrove 2000; McCann and Roberts 1994; Satoh 1998). Between species, plant cell walls display a high degree of diversity in composition and structure, where water is a major integral component (Brett and Hillmann 1985).

Potato tubers are mainly composed of parenchyma cells, with typical thin primary cell walls (Lisinska and Leszczynski 1989; McDougall et al. 1996). Tuber skin (periderm) largely consists of cork phellem (Lisinska and Leszczynski 1989). Cell walls of interior tuber tissues are composed of cellulose (30%) and hemicellulose (11% xyloglucan and 3% mannan), that hold together a vast quantity of pectic polysaccharides (56%) (Vincken et al. 2000). These pectic polysaccharides are rich in rhamnogalacturonan I (RG-I), a branched heteropolymer that accounts for 50-75% of total pectin in potato tubers (Oomen 2003; Vincken et al. 2000). The RG-I backbone polymer consists of repeating disaccharide units of L-rhamnose and D-galacturonic acid: [-α-L-Rha*p*-(1→4)-α-D-GalA*p*-(1→2)] (McNeil et al. 1980). Neutral *β*-(1→4)-D-galactose (galactan) and *α*-(1→5)-L-arabinose (arabinan) side-chains are attached at the *O*-4 positions of rhamnose moieties on the RG-I backbone (Carpita and Gibeaut 1993; Schols and Voragen 1994). Potato RG-I galactan side-chains may be substituted by short chains of galactan or arabinan, also known as type I arabinogalactan (Carpita and Gibeaut 1993; Øbro et al. 2004; Ridley et al. 2001). RG-I galactan side-chains are abundant in potato, where they account for 28-36% of the cell wall (Øbro et al. 2004; Vincken et al. 2000).

Many endeavours have been made to define the biological roles of RG-I galactan side-chains in different species. However, their definitive functions remain a matter of debate. These structures have been suggested to function and maintain open pores in the cell wall matrix through spatial separation of cellulose microfibrils (McCartney et al. 2000; Roach et al. 2011; Baron-Epel

et al. 1988). Potato RG-I galactan side-chains have been implicated to associate with cellulose *in-vitro* (Zykwinska et al. 2005). More recently, these interactions have been shown to be more abundant than previously thought (Wang et al. 2015; Wang et al. 2012). It has also been suggested that RG-I galactan side-chains affect the bio-mechanical properties of cell walls (Dick-Perez et al. 2012; Larsen et al. 2011; McCartney et al. 2000; Tang et al. 1999; Ulvskov et al. 2005), presumably by modulating the hydration capacity of the cell wall matrix. Under hydrated conditions, RG-I galactan side-chains are highly mobile and enhance interactions with water (Ha et al. 2005; Larsen et al. 2011). Water embedded in the cell wall matrix is thought to maintain spatial structures and pores (Makshakova et al. 2018). Intermolecular forces (determined by the architecture and composition of the cell wall), may also bind or entrap water through dipole interactions (Labuza 1968). Dipole interactions (*e.g.* hydrogen bonds) arise from hydrophilic pectin groups that include hydroxyl, carboxyl and amide entities (Matveev et al. 2000). Moreover, pH, (an)ions and drying conditions that disrupt polymer organization (Moore et al. 2008), have been implicated to affect interactions of cell walls with water (Renard et al. 1994; Serena and Knudsen 2007).

It is well established that pectins are important swelling components of the primary cell wall (Thakur et al. 1997). Based on NMR spectroscopy and enzymatic studies, pectic side-chains have been pointed out to regulate the hydration capacity of plant cell walls (Belton 1997; Funami et al. 2011; Larsen et al. 2011; Ramasamy et al. 2015; Ramaswamy et al. 2013). To the best of our knowledge however, the specific truncation of RG-I galactan side-chains in cell walls has not been studied with regard to the water-binding capacity (WBC). In this study, we evaluated the effects of *in-vivo* and *in-vitro* exo-truncation of RG-I β-(1→4)-D-galactan side-chains on the WBC of potato cell walls (PCW).

Materials and methods

Plant material

Three transgenic potato lines (β-GAL) expressing β-(1,4)-galactosidase from chickpea (*Cicer arietinum*) were used (Martín et al. 2005). Expression of β-galactosidase was driven by the potato granule-bound starch synthase (GBSS) promoter. The tetraploid ($2n = 4x = 48$) starch variety *Karnico* (Averis Seeds, Valthermond, The Netherlands), served as the genetic background of the β-GAL lines and control (wild type). The potato plants were grown in pots in an outdoor screen cage (Unifarm, Wageningen UR, Wageningen, The Netherlands), during the potato growing season in The Netherlands (April to September, 2016). At the end of the growing season, tubers were harvested from 50 individual plants of the β-GAL lines and the control. Tuber fresh weight and dry matter content were determined directly after harvest. Tuber dry matter content was determined, after drying the samples at 105°C until constant weight. The harvested tubers were stored for 5 weeks at 5°C prior to the extraction of raw PCW.

Extraction of raw PCW

To extract raw PCW, tubers were processed in a set-up that mimicked the industrial potato starch recovery process as reported earlier (Ramasamy et al. 2014). Tuber batches (3-5 kg) were processed sequentially. Prior to milling, the tubers were washed in a rotating drum to remove traces of soil. Milling was performed by using a spinning cylindrical teeth grinder (type: RU 40-260, Nivoba, Veendam, The Netherlands). To inhibit enzymatic browning of the slurry, 0.1% (v/w) of 10% sodium metabisulphite (w/v) was added during the milling process. After milling, the slurry was filtered four times over a 90 µm centrifugal sieve (Larssons, Bromolla, Sweden). This step was carried out to wash out free starch granules and to acquire raw PCW samples. Raw PCW samples were lyophilized and used for further experiments.

Starch degradation

To acquire PCW with a low (or no) starch content, residual starch was hydrolysed enzymatically according to the following protocol. Samples of 5 g PCW (dry matter) were mixed in a 500 mL sodium acetate buffer solution (0.2 M, pH = 5.6) and homogenized with a Ultra-turrax disperser. To gelatinize starch and inactivate endogenous enzymes, the mixtures were heated for 30 min at 80°C. Next, the mixtures were incubated and gently stirred (120 rpm) for 4 h at 40°C, together with a dose of 2000 U α-amylase from porcine pancreas (Sigma-Aldrich, St. Louis, MO, USA). Afterwards, the pH was lowered to 4.6 by adding acetic acid (glacial: 99.9%). Next, 500 U amyloglucosidase of *Rhizopus sp.* (Megazyme, Bray, Ireland) was added. The mixtures were incubated for 2 h at 40°C under gentle stirring conditions (120 rpm). Subsequently, the polysaccharides were precipitated using ethanol 70% (v/v). After 15 min of precipitation, the residues were collected and repeatedly subjected to another two cycles of heating and enzymatic degradation. The final residues were lyophilized to acquire de-starched PCW.

Scanning electron microscopy (SEM)

To inspect for starch in PCW, samples were visualized microscopically by using scanning electron microscopy (SEM) (Phenom™, FEI, Eindhoven, The Netherlands) as previously described (Xu et al. 2017).

In-vitro β-galactosidase treatment

De-starched PCW samples (300 mg, dry matter) were treated with β-(1,4)-galactosidase (EC 3.2.1.23, *A. niger*) (Megazyme, Bray, Ireland). Incubations were carried out in 30 mL sodium acetate buffer solutions (0.1 M) at a pH of 4.5 for 48 h at 40°C. Gentle stirring took place during incubation (60 rpm). Doses of β-galactosidase (600 U) were added at the start and again after 24 h during the incubation process.

Starch content

Starch content (w/w) was determined using a commercially available starch quantification kit (R-Biopharm AG, Darmstadt, Germany).

Monosaccharide composition

The PCW samples were pre-hydrolysed for 60 min at 30°C using 72% (w/w) sulphuric acid. Subsequently, the mixtures were diluted to a 4% (w/w) sulphuric acid concentration using deionized water. The samples were further hydrolysed for 180 min at 100°C. Afterwards, the samples were centrifuged at 15,000 × *g* for 15 min and the supernatant phases were collected. Dilutions were made for determining the content of glucose (dilution factor 100) and the contents of rhamnose, arabinose, galactose, mannose, xylose and uronic acids (dilution factor 7). The monosaccharide contents were quantified using high-performance anion exchange chromatography, with pulsed amperometric detection (HPAEC-PAD) and HPLC-Dionex™ ICS-5000$^+$ DC (Thermo Fischer Scientific, Waltham, MA, USA). HPAEC-PAD runs were carried out using a Dionex CarboPac™ PA1 guard column (2 × 250mm). Eluent solutions were used as solvents: 0.1 M NaOH, 1 M sodium acetate (NaAc) in 0.1 M NaOH and deionized water. Volumes of 2.5-5 µL passed through the system at a flow rate of 250 µL per minute at 30°C. Recovery standards were employed to correct for monosaccharide losses, due to destruction by acid hydrolysis (Sluiter et al. 2008). Galacturonic acid content was inferred from total uronic acids using HPAEC-PAD. The length of RG-I galactan side-chains and ratio of homogalacturonan (HG) to RG-I were calculated as follows (Huang et al. 2017):

$$Galactan\ side\ chain\ length\ of\ RGI = \frac{galactose}{rhamnose} \qquad (1),$$

$$Ratio\ HG\ to\ RGI = \frac{galacturonic\ acid - rhamnose}{2 \times rhamnose} \qquad (2).$$

Water-binding capacity (WBC)

The water-binding capacity of PCW was quantified according to a modified centrifugation method (Pustjens et al. 2012). PCW samples (250 mg, dry matter) were added to 30 mL deionized water and stirred for 5 min at 500 rpm. Next, the samples were centrifuged using nylon centrifugal filters to remove bound water (pore size: 0.45 µm, F2519-4, Thermo Fischer Scientific, Waltham, MA, USA). After centrifugation (1328 × *g* for 5 min), the wet PCW weight was measured gravimetrically. To quantify the dry weight of the samples, wet samples were lyophilized for 48 h. All steps were carried out at

room temperature. The water-binding capacity (WBC) was expressed in millilitre (mL) water per gram (g) dry PCW (Thibault et al. 2000). WBC was calculated as follows:

$$Water\ binding\ capacity\ (WBC) = \frac{wet\ PCW\ weight\ (g)}{dry\ PCW\ weight\ (g)} \qquad (3).$$

Results and discussion

Plant performance

To assess the potential impact of β-galactosidase on yield, several tuber properties of the β-GAL lines were compared to the control. Earlier work by Martín et al. (2005) showed that β-galactosidase expression levels were high, moderate and low in β-GAL lines 7, 14 and 27 respectively. Tuber yield (fresh weight) was not affected (**Table 1**); although the yield of β-GAL line 27 was lower than the control, but not statistically significant. For line β-GAL-27, tuber dry matter content was lower ($P < 0.05$), whilst the total amount of extracted PCW was higher ($P < 0.05$) relative to the control. These effects were not observed for the other two β-GAL lines. No significant differences were observed for starch content in the tubers and raw PCW. Our findings are in line with earlier studies, reporting that *in-vivo* expression of β-galactosidase does not (clearly) impair tuber yield, nor does it induce noticeable phenotypic changes (Huang et al. 2017; Mayer and Hillebrandt 1997; Meyer et al. 2009).

Table 1 Potato tuber and the raw PCW properties of the *β*-GAL lines and control

Line	Tuber yield (g FW per plant)	Tuber dry matter content (% FW)	Starch content in tubers (% DM)	Raw PCW content in tubers (% DM)	Starch content in raw PCW (% DM)
β-GAL-7	228 ± 79	25.5 ± 0.9	74.7 ± 2.8	1.49 ± 0.15	33.0 ± 2.9
β-GAL-14	236 ± 36	25.5 ± 0.3	74.7 ± 1.7	1.46 ± 0.32	32.7 ± 2.1
β-GAL-27	207 ± 60	22.2 ± 1.6 *	74.0 ± 2.6	2.72 ± 0.48 *	29.9 ± 1.5
Control (WT)	236 ± 82	25.9 ± 0.7	75.3 ± 2.2	1.93 ± 0.40	29.9 ± 1.8

Data (mean ± SD). Data for tuber yield and tuber dry matter content were collected from five biological replicates ($N = 5$). Starch content was measured in samples from three biological replicates ($N = 3$), each measured in four technical replicates ($N = 4$). PCW = potato cell walls. *β*-GAL = *β*-galactosidase transgenic line. FW = fresh weight. DM = dry matter. Control (WT) = wild type (untransformed *Karnico*). Asterisks' denote significant differences between the *β*-GAL lines and control at $\alpha = 0.05$, derived from one-way ANOVA post-hoc (LSD). SD = standard deviation.

Monosaccharide composition

To assess potential modifications of the cell wall composition due to *β*-galactosidase activity, monosaccharides were quantified in PCW samples of the *β*-GAL lines and compared to the control. Galactose levels (µg/g) were clearly reduced by 32-67% for the *β*-GAL lines, whereas the levels of rhamnose were slightly higher (**Table 2**). Huang et al. (2016) showed that the transgene most significantly reduced the length of galactan side-chains in the hot buffer soluble solids (HBSS) extracts, where the non-bound HBSS sub-species was strongly affected. As described in Section 3.3, *β*-galactosidase activity *in-vivo* may have induced other (pleiotropic) effects by modifying the composition and architecture of other cell wall polysaccharides such as xyloglucan. Changes in galactose and arabinose were in line with previous studies regarding these *β*-GAL lines (Huang et al. 2016; Huang et al. 2017; Martín et al. 2005). Our findings confirmed that *in-vivo* expression of *β*-galactosidase strongly reduced galactose levels in PCW of the *β*-GAL lines. The length of RG-I galactan side-chains were shortened, as shown by the reduced ratio of galactose to rhamnose by 34-71%. Structural modification of the ratio homogalacturonan (HG) to rhamno-galacturonan I (RG-I) was not observed.

Table 2 Monosaccharide composition (µg/g) of PCW from the tubers of the *β*-GAL lines and control

	µg/g PCW (dry basis)								
PCW sample	Rha	Ara	Gal	Glc$^{\Delta}$	Man	Xyl	GalA	Gal:Rha	HG:RG-I
β-GAL-7	349 ± 34	801 ± 66	1053 ± 73	6671 ± 353	326 ± 33	363 ± 36	2367 ± 180	3	2.9
β-GAL-14	330 ± 18	934 ± 41	1580 ± 44	6795 ± 122	320 ± 17	446 ± 24	2514 ± 109	4.8	3.3
β-GAL-27	319 ± 18	961 ± 40	2141 ± 55	6582 ± 168	299 ± 16	380 ± 20	2255 ± 142	6.7	3
Control (WT)	309 ± 25	981 ± 50	3150 ± 83	6497 ± 161	303 ± 19	365 ± 26	2249 ± 116	10.2	3.2
Control (WT) after starch degradation	427 ± 26	1250 ± 58	3931 ± 97	4758 ± 161	300 ± 14	477 ± 31	2625 ± 130	9.2	2.6

Data (mean ± SD). *β*-GAL = *β*-galactosidase transgenic line. PCW = potato cell walls. Control = untransformed control (wild type). Rha = rhamnose; Ara = arabinose; Gal = galactose; Man = mannose; Xyl = xylose; GalA = galacturonic acid. Gal:Rha = ratio galactose to rhamnose (RG-I galactan side-chain length). HG:RG-I = ratio homogalacturonan (HG) to RG-I. Δ = includes glucose in starch. Data were collected from three ($N = 3$) biological replicates, each measured in two ($N = 2$) technical replicates. SD = standard deviation.

In PCW from the wild type control, starch granules were present as loose entities or were entrapped in partially erupted or intact cells (**Figure 1**). After enzymatic starch degradation, the granules were not visible anymore and the starch content was reduced from 29.9% to 1.9% (w/w).

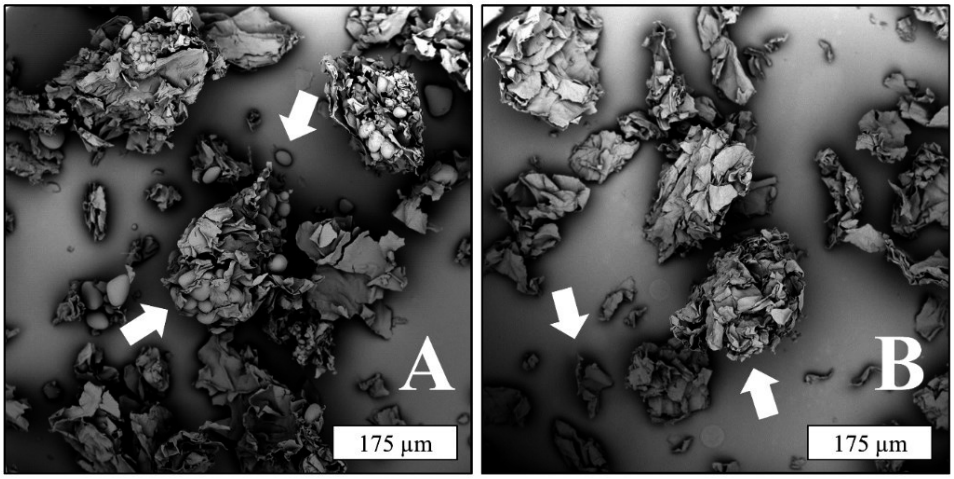

Figure 1 Electron micrographs of coalesced potato cell walls (PCW) from the tubers from the wild type control (untransformed *Karnico*) (**A**) before starch degradation showing both entrapped and loose starch granules and (**B**) after enzymatic starch degradation that removed all starch granules. Scanning election microscopy (SEM) was used to capture the micrographs of the lyophilized PCW samples. The white arrows show the presence and absence of starch granules in the PCW samples.

β-galactosidase affected WBC *in-vivo* and *in-vitro*

To study the effect of RG-I galactan side-chains on the WBC of PCW, we degraded these side chains in both *in-vivo* and *in-vitro* conditions. Expression of *β*-galactosidase *in-vivo* structurally reduced the WBC by 6-9% for all three *β*-GAL lines (**Figure 2**). This reduction corresponded to a clear decrease in galactose content and shorter (or less abundant) RG-I galactan side-chains (**Table 2**). In contrast, *in-vitro β*-galactosidase degradation increased the WBC by 13% compared to the control (**Figure 3**). This increase corresponded to a release of mostly galactose and traces of glucose and galacturonic acid (**Figure 4**). Here, we show that *β*-galactosidase affected the WBC of PCW in both *in-vivo* and *in-vitro* conditions, but with contrasting effects. Contrasting effects may be encountered when *in-vivo* versus *in-vitro* systems are compared. For instance, it has been reported that de-esterification of HG from plant cell walls led to contrasting biophysical properties in *in-vivo* and *in-vitro* setups (Braybrook and Peaucelle 2013; Goldberg et al. 1986; Peaucelle et al. 2011; Peaucelle et al. 2015; Zhao et al. 2008).

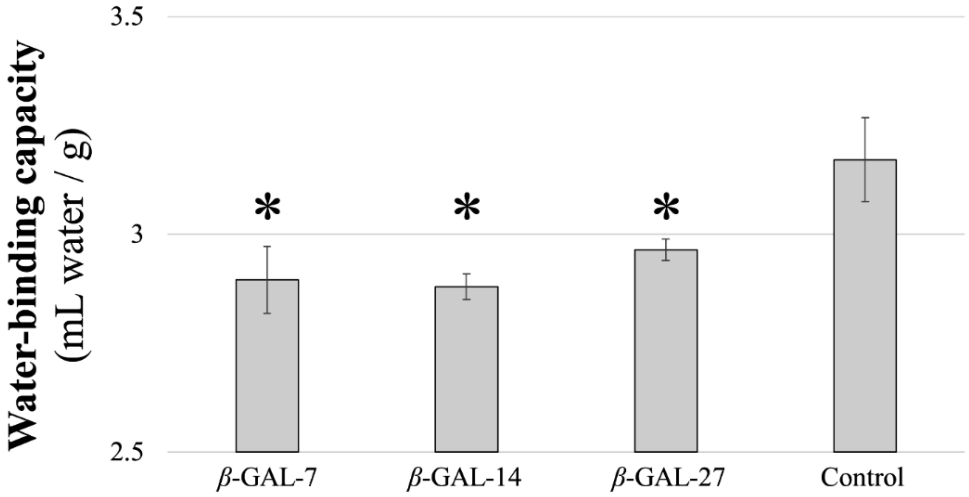

Figure 2 WBC of PCW from the tubers of the *β*-GAL lines and the wild type control (untransformed *Karnico*). Data (mean ± SD) were collected from three biological replicates (N = 3), each measured in four technical replicates (N = 4). Asterisks' denote significant differences at α = 0.05 between the *β*-GAL lines and the control, derived from one-way ANOVA post-hoc (LSD) tests. SD = standard deviation.

Cell wall properties are governed by multiple (interacting) factors, therefore it is not straight-forward to ascribe definitive functions to specific moieties. Although we observed that RG-I galactan side-chains influenced the WBC of PCW, it remains challenging to pinpoint the (relative effects of the) underlying causal factors. *In-vivo* systems are prone to indirect effects, as the plant may attempt to compensate for changes to create a functional cell wall. It has been hypothesized that *in-vivo* degradation of cell wall components may activate integrity sensing pathways or defence responses in plants, as shown for HG degradation (Ferrari et al. 2013). Moreover, targeted degradation of specific components *in-vivo* may elicit changes in non-targeted components that may indirectly affect a trait of interest. Huang et al. (2017) observed that increased pectic methyl-esterification and altered xyloglucan structures (from XXGG to a XXXG permutation) were indirect effects of *β*-galactosidase *in-vivo* in *β*-GAL-14 line. Pectin methylation affects ionic (calcium) crosslinking between pectin chains. This mechanism has been proposed to modify textural firmness properties of potato and carrot cell walls (Ross et al. 2011; Sila et al. 2006).

Changes in xyloglucan structures, as a result of *β*-galactosidase expression *in-vivo*, may also have altered these interactions that ultimately affected the WBC. Xyloglucans bind tightly to cellulose microfibrils, thereby reducing the elasticity of the cell wall (Abasolo et al. 2009). Therefore, interactions between cellulose, xyloglucan and RG-I galactan side-chains may be crucial to maintain a functional cell wall. For instance, to create sufficient space in the matrix to allow water and electrolytes to manoeuvre through (apoplastic transport). These indirect effects may influence the properties of the cell wall as suggested earlier (Cosgrove 2016). Although at this point, no direct link can be made between truncated RG-I galactan side-chains *in-vivo* and reduced WBC.

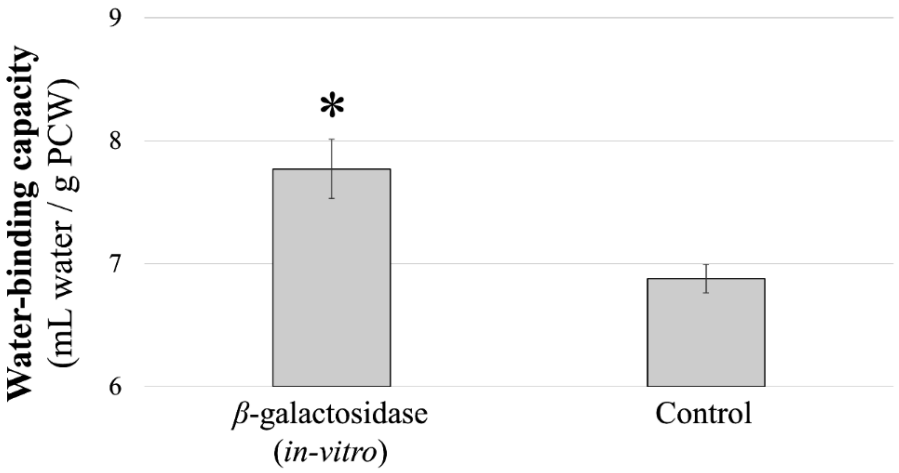

Figure 3 WBC of de-starched PCW from tubers of the wild type control (untransformed *Karnico*) after *β*-galactosidase treatment *in-vitro* compared to the control (blank treatment: no enzyme). Data (mean ± SD) were derived from three biological replicates ($N = 3$), each measured in one technical replicate ($N = 1$). The asterisk denotes a significant difference at $\alpha = 0.05$ (two-samples Student's t test) relative to the control. SD = standard deviation.

The WBC increased by 13% after *β*-galactosidase treatment *in-vitro* (**Figure 3**). This coincided with a minor release of galactose (1.98% w/w) after quantification of the monosaccharides in the supernatant phase (**Figure 4**). Physical barriers in the cell wall matrix may have limited the accessibility of

the enzyme to effectively cleave off more galactose molecules from the non-reducing ends of the galactan chains (Zykwinska et al. 2007).

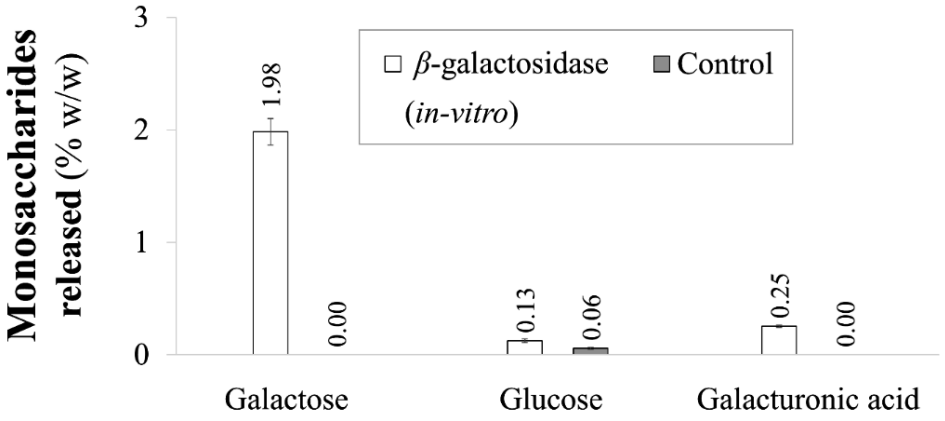

Figure 4 Monosaccharides released in the supernatant from de-starched PCW from the tubers of the wild type control (untransformed *Karnico*) after *β*-galactosidase treatment *in-vitro* compared to the control (blank treatment: no enzyme). Data (mean ± standard deviation) were derived from three biological replicates ($N = 3$) and one technical replicate ($N = 1$).

The levels of galacturonic acid, rhamnose and arabinose were reduced in PCW (residue) samples from both the enzyme treatment and the blank control (**Table 3**), when compared to PCW used as input material (**Table 2**). Numerous pectic fragments from PCW are soluble (Meyer et al. 2009; Ramasamy 2014), therefore solubilisation of arabinans and stretches of RG-I and HG may have caused these changes in our samples. A clear difference between galactose content in PCW from the enzyme treated and the control residues was not observed (**Table 3**). The relatively low release of galactose (1.98% w/w) by *β*-galactosidase *in-vitro* may underlie this observation (**Figure 4**). Both solubilisation of pectic fragments and degradation of galactan side-chains *in-vitro* may have distorted the mediation of intermolecular interactions between cell wall structures that could have resulted in a different organization of the matrix. For instance, the water-holding (and swelling) capacity of insoluble cell wall fractions from wheat flour were increased by *in-vitro* xylanase degradation (Gruppen et al. 1993).

The authors proposed that the minor degradation of xylans may have loosened cell wall structures, consequently allowing greater swelling and water retention in the cell wall matrix. Changes in the cell wall organization may affect the porosity and packing of polysaccharides that may extend the wall (Fujino and Itoh 1998; Vincken 2003). An alternative explanation to the increased WBC by *β*-galactosidase *in-vitro*, may be related to the electro-charge of the cell wall matrix. The predominant cleavage of neutral RG-I galactan side-chains may have increased the proportion of negatively-changed pectins that stimulated the formation of gelling zones in the cell wall (Sørensen et al. 2000; Willats et al. 2006). The potential increased abundance of gelling zones may have affected the WBC. Although we observed that truncated RG-I galactan side-chains *in-vitro* increased the WBC of PCW, a causal link between galactans and WBC cannot be established at this point.

Table 3 Monosaccharide composition (µg/g) of *β*-galactosidase treated PCW (*in vitro*) and control (both from the untransformed wild type)

	µg / g PCW dry basis								
PCW sample	Rha	Ara	Gal	Glc	Man	Xyl	GalA	Gal:Rha	HG:RG-I
β-galactosidase (*in-vitro*)	128 ± 21	747 ± 70	3831 ± 265	4509 ± 367	315 ± 34	499 ± 40	607 ± 133	29.9	1.9
Control	124 ± 22	751 ± 55	3891 ± 343	4415 ± 355	321 ± 14	500 ± 15	592 ± 132	31.4	1.9

Data (mean ± SD). SD = standard deviation. PCW = potato cell walls from the untransformed control (wild type). *β*-galactosidase (*in vitro*) = hydrolysed PCW residue after *β*-galactosidase (*A. niger*) treatment *in vitro*. Control = hydrolysed blank treatment (no enzyme). Rha = rhamnose; Ara = arabinose; Gal = galactose; Man = mannose; Xyl = xylose; GalA = galacturonic acid. Gal:Rha = ratio galactose to rhamnose (*i.e.* RG-I galactan side-chain length). HG:RG-I = ratio homogalacturonan (HG) to RG-I. Data were derived from three ($N = 3$) biological replicates and one ($N = 1$) technical replicate.

Conceptual model: cell wall porosity and polysaccharide spacing potentially affect WBC of PCW

Based on our results and findings from literature, we propose that the porosity and spacing between cell wall polysaccharides are important factors that modulate the WBC of PCW. RG-I galactan side-chains have been described to interact with xyloglucan and cellulose microfibrils (Carpita and Gibeaut 1993; McCann and Roberts 1991; Talbott and Ray 1992). These interactions may regulate polymer separation and porosity in the apoplastic space (Carpita

and Gibeaut 1993; Hayashi 1989; O'Neill and York 2003). As RG-I galactan side-chains are abundant in PCW, we expect that these entities may function to maintain open and well-spaced pores between charged polysaccharides that would otherwise form tight aggregated complexes that compress the cell wall matrix (**Figure 5**). In our view, RG-I galactan side-chains buffer the deleterious consequences of cell wall dehydration by inhibiting the (irreversible) adhesion of polysaccharides. These include "egg box" junctions between HG and (calcium) ions and strong hydrogen bonds between skeletal cellulose microfibrils and xyloglucan. A better understanding of cell wall polysaccharide functions and their potential interactions, will pave the way to improve the properties of cell wall by-products for high-value valorisation in food and biobased applications.

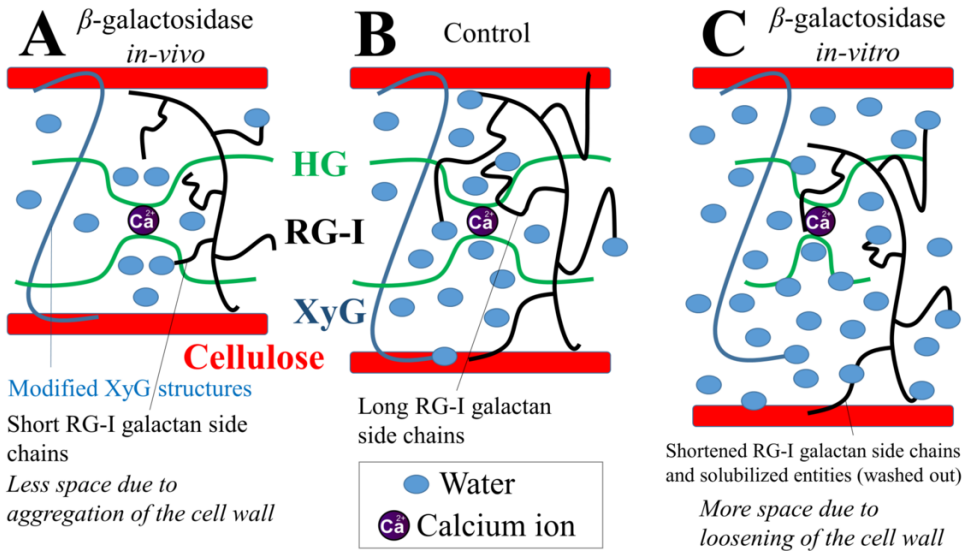

Figure 5 A simplified conceptual model for the primary cell wall matrix of PCW, showing RG-I, RG-I galactan side-chains, cellulose, xyloglucan (XyG) and homogalacturonan (HG). The presence of calcium ions induce cooperative binding of free carboxyl groups from smooth HG stretches to form hydrophilic gelling zones. (**A**) A remodelled, compressed and stiffer network due to *in-vivo* expression of *β*-galactosidase that shortened RG-I galactan side-chains (direct effect) and modified xyloglucan structures as an indirect effect. These effects most likely altered arrangements and interactions between the cell wall components, potentially reducing the WBC. (**B**) The network of the wild type control, shows longer RG-I galactan side-chains that maintain open pores that embody free water. (**C**) A loosened and more spacious network (increased porosity) due to *β*-galactosidase treatment *in-vitro* with shorter RG-I galactan side-chains and washed out (solubilized) fragments that potentially affected intermolecular interactions. Further opening of the apoplastic space may increase the WBC.

Acknowledgements

The authors thank *AVEBE* for kindly providing the milling and sieving machinery to extract the cell walls from the potato tubers. We are grateful for assistance by Nick de Vetten, Piet Buwalda, Marc Laus, Johan Krikken, Maarten Wilbrink, Annemarie Dechense, Dirk Jan Huigen, José Anton Abelenda Vila, Oene Dolstra and Yannick Schrik.

References

Abasolo W, Eder M, Yamauchi K, Obel N, Reinecke A, Neumetzler L, Dunlop JW, Mouille G, Pauly M, Höfte H (2009) Pectin may hinder the unfolding of xyloglucan chains during cell deformation: implications of the mechanical performance of Arabidopsis hypocotyls with pectin alterations. Molecular Plant 2 (5):990-999

Aldington S, Fry SC (1993) Oligosaccharins. Advances in Botanical Research 19:1-101

Baron-Epel O, Gharyal PK, Schindler M (1988) Pectins as mediators of wall porosity in soybean cells. Planta 175 (3):389-395

Belton P (1997) NMR and the mobility of water in polysaccharide gels. International Journal of Biological Macromolecules 21 (1):81-88

Braybrook SA, Peaucelle A (2013) Mechano-chemical aspects of organ formation in Arabidopsis thaliana: the relationship between auxin and pectin. PloS One 8 (3):e57813

Brett C, Hillmann J (1985) Biochemistry of Plant Cell Walls. Cambridge, UK: Cambridge University Press

Carpita NC, Gibeaut DM (1993) Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. The Plant Journal 3 (1):1-30

Cosgrove DJ (2000) Loosening of plant cell walls by expansins. Nature 407 (6802):321

Cosgrove DJ (2016) Catalysts of plant cell wall loosening. F1000 Research, 5

Dick-Perez M, Wang T, Salazar A, Zabotina OA, Hong M (2012) Multidimensional solid-state NMR studies of the structure and dynamics of pectic polysaccharides in uniformly 13C-labeled Arabidopsis primary cell walls. Magnetic Resonance in Chemistry 50 (8):539-550

Ferrari S, Savatin DV, Sicilia F, Gramegna G, Cervone F, De Lorenzo G (2013) Oligogalacturonides: plant damage-associated molecular patterns and regulators of growth and development. Frontiers in Plant Science 4:49

Fujino T, Itoh T (1998) Changes in pectin structure during epidermal cell elongation in pea (Pisum sativum) and its implications for cell wall architecture. Plant and Cell Physiology 39 (12):1315-1323

Funami T, Nakauma M, Ishihara S, Tanaka R, Inoue T, Phillips GO (2011) Structural modifications of sugar beet pectin and the relationship of structure to functionality. Food Hydrocolloids 25 (2):221-229.

Goldberg R, Morvan C, Roland JC (1986) Composition, properties and localisation of pectins in young and mature cells of the mung bean hypocotyl. Plant and Cell Physiology 27 (3):417-429

Gruppen H, Kormelink F, Voragen A (1993) Enzymic degradation of water-unextractable cell wall material and arabinoxylans from wheat flour. Journal of Cereal Science 18 (2):129-143

Ha M-A, Viëtor RJ, Jardine GD, Apperley DC, Jarvis MC (2005) Conformation and mobility of the arabinan and galactan side-chains of pectin. Phytochemistry 66 (15):1817-1824

Hayashi T (1989) Xyloglucans in the primary cell wall. Annual Review of Plant Biology 40 (1):139-168

Huang J, Kortstee A, Dees DC, Trindade LM, Schols HA, Gruppen H (2016) Modification of potato cell wall pectin by the introduction of rhamnogalacturonan lyase and β-galactosidase transgenes and their side effects. Carbohydrate Polymers 144:9-16

Huang J, Kortstee A, Dees DC, Trindade LM, Visser RG, Gruppen H, Schols HA (2017) Evaluation of both targeted and non-targeted cell wall polysaccharides in transgenic potatoes. Carbohydrate Polymers 156:312-321

Labuza TP (1968) Sorption phenomena in foods. Food Technology 22:15-19

Larsen FH, Byg I, Damager I, Diaz J, Engelsen SB, Ulvskov P (2011) Residue specific hydration of primary cell wall potato pectin identified by solid-state 13C single-pulse MAS and CP/MAS NMR spectroscopy. Biomacromolecules 12 (5):1844-1850

Lisinska G, Leszczynski W (1989) Potato starch processing. In: Lisinska, G., Leszczynski, W. (Eds.). Potato science and technology (pp. 281-346). Essex, UK: Elsevier

Makshakova ON, Faizullin DA, Mikshina PV, Gorshkova TA, Zuev YF (2018) Spatial structures of rhamnogalacturonan I in gel and colloidal solution identified by 1D and 2D-FTIR spectroscopy. Carbohydrate Polymers 192:231-239

Martín I, Dopico B, Muñoz FJ, Esteban R, Oomen RJ, Driouich A, Vincken J-P, Visser R, Labrador E (2005) In vivo expression of a Cicer arietinum β-galactosidase in potato tubers leads to a reduction of the galactan side-chains in cell wall pectin. Plant and Cell Physiology 46 (10):1613-1622

Matveev YI, Grinberg VY, Tolstoguzov VB (2000) The plasticizing effect of water on proteins, polysaccharides and their mixtures. Glassy state of biopolymers, food and seeds. Food Hydrocolloids 14 (5):425-437.

Mayer F, Hillebrandt J-O (1997) Potato pulp: microbiological characterization, physical modification, and application of this agricultural waste product. Applied Microbiology and Biotechnology 48 (4):435-440

McCann M, Roberts K (1991) The cytoskeletal basis of plant growth and form. In: C.W. Lloyd (Eds.) Architecture of the primary cell wall (pp. 109-129). London, UK: Academic press

McCann MC, Roberts K (1994) Changes in cell wall architecture during cell elongation. Journal of Experimental Botany:1683-1691

McCartney L, Ormerod AP, Gidley MJ, Knox JP (2000) Temporal and spatial regulation of pectic (1→ 4)-β-D-galactan in cell walls of developing pea cotyledons: implications for mechanical properties. The Plant Journal 22 (2):105-113

McDougall GJ, Morrison IM, Stewart D, Hillman JR (1996) Plant cell walls as dietary fibre: range, structure, processing and function. Journal of the Science of Food and Agriculture 70 (2):133-150

McNeil M, Darvill AG, Albersheim P (1980) Structure of plant cell walls: X. Rhamnogalacturonan I, a structurally complex pectic polysaccharide in the walls of suspension-cultured sycamore cells. Plant Physiology 66 (6):1128-1134

Meyer AS, Dam BP, Lærke HN (2009) Enzymatic solubilization of a pectinaceous dietary fiber fraction from potato pulp: Optimization of the fiber extraction process. Biochemical Engineering Journal 43 (1):106-112.

Moore JP, Farrant JM, Driouich A (2008) A role for pectin-associated arabinans in maintaining the flexibility of the plant cell wall during water deficit stress. Plant Signaling & Behavior 3 (2):102-104

O'Neill MA, York WS (2003) The composition and structure of plant primary cell walls. The Plant Cell Wall:1-54

Øbro J, Harholt J, Scheller HV, Orfila C (2004) Rhamnogalacturonan I in Solanum tuberosum tubers contains complex arabinogalactan structures. Phytochemistry 65 (10):1429-1438

Oomen RJFJ, Vincken, J.P., Bush, M.S., Skjot, M., Voragen, C.H.L., Ulvskov, P., et al. (2003) Towards unravelling the biological significance of individual components of pectic hairy regions in plants. In F. Voragen, H. Schols & R. Visser (Eds.) Advances in pectin and pectinase research: 2nd international symposium on pectins and pectinases (pp. 15-34). Dordrecht, The Netherlands: Kluwer Academic Publishers

Peaucelle A, Braybrook SA, Le Guillou L, Bron E, Kuhlemeier C, Höfte H (2011) Pectin-induced changes in cell wall mechanics underlie organ initiation in Arabidopsis. Current Biology 21 (20):1720-1726

Peaucelle A, Wightman R, Höfte H (2015) The control of growth symmetry breaking in the Arabidopsis hypocotyl. Current Biology 25 (13):1746-1752

Pustjens AM, de Vries S, Gerrits WJ, Kabel MA, Schols HA, Gruppen H (2012) Residual carbohydrates from in vitro digested processed rapeseed (Brassica napus) meal. Journal of Agricultural and Food Chemistry 60 (34):8257-8263

Ramasamy UR (2014) Water holding capacity and enzymatic modification of pressed potato fibres. Ph.D. dissertation. Chair group: Food Chemistry. Wageningen University. Wageningen, The Netherlands. Retrieved from http://edepot.wur.nl/304353

Ramasamy UR, Gruppen H, Kabel MA (2015) Water-holding capacity of soluble and insoluble polysaccharides in pressed potato fibre. Industrial Crops and Products 64:242-250

Ramasamy UR, Lips S, Bakker R, Gruppen H, Kabel MA (2014) Improved starch recovery from potatoes by enzymes and reduced water holding of the residual fibres. Carbohydrate Polymers 113:256-263

Ramaswamy UR, Kabel MA, Schols HA, Gruppen H (2013) Structural features and water holding capacities of pressed potato fibre polysaccharides. Carbohydrate Polymers 93 (2):589-596.

Renard C, Crépeau M-J, Thibault J-F (1994) Influence of ionic strength, pH and dielectric constant on hydration properties of native and modified fibres from sugar-beet and wheat bran. Industrial Crops and Products 3 (1-2):75-84

Ridley BL, O'Neill MA, Mohnen D (2001) Pectins: structure, biosynthesis, and oligogalacturonide-related signaling. Phytochemistry 57 (6):929-967

Roach MJ, Mokshina NY, Badhan A, Snegireva AV, Hobson N, Deyholos MK, Gorshkova TA (2011) Development of cellulosic secondary walls in flax fibers requires β-galactosidase. Plant Physiology 156 (3):1351-1363

Ross HA, Morris WL, Ducreux LJ, Hancock RD, Verrall SR, Morris JA, Tucker GA, Stewart D, Hedley PE, McDougall GJ (2011) Pectin engineering to modify product quality in potato. Plant Biotechnology Journal 9 (8):848-856

Satoh S (1998) Functions of the cell wall in the interactions of plant cells: analysis using carrot cultured cells. Plant and Cell Physiology 39 (4):361-368

Schols HA, Voragen AG (1994) Occurrence of pectic hairy regions in various plant cell wall materials and their degradability by rhamnogalacturonase. Carbohydrate Research 256 (1):83-95

Serena A, Knudsen KB (2007) Chemical and physicochemical characterisation of co-products from the vegetable food and agro industries. Animal Feed Science and Technology 139 (1):109-124

Sila DN, Doungla E, Smout C, Van Loey A, Hendrickx M (2006) Pectin fraction interconversions: insight into understanding texture evolution of thermally processed carrots. Journal of Agricultural and Food Chemistry 54 (22):8471-8479

Sluiter A, Hames B, Ruiz R, Scarlata C, Sluiter J, Templeton D, Crocker D (2008) Determination of structural carbohydrates and lignin in biomass. Laboratory Analytical Procedure 1617:1-16

Sørensen SO, Pauly M, Bush M, Skjøt M, McCann MC, Borkhardt B, Ulvskov P (2000) Pectin engineering: modification of potato pectin by in vivo expression of an endo-1, 4-β-D-galactanase. Proceedings of the National Academy of Sciences 97 (13):7639-7644

Talbott LD, Ray PM (1992) Changes in molecular size of previously deposited and newly synthesized pea cell wall matrix polysaccharides: effects of auxin and turgor. Plant Physiology 98 (1):369-379

Tang H, Belton PS, Ng A, Ryden P (1999) 13C MAS NMR studies of the effects of hydration on the cell walls of potatoes and Chinese water chestnuts. Journal of Agricultural and Food Chemistry 47 (2):510-517

Thakur BR, Singh RK, Handa AK, Rao M (1997) Chemistry and uses of pectin—a review. Critical Reviews in Food Science & Nutrition 37 (1):47-73

Thibault J, Renard C, Guillon F (2000) Sugar beet fiber: production, composition, physicochemical properties, physiological effects, safety and food applications. In: S.C.A.M. Susan, L. Dreher (Eds.). Handbook of dietary fiber (pp. 553-582). Food Science and Technology. New York, NY: Marcel Dekker

Ulvskov P, Wium H, Bruce D, Jørgensen B, Qvist KB, Skjøt M, Hepworth D, Borkhardt B, Sørensen SO (2005) Biophysical consequences of remodeling the neutral side chains of rhamnogalacturonan I in tubers of transgenic potatoes. Planta 220 (4):609-620

Vincken JP, Borkhardt B, Bush M, Doesdijk-Voragen C, Dopico B, Labrador E, Visser R (2000) Remodelling pectin structure in potato. Conference Proceedings of Phytosfere '99 European Plant Biotechnology Network, 23-34

Vincken JP, Schols, H.A., Oomen, R.J.F.J., McCann, M.C., Ulvskov, P., Voragen, A.G.J., Visser, R.G.F. (2003) If homogalacturonan were a side chain of rhamnogalacturonan I. Implications for cell wall architecture. Plant Physiology 132 (4):1781-1789

Wang T, Park YB, Cosgrove DJ, Hong M (2015) Cellulose-pectin spatial contacts are inherent to never-dried Arabidopsis primary cell walls: evidence from solid-state nuclear magnetic resonance. Plant Physiology 168 (3):871-884

Wang T, Zabotina O, Hong M (2012) Pectin–cellulose interactions in the Arabidopsis primary cell wall from two-dimensional magic-angle-spinning solid-state nuclear magnetic resonance. Biochemistry 51 (49):9846-9856

Willats WG, Knox JP, Mikkelsen JD (2006) Pectin: new insights into an old polymer are starting to gel. Trends in Food Science & Technology 17 (3):97-104

Xu X, Dees D, Dechesne A, Huang X-F, Visser RGF, Trindade LM (2017) Starch phosphorylation plays an important role in starch biosynthesis. Carbohydrate Polymers 157:1628-1637.

Zhao Q, Yuan S, Wang X, Zhang Y, Zhu H, Lu C (2008) Restoration of mature etiolated cucumber hypocotyl cell wall susceptibility to expansin by pretreatment with fungal pectinases and EGTA in vitro. Plant Physiology 147 (4):1874-1885

Zykwinska A, Thibault J-F, Ralet M-C (2007) Organization of pectic arabinan and galactan side chains in association with cellulose microfibrils in primary cell walls and related models envisaged. Journal of Experimental Botany 58 (7):1795-1802

Zykwinska AW, Ralet M-CJ, Garnier CD, Thibault J-FJ (2005) Evidence for in vitro binding of pectin side chains to cellulose. Plant Physiology 139 (1):397-407

Chapter 6

General discussion

Valorisation of protein and fibre in potato

In circular agriculture, resources (*e.g.* fertilizers and energy) are used sparingly to obtain substantial yields and product quality, whilst reducing the impact on the environment. In this light, the potato starch industry aims to valorise all compounds in potatoes to create value and reduce waste. In this dissertation, we made use of genetic and molecular approaches to valorise protein and fibre in potato. One way to contribute to this goal is to increase the level of total soluble protein in tubers and to reduce the hydrophilic nature of potato fibre (*i.e.* potato cell walls). However, these traits are modulated by multiple factors that are poorly understood. Therefore, the aim of the study described in this dissertation was to gain a better understanding of the genetics and the biological processes underlying tuber protein content and the water-binding capacity (WBC) of potato cell walls. In this chapter, the findings that were generated in our experimental research are evaluated in a broader context and implications are addressed. The last part of this chapter describes future prospects for science and industry.

Breeding for tuber protein content in potato is feasible

From a breeders perspective, the heritability of a trait is of central importance to assess the magnitude of genetic improvement (Dudley and Moll 1969). We used REML and ANOVA-based approaches to compute moderate levels of broad-sense heritability for tuber protein content in a variety panel (H^2 = 0.49) (**Chapter *2***) and in a bi-parental population (H^2 = 0.40-0.74) (**Chapter *3***). Using these approaches, that take into account within and between variance of the genotypes, we estimated values that are comparable to earlier reports based on diploid research populations (Lu et al. 2012; Werij 2011). As heritability estimates are specific for the datasets analysed, one should take care to interpret their value for breeding. Moreover, heritability values come with error and reference values do not exist. However, the heritability concept can be used in breeding when it is used appropriately by eliminating biases. To compensate for biases in practical situations, that include non-random selection of genotypes and potential environmental factors, phenotypic data from multiple locations and years should be used to reduce the effects of genotype-by-environment interactions that are commonly encountered for complex traits (Nyquist and Baker 1991). Moreover, technical error from analytical methods should be minimized to better estimate the genetic effects. For example, the level of technical error was higher for the bicinchoninic acid assay (**Chapter *2***) than for the Sprint assay (**Chapter *3***) to quantify protein content (data not shown).

When evaluating our H^2 values and those from literature, it seems plausible to assume that tuber protein content in potato will respond to selection (breeding), although concrete generalisations cannot be made as H^2 estimates always refer to the genotypes and circumstances investigated (Comstock and Moll 1963; Dudley and Moll 1969). Therefore, breeders that aim to improve tuber protein content could compare our findings by estimating H^2 in their own genepool.

Nowadays, starch potato breeders aim to improve both protein and starch content in their varieties. When analysing the phenotypic data of the variety panel – that consisted of the sub-populations "Processing", "Starch" and "Other" – we observed that tuber protein content displayed a positive correlation ($r = 0.639$, $P < 0.001$, $N = 277$) with tuber under water weight (a proxy for starch content, a highly heritable trait as reported in **Chapter *2***). This finding suggests that the underlying system(s) of these two traits (partially) complement each other. This phenomenon was also observed in the sub-population "Starch" (**Chapter *2***), where tuber protein content and starch content are key traits for breeding. As both tuber protein content and starch content are heritable and complementary, breeding tuber protein content (in combination with starch content) seems feasible.

Thoughtful experimental design will enhance the dissection of tuber protein content as a complex trait

In genetic studies, a single experimental design is frequently used to analyse multiple traits simultaneously. Although a single design provides practical advantages, it may be sub-optimal to analyse traits of which the genetic architectures are complex or when traits are only relevant for specific genepools that are not correctly represented in the design. For example, in **Chapter *3*** we analysed tuber protein content in the *Altus* × *Colomba* population, *i.e.* a diverse cross between a starch variety and consumption variety. This population has been useful to estimate heritability and to assess transgressive segregation. However, the results may not be directly applicable for breeders as the allelic contributions from the consumption variety *Colomba* cannot be exploited in the starch potato genepool where these are (mostly) absent. We characterized tuber protein content as a moderately heritable trait with a typical complex genetic architecture. Therefore, environmental effects of target environments, that include the soil properties, agronomic practices (*e.g.* fertilizer application rates) and weather conditions (*e.g.* precipitation), likely influenced the phenotypic values. Taking these factors into account in the experimental design is relevant to improve the genetic analysis and effective use of resources. Surely

replication is a crucial factor to reduce experimental variation and to get (more) robust results. For instance, in **Chapter *2*** more biological replicates and environments were included than in **Chapter *3***. This contributed to higher year-to-year correlations of the phenotypic values (*i.e.* r = 0.508-0.713 versus r = 0.300-0.361). Other important design criteria that have been suggested to increase the statistical power are related to the selection of genotypes, distributions of phenotypic values (for sub-populations), sample sizes, ascertain bias of SNPs included in marker arrays and the allele frequency of SNPs or alleles (Vos 2016; Willemsen 2018). To conclude, thoughtful experimental design will likely improve the genetic analysis to better understand the biology of tuber protein content for science and industry (see final section of this chapter for more details).

Reproducibility of marker-trait associations and missing heritability

In this study, forward genetic approaches (**Figure 1**) were used to dissect the genetic architecture of tuber protein content in tetraploid potato. In literature it is known that the accumulation of storage proteins in vegetative tissues (such as potato tubers) is influenced by the plant's source and sink capacities (*i.e.* sink strength) (Nsimba-Lubaki and Peumans 1986; Staswick 1989; van Cleve and Apel 1993; Wetzel et al. 1989). Sink strength has been suggested to be controlled by a combination of physiological, developmental and partitioning factors (**Figure 2**) (White et al. 2015). Therefore, the discovery of multiple minor-effect QTLs for tuber protein content (on chromosomes *2*, *3*, *5*, *7*, *9* and *12*) was not completely unexpected. However, the QTLs on chromosomes *3* and *5* were observed in both the variety panel (**Chapter *2***) and in the bi-parental population (**Chapter *3***). Previously, these loci have also been reported to be associated with tuber protein content in a diploid potato research population (Acharjee et al. 2018; Werij 2011). Using multiple tools, we identified two unique haplotypes that tagged the QTLs at the start of chromosome *5* (**Chapter *2***), *i.e.* a haplotype for *StCDF1.1* (early maturity) and the so-called *GPa5* haplotype (that originates from an introgression segment) that confers resistance to potato cyst nematodes. The *GPa5* haplotype also carries a late *StCDF1* maturity allele. However, these two haplotypes could not be validated in the bi-parental population that also showed a QTL for tuber protein content at the same locus. A similar situation where variants underlying QTLs could not be reproduced between different experimental datasets has been reported before (Vos 2016). For complex traits, different genetic backgrounds may affect how a trait is expressed. Follow-up studies are therefore relevant to conclude or rule out if these haplotypes truly affect tuber protein content in potato.

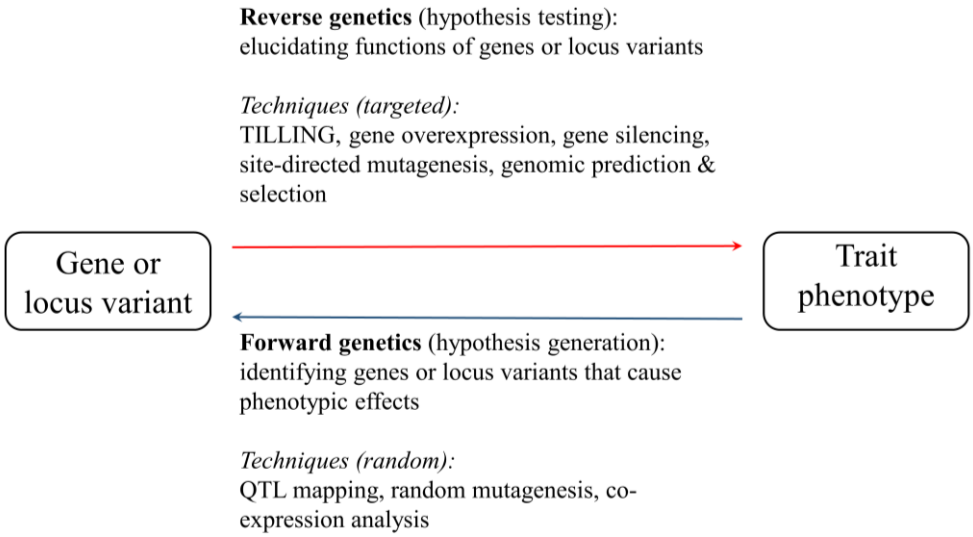

Figure 1 Forward and reverse genetics approaches that are used to analyse relationships between genes (or locus variants) and phenotypes, and vice versa. TILLING = targeted induced local lesions in genomes. QTL = quantitative trait locus.

Another observation that we made was related to missing heritability for tuber protein content, a situation where the cumulative effect of significant QTLs only explain a (minor) portion of the total phenotypic variation. It is plausible to assume that many loci with small additive and non-additive (*i.e.* epistatic) effects influence tuber protein content, but that their phenotypic contributions are too small for QTL discovery due to limited statistical power. A well-known study in yeast demonstrated that a considerable proportion of missing heritability can only be accounted for when epistatic interactions are considered (Bloom et al. 2013), suggesting that the inclusion of epistatic effects can reduce the gap of missing heritability. Similarly, a recent study demonstrated that QTLs for yield traits could only be identified in diploid potato when epistatic interactions were taken into account (Marand et al. 2019). Uncovering and understanding epistatic interactions is more complex in tetraploids than in diploids, as many more allelic combinations are possible. Other factors that may cause missing heritability include interactions between intra-locus variants from an allelic series and linkage disequilibrium (LD) of genetic markers.

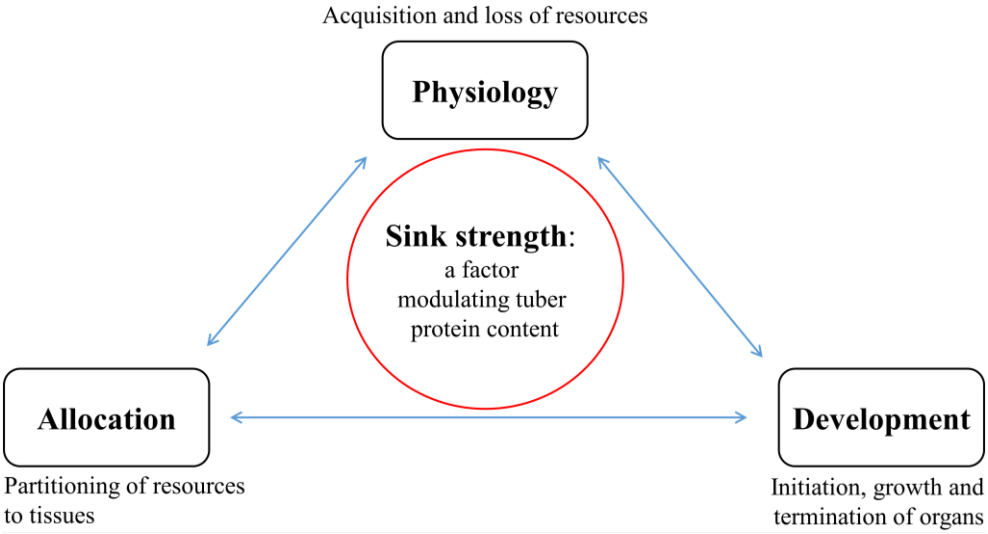

Figure 2 Interactions between physiological, developmental and resource allocation (partitioning) factors that control sink strength and (indirectly) tuber protein content in potato (adapted from White et al. (2015)).

Matching nutrient uptake, metabolism and storage to improve tuber protein content in potato

Genes involved in the central scheme for nitrogen (N) assimilation in plants (**Figure 3**) have been pinpointed as promising targets (*e.g.* nitrate and ammonium transporters) to optimize crop production and quality (Tegeder and Masclaux-Daubresse 2018). In this dissertation, the generic (35S) overexpression of a putative nitrate transporter (*StNPF1.11*), that was selected as a candidate gene underlying the discovered QTLs on chromosome *5* (**Chapter *2*** and **Chapter *3***), increased tuber protein content and leaf chlorophyll content in young potato plants (**Chapter *4***). Intriguingly, overexpression of *StNPF1.11* did not clearly increase protein content in mature tubers. This finding may have resulted from imbalances between N uptake, transport and metabolism during the phase of tuber bulking. *StNFP1.11* may not have functioned (sufficiently) during later stages of plant development or could have altered internal reserves of N metabolites (*e.g.* substrates for amino acid synthesis) that may have reduced protein biosynthesis (Thomsen et al. 2014). Tegeder and Masclaux-Daubresse (2018) stress that the uptake of nutrients, metabolism and storage of both N and carbon assimilates should be matched evenly to improve nitrogen use in plants. Selecting the 'correct' promoter for a particular trait and situation can determine the outcome of experiments. For instance, the type of

promoter used to overexpress an alanine aminotransferase was of crucial importance to induce nitrogen-use efficiency in rice (Beatty et al. 2013; Good et al. 2007; Miyashita et al. 2007; Shrawat et al. 2008). Also in pea, an increase in seed yield and seed protein content was prompted by modifying the loading and unloading of amino acids specifically in phloem tissue (Zhang et al. 2015). These findings point to the importance of using specific promoters (in combination with particular genes) to alter key biological processes at crucial stages of plant development. It is widely accepted that optimal plant growth and development relies on carbon (C) and nitrogen (N) homeostasis. Therefore, it is likely that multiple layers of metabolic steady states must be identified and engineered to enhance tuber protein content in potato. Therefore, a holistic approach that takes into account nutrient uptake, transport and storage seems promising to improve protein content in potato tubers.

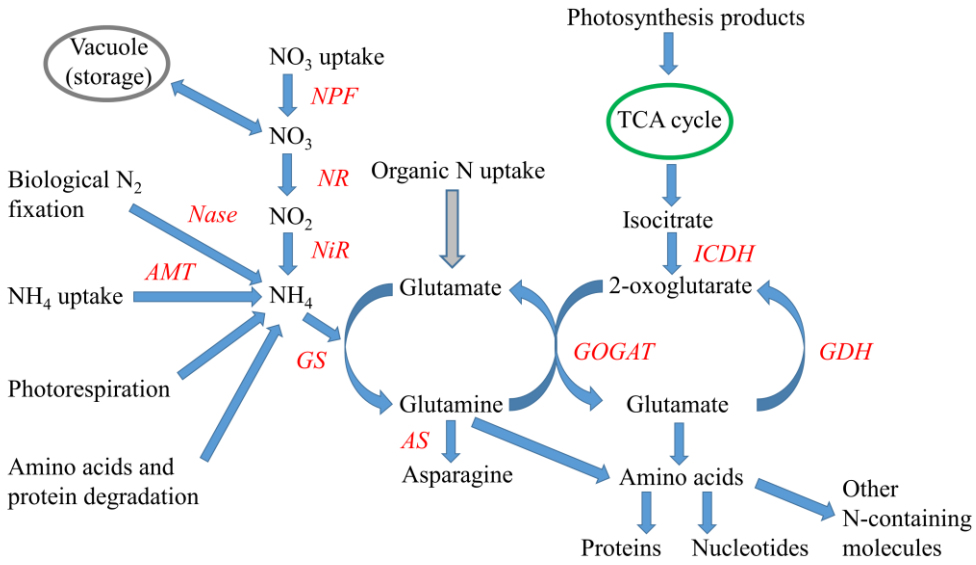

Figure 3 A simplified metabolic scheme for nitrogen (N) assimilation in higher plants adapted from Lu et al. (2016) (Lu et al. 2016). Inorganic nitrate is taken up from the soil by nitrate transporters (*NPF*) that are located at the surfaces of root cells, after which it is utilized directly or temporarily stored in vacuoles. Nitrate is subsequently converted into ammonium, after which it is incorporated (together with assimilated carbon that passes through the tricarboxylic acid (TCA) cycle in the form of 2-oxoglutarate) into glutamine and glutamate. Glutamine and glutamate serve as the basic metabolites for the biosynthesis of amino acids, proteins, nucleotides and other nitrogen-containing molecules (*e.g.* chlorophyll). Ammonium may also be taken up directly from the soil by ammonium transporters (*AMT*). Likewise, glutamate and other organic amino acids and peptides may also be taken up directly from the soil by transporters. Abbreviations that are indicated red and italics denote *NPF:* nitrate

transporters, *NR:* nitrate reductase, *NiR:* nitrite reductase, *Nase:* nitrogenase, *AMT:* ammonium transporters, *GS:* glutamine synthetase, *GOGAT:* glutamate synthase, *AS:* asparagine synthetase, *GDH:* glutamate dehydrogenase and *ICDH:* isocitrate dehydrogenase.

Influence of *StCDF1* on tuber protein content by affecting the bulking dynamics of storage compounds in potato

We identified associations between the alleles of the potato transcription factor *CYCLING DOF FACTOR1* (*StCDF1*) and tuber protein content (**Chapter *2***). In both potato and other species, these types of transcription factors have been shown to be involved in the accumulation of carbon and N storage compounds (*i.e.* sink-source dynamics). For instance, *AtCDF1* in *Arabidopsis* has been shown to function in the activation and repression of multiple genes involved in N-uptake such as nitrate transporters (**Figure 3**) (Varala et al. 2018). Moreover, overexpression of *ZmDOF1* from maize increased ammonia and elementary N content in *Arabidopsis* plants (Yanagisawa et al. 2004). These findings suggest that *DOF* factors may be involved in the modulation of storage proteins. In potato, *StCDF1* indirectly affects multiple traits and physiological processes, including the expression of the gene *SELF-PRUNING6A* (*StSP6A*) which strongly controls tuberization in potato (Navarro et al. 2011). High levels of *SP6A* maintain a high rate of tuber sink metabolism and also play an important role in sucrose allocation in tubers, *e.g.* by signalling a switch from apoplastic to symplastic transport (Viola et al. 2001). In potato tubers, the patatin protein group accounts for approximately 40% of total soluble protein (Paiva et al. 1983; Park et al. 1983; Racusen and Foote 1980). Potato patatin genes are highly expressed in tubers under high sucrose conditions (Pikkard et al. 1987; Rocha-Sosa et al. 1989). In tubers, sucrose also functions as a donor for ADP-glucose *i.e.* the substrate for starch polymer biosynthesis. As the levels of *SP6A* gradually increase during tuber development, late maturing varieties (*i.e.* starch potato varieties) may have longer and more productive phases of plant growth (under high source conditions) to maintain high accumulation rates of both patatin (protein) and starch in tubers. Taken together, it is conceivable to state that the bulking dynamics of tuber protein content in potato is indirectly affected by *StCDF1* (**Figure 2**).

Roles of RG-I galactan side-chains in cell wall development and yield

Many studies have been carried out in an attempt to define the biological functions of pectin and its components (*i.e.* structure-function relationships). This is challenging as the pectin macromolecule is structurally complex and because it interacts with other cell wall components. Therefore, clear definitions of their

biological functions in plants are lacking. However, associations between RG-I galactans and cell wall biomechanical properties have been reported that include tissue firmness and flexibility (Bush et al. 2001; Jones et al. 1997; McCartney et al. 2000; Redgwell et al. 1997; Vicré et al. 1998; Willats et al. 1999). RG-I galactan side-chains have been hypothesized to affect the hydration properties of cell walls, however very few studies have been devoted to analyse this in detail (Einhorn-Stoll 2018). We showed that *in-vivo* and *in-vitro* exo-truncation of pectic rhamnogalacturonan I (RG-I) galactan side-chains led to an altered level of water-binding capacity (WBC) of potato cell walls (PCW) (**Chapter *5***). Mechanistically, we proposed that truncation of these galactans *in-vivo* affected the porosity of the cell wall matrix due to cell wall remodelling (*e.g.* xyloglucan modification) and loosening (*i.e.* creation of voids). Findings from literature (Martín et al. 2005; Sørensen et al. 2000; Ulvskov et al. 2005), as well as our own results show that potato tubers can tolerate short (or less abundant) RG-I galactan side-chains *in-vivo*. Seemingly, an altered level of cell wall hydration *in-vivo* – that may also have affected other cell wall biomechanical properties (Ulvskov et al. 2005) – does not cause detrimental effects in potato *per se*. In literature it is known that altered biomechanical properties of cell walls can reduce plant growth (Cosgrove 1997). Intriguingly, we observed in **Chapter *5*** that both the yield and tuber dry matter content of line *β*-GAL-27 (that exhibited the highest level of *β*-galactosidase expression of all the transformants (Martín et al. 2005)) were reduced compared to the control (whereas this was not observed for the other transformants). In line *β*-GAL-27, high *β*-galactosidase activity may have negatively affected the cell wall architecture during critical stages of stolon or tuber development that caused long-lasting effects. Based on epitope studies, the abundance of galactan chains increase gradually as potato tubers increase in size (Bush et al. 2001), where these structures have been suggested to influence the size of pores in the cell wall matrix (Fenwick et al. 1999; Foster et al. 1996). Cell wall expansin-mediated polymer creep is thought to be (non-exclusively) regulated by pectinases (Cosgrove 1997). Therefore, an altered pore size due to *β*-galactosidase activity may have hampered the accessibility of enzymes or other compounds to modify the cell wall (*e.g.* expansins). Therefore, one can speculate that high *β*-galactosidase activity may influence matrix loosening that may negatively impact cell wall extensibility to correctly respond to turgor pressure (Carpita and Gibeaut 1993; McCann and Roberts 1994). Consequently, this may have hindered isodiametric growth of potato tubers in line *β*-GAL-27 where the expression of *β*-galactosidase was high. However, a causal link between the factors

cannot be established due to indirect effects that may also have played a role (Huang et al. 2017).

Future perspectives

Tuber protein content and the hydration of cell walls are both complex traits in potato. Hence, a better understanding of the factors that modulate these traits will allow new improvement strategies to be formulated. In the next section, we present prospects for future research.

Dissecting complex traits in diploid potato simplifies analysis

The use of potato diploid (inbred) lines certainly presents a promising opportunity to better study the complex (*i.e.* polygenic) nature of tuber protein content. Ample genetic analyses in potato are performed using outcrossing heterozygous tetraploid varieties. In tetraploids, alleles can segregate into five categories (0000, 0001, 0011, 0111 and 1111) whereas in diploids there are only three categories (00, 01 and 11). In the genetic analysis of diploids, the phenotypic values are divided into only three categories instead of five for tetraploids, which will increase statistical power to uncover associations. Also, the number of possible interactions between the alleles at a single locus and those between loci (epistasis) are lower in diploids in comparison to tetraploids. Because of these advantages, the dissection of complex traits in diploid potato will likely improve genetic analyses to obtain more meaningful and robust results. We foresee that epistasis will probably emerge as a relevant topic to better analyse the genetics of complex traits in potato. Inbred diploid systems provide a new system to more easily identify and fixate favourable alleles, *e.g.* by backcrossing inbred lines. Although diploids provide advantages for genetic analyses, it is still unclear whether diploid potato varieties will prevail over tetraploids in terms of yield and quality in future production systems. Advantages of tetraploids versus diploids are increased vigour (yield) and the formation of larger cells (Hutten et al. 1994). Also, a higher allelic diversity in heterozygous tetraploids may lead to a higher adaptability level than for diploids. On the other hand, the removal and fixation of favourable alleles in diploid systems can be achieved more easily. Therefore, combining the advantages of both systems may be the best approach to improve complex traits and yield in potato. For instance, by using unreduced gametes from inbred diploids (*e.g.* using alkaloid colchicine) to develop synthetic tetraploids with a lower genetic load.

Improving tuber protein content by breeding

Route 1 Recurrent selection at the phenotypic level provides a practical approach to improve polygenic traits by increasing the frequency of alleles with favourable effects (*i.e.* reduce genetic load), as demonstrated in a long-term selection experiment that dramatically increased maize kernel protein content (Dudley 2007). Tetraploid starch varieties with high levels of tuber protein content (*e.g. Altus*, *Aveka*, *Festien*, *Kartel* and *Seresta*) can be used as breeding material to carry out this procedure. In a separate breeding program, recurrent selection for tuber protein content can be performed to study complementation and trade-off effects with other (agronomic) traits (*e.g.* contents starch and glycoalkaloids). Once enhanced genotypes have been developed, these can potentially be exploited in commercial breeding to enrich crosses with favourable alleles to enhance tuber protein content.

Route 2 It is clear that marker-assisted selection (MAS) can enhance potato breeding. MAS is especially effective for improving traits that are controlled by alleles with strong effects (*e.g.* resistances against pathogens). On the other hand, tuber protein content is a complex trait that is affected by multiple loci with minor-effects. Therefore, to use MAS effectively, experimental crosses between starch varieties with high (*e.g. Seresta*) and low levels of tuber protein content (*e.g. Axion*) could be made. Minimizing factors that cause noise will be advantageous, *e.g.* by taking into account the segregation of *StCDF1* alleles to reduce the variation of maturity levels in F_1 progenies. After the effects of alleles are validated, MAS can be used to introduce or increase their frequencies in the genepool. MAS could also be used in combination with recurrent selection (see previous section) and genomic selection (see next section).

Route 3 Genomic selection (GS) has been used as a technique to improve complex traits by combining genotype, phenotype and pedigree/kinship information (if available). GS is routinely used to improve polygenic traits in other major food crops *e.g.* maize, wheat and soybean. In potato, the application of GS is in its infancy but certainly provides an alternative approach to improve tuber protein content. An important consideration point is the prediction accuracy that can be achieved. Only if the prediction accuracy is valid and high enough, will this technique provide value for potato breeders. GS may also be combined with MAS (see previous section).

Molecular approaches to improve tuber protein content

After key molecular regulators, bottlenecks and junctions in metabolic pathways have been identified, these can be targeted. Both traditional and modern biotechnology tools (*e.g.* CRISPR/Cas9 knock-out) can be used. These tools can be used to modify the level of total protein content, isoform protein groups (*e.g.* patatin) or individual proteins (*e.g.* protease inhibitor II). These approaches have already been applied for more than 30 years, for example to develop the mutant amylose-free starch potato (Hovenkamp-Hermelink et al. 1987). In future endeavours, decision makers should take care to take into account the societal perception with regard to these technologies.

RG-I galactan side-chains as a potential target to modulate cell wall properties

Pectin is a major component of potato cell walls (PCW). We showed that RG-I galactan side-chains are involved in the water-binding capacity (WBC) of PCW (**Chapter *5***), although we could not establish a direct link due to indirect effects *in-vivo* (*e.g.* xyloglucan) or *in-vitro* (*e.g.* loosening of the cell wall matrix). Follow-up studies are therefore needed to unravel the precise relationship between RG-I galactan side-chains and WBC. It is evident that short (or less abundant) RG-I galactan side-chains do not necessarily cause major negative pleiotropic effects in potato (**Chapter *5***). Cell wall analysis of several potato varieties shows that these galactans seem to display a degree of variation (Huang et al. 2017). Therefore, one may expect to observe larger phenotypic variation in potato genepools. By quantifying the levels of WBC and RG-I galactan side-chains of varieties, it may be possible to better understand how these factors relate to each other. If a robust correlation is found, breeding may be attainable. Alternatively, post-harvest modification of RG-I galactans could be carried out to lower the WBC of PCW. Physical treatment of PCW using ultrasound-assisted extraction (UAE) is a known technique to degrade pectic side-chains (Zhang et al. 2013). Also, enzymes could be used but are more costly. Ultimately, the desired properties and the costs will determine the applicability of such downstream processing techniques.

Final note: work together to create value

Improving complex traits in potato, or any other crop or species, is challenging. A multidisciplinary approach is therefore needed to improve our understanding of the system for discovering key factors that must be identified to define clear improvement strategies. Only by working together will breeders, farmers, process

technologists and scientists be able to pave new ways to effectively valorise potato protein and fibre.

References

Acharjee A, Chibon P-Y, Kloosterman B, America T, Renaut J, Maliepaard C, Visser RG (2018) Genetical genomics of quality related traits in potato tubers using proteomics. BMC Plant Biology 18:20

Beatty PH, Carroll RT, Shrawat AK, Guevara D, Good AG (2013) Physiological analysis of nitrogen-efficient rice overexpressing alanine aminotransferase under different N regimes. Botany 91:866-883

Bloom JS, Ehrenreich IM, Loo WT, Lite T-LV, Kruglyak L (2013) Finding the sources of missing heritability in a yeast cross. Nature 494:234

Bush MS, Marry M, Huxham MI, Jarvis MC, McCann MC (2001) Developmental regulation of pectic epitopes during potato tuberisation. Planta 213:869-880

Carpita NC, Gibeaut DM (1993) Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. The Plant Journal 3:1-30

Comstock R, Moll RH (1963) Genotype-environment interactions. Statistical Genetics and Plant Breeding 982:164-196

Cosgrove DJ (1997) Assembly and enlargement of the primary cell wall in plants. Annual Review of Cell and Developmental Biology 13:171-201

Dudley J (2007) From means to QTL: The Illinois long-term selection experiment as a case study in quantitative genetics. Crop Science 47:S-20-S-31

Dudley J, Moll R (1969) Interpretation and use of estimates of heritability and genetic variances in plant breeding 1. Crop Science 9:257-262

Einhorn-Stoll U (2018) Pectin-water interactions in foods – From powder to gel. Food Hydrocolloids 78:109-119

Fenwick K, Apperley D, Cosgrove DJ, Jarvis M (1999) Polymer mobility in cell walls of cucumber hypocotyls. Phytochemistry 51:17-22

Foster TJ, Ablett S, McCann MC, Gidley MJ (1996) Mobility-resolved 13C-NMR spectroscopy of primary plant cell walls. Biopolymers 39:51-66

Good AG, Johnson SJ, De Pauw M, Carroll RT, Savidov N, Vidmar J, Lu Z, Taylor G, Stroeher V (2007) Engineering nitrogen use efficiency with alanine aminotransferase. Botany 85:252-262

Hovenkamp-Hermelink J, Jacobsen E, Ponstein A, Visser R, Vos-Scheperkeuter G, Bijmolt E, De Vries J, Witholt B, Feenstra W (1987) Isolation of an amylose-free starch mutant of the potato (Solanum tuberosum L.). Theoretical and Applied Genetics 75:217-221

Huang J, Kortstee A, Dees DC, Trindade LM, Visser RG, Gruppen H, Schols HA (2017) Evaluation of both targeted and non-targeted cell wall polysaccharides in transgenic potatoes. Carbohydrate Polymers 156:312-321

Hutten RCB, Schippers MGM, Hermsen JGT, Ramanna MS (1994) Comparative performance of FDR and SDR progenies from reciprocal 4x-2x crosses in potato. Theoretical and Applied Genetics 89:545-550

Jones L, Seymour GB, Knox JP (1997) Localization of pectic galactan in tomato cell walls using a monoclonal antibody specific to (1 [->] 4)-[beta]-D-galactan. Plant Physiology 113:1405-1412

Lu J, Zhang L, Lewis RS, Bovet L, Goepfert S, Jack AM, Crutchfield JD, Ji H, Dewey RE (2016) Expression of a constitutively active nitrate reductase variant in tobacco reduces tobacco-specific nitrosamine accumulation in cured leaves and cigarette smoke. Plant Biotechnology Journal 14:1500-1510

Lu W, Yu M, Bai Y, Li W, Xu X (2012) Crude Protein Content in Diploid Hybrid Potato Clones of Solanum phureja–S. stenotomum. Potato Research 55:315-322

Marand AP, Jansky SH, Gage JL, Hamernik AJ, de Leon N, Jiang J (2019) Residual Heterozygosity and Epistatic Interactions Underlie the Complex Genetic Architecture of Yield in Diploid Potato. Genetics 212:317-332

Martín I, Dopico B, Muñoz FJ, Esteban R, Oomen RJ, Driouich A, Vincken J-P, Visser R, Labrador E (2005) In vivo expression of a Cicer arietinum β-galactosidase in potato tubers leads to a reduction of the galactan side-chains in cell wall pectin. Plant and Cell Physiology 46:1613-1622

McCann MC, Roberts K (1994) Changes in cell wall architecture during cell elongation. Journal of Experimental Botany:1683-1691

McCartney L, Ormerod AP, Gidley MJ, Knox JP (2000) Temporal and spatial regulation of pectic (1→ 4)-β-D-galactan in cell walls of developing pea cotyledons: implications for mechanical properties. The Plant Journal 22:105-113

Miyashita Y, Dolferus R, Ismond KP, Good AG (2007) Alanine aminotransferase catalyses the breakdown of alanine after hypoxia in Arabidopsis thaliana. The Plant Journal 49:1108-1121

Navarro C, Abelenda JA, Cruz-Oró E, Cuéllar CA, Tamaki S, Silva J, Shimamoto K, Prat S (2011) Control of flowering and storage organ formation in potato by FLOWERING LOCUS T. Nature 478:119

Nsimba-Lubaki M, Peumans WJ (1986) Seasonal fluctuations of lectins in barks of elderberry (Sambucus nigra) and black locust (Robinia pseudoacacia). Plant Physiology 80:747-751

Nyquist WE, Baker R (1991) Estimation of heritability and prediction of selection response in plant populations. Critical Reviews in Plant Sciences 10:235-322

Paiva E, Lister RM, Park WD (1983) Induction and accumulation of major tuber proteins of potato in stems and petioles. Plant Physiology 71:161-168

Park WD, Blackwood C, Mignery GA, Hermodson MA, Lister RM (1983) Analysis of the heterogeneity of the 40,000 molecular weight tuber glycoprotein of potatoes by immunological methods and by NH2-terminal sequence analysis. Plant Physiology 71:156-160

Pikkard CS, Brusca JS, Hannapel DJ, Park WD (1987) The tow classes of genes for the major potato tuber protein, patatin, are differentially expressed in tubers and roots. Nucleic Acids Research 15:1979-1994

Racusen D, Foote M (1980) A major soluble glycoprotein of potato tubers. Journal of Food Biochemistry 4:43-52

Redgwell RJ, MacRae E, Hallett I, Fischer M, Perry J, Harker R (1997) In vivo and in vitro swelling of cell walls during fruit ripening. Planta 203:162-173

Rocha-Sosa M, Sonnewald U, Frommer W, Stratmann M, Schell J, Willmitzer L (1989) Both developmental and metabolic signals activate the promoter of a class I patatin gene. The EMBO Journal 8:23

Shrawat AK, Carroll RT, DePauw M, Taylor GJ, Good AG (2008) Genetic engineering of improved nitrogen use efficiency in rice by the tissue-specific expression of alanine aminotransferase. Plant Biotechnology Journal 6:722-732

Sørensen SO, Pauly M, Bush M, Skjøt M, McCann MC, Borkhardt B, Ulvskov P (2000) Pectin engineering: modification of potato pectin by in vivo expression of an endo-1, 4-β-D-galactanase. Proceedings of the National Academy of Sciences 97:7639-7644

Staswick PE (1989) Developmental regulation and the influence of plant sinks on vegetative storage protein gene expression in soybean leaves. Plant Physiology 89:309-315

Tegeder M, Masclaux-Daubresse C (2018) Source and sink mechanisms of nitrogen transport and use. New Phytologist 217:35-53

Thomsen HC, Eriksson D, Møller IS, Schjoerring JK (2014) Cytosolic glutamine synthetase: a target for improvement of crop nitrogen use efficiency? Trends in Plant Science 19:656-663

Ulvskov P, Wium H, Bruce D, Jørgensen B, Qvist KB, Skjøt M, Hepworth D, Borkhardt B, Sørensen SO (2005) Biophysical consequences of remodeling the neutral side chains of rhamnogalacturonan I in tubers of transgenic potatoes. Planta 220:609-620

van Cleve B, Apel K (1993) Induction by nitrogen and low temperature of storage-protein synthesis in poplar trees exposed to long days. Planta 189:157-160

Varala K, Marshall-Colón A, Cirrone J, Brooks MD, Pasquino AV, Léran S, Mittal S, Rock TM, Edwards MB, Kim GJ (2018) Temporal transcriptional logic of dynamic regulatory networks underlying nitrogen signaling and use in plants. Proceedings of the National Academy of Sciences 115:6494-6499

Vicré M, Jauneau A, Knox JP, Driouich A (1998) Immunolocalization of β-(1→ 4) and β-(1→ 6)-D-galactan epitopes in the cell wall and Golgi stacks of developing flax root tissues. Protoplasma 203:26-34

Viola R, Roberts AG, Haupt S, Gazzani S, Hancock RD, Marmiroli N, Machray GC, Oparka KJ (2001) Tuberization in potato involves a switch from apoplastic to symplastic phloem unloading. The Plant Cell 13:385-398

Vos P (2016) Development and application of a 20K SNP array in potato. PhD dissertation. Chair group: Plant breeding. Wageningen University, Wageningen, The Netherlands. Retrieved from http://edepot.wur.nl/392278

Werij JS (2011) Genetic analysis of potato tuber quality traits. PhD dissertation. Chair Group: Plant breeding. Wageningen University, Wageningen, The Netherlands. Retrieved from http://edepot.wur.nl/183746

Wetzel S, Demmers C, Greenwood J (1989) Seasonally fluctuating bark proteins are a potential form of nitrogen storage in three temperate hardwoods. Planta 178:275-281

White AC, Rogers A, Rees M, Osborne CP (2015) How can we make plants grow faster? A source–sink perspective on growth rate. Journal of Experimental Botany 67:31-45

Willats WG, Steele-King CG, Marcus SE, Knox JP (1999) Side chains of pectic polysaccharides are regulated in relation to cell proliferation and cell differentiation. The Plant Journal 20:619-628

Willemsen JH (2018) The identification of allelic variation in potato. PhD dissertation. Chair group: Plant breeding. Wageningen University, Wageningen, The Netherlands

Yanagisawa S, Akiyama A, Kisaka H, Uchimiya H, Miwa T (2004) Metabolic engineering with Dof1 transcription factor in plants: improved nitrogen assimilation and growth under low-nitrogen conditions. Proceedings of the National Academy of Sciences 101:7833-7838

Zhang L, Garneau MG, Majumdar R, Grant J, Tegeder M (2015) Improvement of pea biomass and seed productivity by simultaneous increase of phloem and embryo loading with amino acids. The Plant Journal 81:134-146

Zhang L, Ye X, Ding T, Sun X, Xu Y, Liu D (2013) Ultrasound effects on the degradation kinetics, structure and rheological properties of apple pectin. Ultrasonics Sonochemistry 20:222-231

Summary

On a global scale, potato is an important crop for both consumption and the starch industry. At present, the starch industry is keen to create economic value and to reduce its impact on the environment by valorising all resources in potatoes. One way to contribute towards this goal is to develop new potato varieties with elite traits, *i.e.* a high level of tuber protein content and a reduced level of fibre hydration. Improving these traits is however challenging due to their complex nature. Therefore, a better understanding of the genetics and biological processes underlying tuber protein content and fibre hydration are relevant. The objectives of this study were to shed light on the genetic and molecular architectures of tuber protein content and fibre hydration and to pinpoint key factors (*i.e.* biological processes, molecular structures, genes and alleles) that are involved in modulation of these traits.

Chapter *1* provides a description of the potato crop and the starch industry. We present protein and fibre as key resources, where improvement of their quantity and quality will lead to a better valorisation of the potato crop. The principles of potato genetics, potato breeding and modern tools for trait improvement are described.

In **Chapter *2*** we studied the genetics of tuber protein content in a panel of tetraploid potato varieties. We estimated a moderate level of trait heritability, identified marker-trait associations (QTLs), haplotypes and candidate genes. Our findings showed that alleles of *StCDF1* were associated with tuber protein content. The results provide resources for genomics-enabled breeding.

In **Chapter *3*** we performed a multi-allelic QTL analysis of tuber protein content in a large bi-parental population of tetraploid potato. We estimated a moderate level of trait heritability and identified QTLs. The alleles underlying the QTLs provided both positive and negative effects on the level of tuber protein content. Our results showed that tuber protein content is a complex trait in potato.

In **Chapter *4*** we studied the overexpression effect of a putative nitrate transporter gene (*StNPF1.11*) on tuber protein content in potato.

Overexpression of *StNPF1.11* increased tuber protein content, leaf chlorophyll content and plant height of young potato plants. A pleiotropic effect on tuber dry matter content (a proxy for starch content), suggests that the nitrogen status may affect tuber starch accumulation in potato *in vivo*.

In **Chapter *5*** we studied the role of pectic rhamnogalacturonan (RG-I) galactan side-chains on the water-binding capacity (WBC) of potato cell walls. Both *in-vivo* and *in-vitro* truncation of RG-I β-(1→4)-D-galactan side-chains altered the WBC, but with contrasting effects. Our results reinforce the view that RG-I galactan side-chains play a role in modulating the WBC of potato cell walls.

In **Chapter *6*** the insights that were generated in the experimental chapters are evaluated and discussed in a broader context. Finally, implications and prospects for future research are presented.

Acknowledgements

After six years I have completed my PhD journey. Looking back, I have enjoyed many different moments, from planting potatoes to publishing papers. It is needless to say that contributions of others have facilitated me to complete my dissertation. In this section I want to sincerely thank all whom have aided me. In case I have omitted anyone's one, my apologies.

At Wageningen University, I firstly want to thank my first promotor Luisa Trindade. We have known each other since 2010, where you supervised my master thesis. Thank you for your scientific and professional guidance, and your diligent support. One of the many things you learnt me is how to publish research work in scientific journals. Last year you were promoted to full professor. This was a memorable moment where I celebrated that you were one of the women to break grounds in the mostly male dominated science world. You make a great example for more to follow. I am also very grateful to my second promotor Richard Visser. Thank you for the opportunity to start a PhD at your chair group. You have always been supportive and your feedback was constructive. I appreciated our formal and informal contact moments and your presence at my PhD consortium meetings, in spite of your busy schedule. Chris Maliepaard, I want to thank you for your valuable role as co-promotor. Thank you for your fine expertise with regard to the statistics. You were always available for ad hoc meetings in which you provided sharp answers. Thank you Herman van Eck for your contributions and sharing your thorough knowledge on potato genetics. It was also nice to talk about informal matters that took place in Wageningen. Surely we will see each other more often in the future. Harry Gruppen, unfortunately due to a sudden illness, you could no longer fulfil the role as external supervisor. Your clear advice was helpful for making decisions in the first phase of my PhD, for this I am thankful. I hope that you will be able to receive my dissertation.

At Averis Seeds I want to thank Nick de Vetten, Johan Hopman, Hilde Room, Stan Oome and Hellen Lensing for our talks, facilitating excursions and for providing planting material and data. Nick, thank you for the many tips and

advice. At AVEBE I want to thank Piet Buwalda, Marc Laus, Joan Schipper, Johan Krikken, Maarten Wilbrink and Robin Spelbrink for allowing me work in your laboratories, sharing equipment and facilitating excursions at the AVEBE factories. Jerome Vlaanderen, thanks for the pleasant evenings and having me stay over at your place in Groningen.

I want to thank my friends from Plant Breeding: Johan Willemsen, Peter Bourke, Thijs van Dijk and Geert van Geest. We shared many special moments and had a great time together. One of the most splendid experiences was our trip to San Diego. You guys made my PhD journey an enjoyable one. I wish you all the best for the future. I also want to thank Peter Vos, Abe Abelenda Vila, Giorgio Tumino and Arwa Shanin for your company and contributions. I enjoyed spending time with you guys during our breaks. Letty Dijker and Nicole Trefflich, thanks for your administrative support and your swift moves to book the most popular moment during the week for my PhD defence ceremony (Friday at 4 p.m.). Thank you Maria João Paulo, Anne Kortstee, Oene Dolstra, Christian Bachem, Robert van Loo, Ronald Hutten, Gerard van der Linden, Henk Schouten, Rene Smulders and Paul Arens and for our discussions and feedback. I want to express my thankfulness to Dianka Dees, Annemarie Dechesne, Christel Denneboom and Dirk Jan Huigen for your excellent laboratory and greenhouse skills. Henk Meurs, René Alles, Eddy de Boer, Gerard Derks, Wim van der Slikke and Dafydd Timmermans (all from Unifarm), thank you for helping out to conduct the different growth trials and logistics to process the potato samples. I also want to acknowledge the contributions of the master students that I co-supervised: Kiflom Aregawi, Remco Overeem, Pablo Baldeón Mendoza, Jorge Alemán Báez, Michiel Schrijen and Rommel Garrido Jr. Also thanks to my (next door) officemates in Wageningen for our talks: Manos Domazakis, Carolina Aguilera Galvez, Baye Berhihun Getahun, Ehsan Motazedi, Charlotte Prodhomme, Tim van der Weijde, Ernest Aliche, Jordi Petit-Pedro, Andres Torres, Dennis van Muijen, Jarst van Belle, Pauline van Haperen, Aviv Andriani, Yerisf Torres Ascurra, Yinlei Wang, Francesco Pancaldi, Matthijs Brouwer, Eliana Papoutsoglou, Xuan Xu, Behzad Rashidi and Cynara Romero.

At Aeres University of Applied Sciences I want to start to thank Wim van de Weg for allowing me to pursue a PhD within the CBBE program. To me you have been an inspirational and visionary team leader. Corné Kocks, thank you for your support and trust in me. I appreciate your entrepreneurial drive, creativity and people skills to get things done. Of course, I also have to express my gratitude to my former colleagues Han van Kasteren and Annemarie van Leeuwen. We experienced a fantastic time together in generating new concepts for the valorisation of biological resources. Yannick Schrik, thank you for your help. I can clearly recall the moment that we milled potatoes during the winter in Wageningen. Wiggele Oosterhoff, Corne Kempenaar, Peter Kooman and Matthijs Breel, thanks for your reflection and talks which I appreciated. Also thanks to my officemates in Dronten: Mandy van Vugt, Ayella Spaapen, Piet Moonen, Ferry Tigchelhoff, Janne van den Akker and Teatske Pol for a pleasant working environment and the nice moments at the Bakker Bart shop in Dronten. Mariska van Asselt, Hans Corten and Ron Methorst, thanks for the tips and tricks on how to complete a part-time PhD. My carpool mates: Wander Postma, Arjan Ruijssenaars, Corry Uenk, Geronda Klop, Ineke van Meggelen and Annet Velthuis, you surely made the trips to Dronten more enjoyable. Thank you Gerlinde van Vilsteren for your administrative support from the CBBE office.

Thanks to my friends and paranymphs Bart Waalewijn and Hiskia Gerestein for your relaxed pose and perfect articulation of my propositions at the ceremony. We know each other for a very long time, it means a lot to me that you could fulfil this role. I also want to thank my PhD thesis opponents Daniela Bustos-Korts, Jean-Paul Vincken, Johan Sanders and Bas Zwaan for your contributions.

Marleen & Hilko van Hell, thank you very much for your hospitality and that we could dine together with our families and Salem friends every Wednesday evening in Wageningen. Seerp and Maaike Wigboldus, your company was great and I enjoyed our many discussions. Also Jouke and Arianne Spriensma, you guys are great. Thanks to Steffi & Richard Fuchs and Jerke de Vries for the numerous great coffee and cake moments. Thanks to my parents (Willem

Klaassen and Ank Klaassen-Van Marwijk) for always showing your interest in my PhD project. You never stopped asking what I was working on and when I would finish. Also, thanks to my parents in law (Kees Randewijk and Annegina Randewijk-Post) for showing me how to be bold. Renée de Vink and Nina Klaassen, thank you for making the artful potato painting for the cover of this dissertation. You did a great job, I love it!

Finally I want to express deep gratefulness to my wife Annegina and my children Nina, David and Mika. I am blessed to be part of a warm family with you guys. Annegina, thank you for your never ending support and patience. You contributed a lot of time and freedom to take care of our children so that I could go on and complete my PhD. There are no words to thank you for this. Make sure to compensate for this in the future, take a trip soon. Nina, David and Mika, let's go skateboarding, swimming and camping with a big fire more often now.

Michiel

January 2020, Wageningen, The Netherlands

About the author

Michiel Thabiso Klaassen was born on 18 December 1984 in Morija (Lesotho). He attended high school at *Waterford Kamhlaba United World Colleges* (UWC) in Mbabane (Kingdom of eSwatini, formerly known as Swaziland). Because he had to move back to The Netherlands with his parents, he did not graduate from high school. He continued his studies in Electrical Engineering at *ROC ASA* in Utrecht and graduated *cum laude.* During his studies, he worked (part-time) as an engineer to design and manufacture food processing equipment at *Rademaker* in Culemborg. He went on to complete his propaedeutic in Agricultural Systems Management at *Van Hall Larenstein University of Applied Sciences* in Deventer. He continued to study Plant Sciences at *Wageningen University* and married his love Annegina. After graduating he worked as a business consultant on sustainable agriculture and the biobased economy at the firm *Schuttelaar & Partners* in Wageningen and The Hague. Michiel then moved on to work as a researcher and teacher at *Aeres University of Applied Sciences* in Dronten. Here he received the opportunity to carry out a part-time PhD trajectory in collaboration with the *Laboratory of Plant Breeding* of *Wageningen University*, the *Centre for Biobased Economy* (CBBE), *AVEBE* and *Averis Seeds*. From the eight candidates that entered the CBBE PhD program, Michiel was the only individual to graduate. After completing his PhD trajectory, Michiel continued his career at *Aeres University of Applied Sciences*. Together with Annegina and his children (Nina, David & Mika), he enjoys to travel and discover the world.

List of publications

Michiel T. Klaassen, Peter M. Bourke, Chris Maliepaard, Luisa M. Trindade (2019). Multi-allelic QTL analysis of protein content in a bi-parental population of cultivated tetraploid potato. *Euphytica* 215 (2):14. DOI: 10.1007/s10681-018-2331-z

Michiel T. Klaassen, Johan H. Willemsen, Peter G. Vos, Richard G.F. Visser, Herman J. van Eck, Chris Maliepaard, Luisa M. Trindade (2019). Genome-wide association analysis in tetraploid potato reveals four QTLs for protein content. *Molecular Breeding* 39 (11):151. DOI: 10.1007/s11032-019-1070-8

Michiel T. Klaassen, Luisa M. Trindade (2020). RG-I galactan side-chains are involved in the regulation of the water-binding capacity of potato cell walls. *Carbohydrate Polymers* 227:115353. DOI: 10.1016/j.carbpol.2019.115353

Michiel T. Klaassen, Dianka C.T. Dees, Rommel M. Garrido Jr., Jorge Alemán Báez, Michiel Schrijen, Pablo G. Baldeón Mendoza, Luisa M. Trindade. Overexpression of a putative nitrate transporter (*StNPF1.11*) increases plant height, leaf chlorophyll content and tuber protein content of young potato plants. *Functional Plant Biology (Accepted)*

Education Statement of the Graduate School
Experimental Plant Sciences

The Graduate School EXPERIMENTAL PLANT SCIENCES

Issued to: Michiel T. Klaassen
Date: 31 January 2020
Group: Laboratory of Plant Breeding
University: Wageningen University & Research

1) Start-Up Phase	date	cp
► **First presentation of your project**		
Title: Valorisation of potato compounds	09 Sep 2014	1.5
► **Writing or rewriting a project proposal**		
Title: Valorisation of potato compounds	01 Apr 2014	6.0
► **Writing a review or book chapter**		
► **MSc courses**		
Subtotal Start-Up Phase		7.5

2) Scientific Exposure	date	cp
► **EPS PhD student days**		
EPS PhD Student Days, Get2Gether 2016, Soest, NL	28-29 Jan 2016	0.6
EPS PhD Student Days, Get2Gether 2017, Soest, NL	9-10 Feb 2017	0.6
► **EPS theme symposia**		
EPS Theme 3 Symposium 'Metabolism and Adaptation', Utrecht, NL	10 Feb 2015	0.3
EPS Theme 4 Symposium 'Genome Biology', Amsterdam, NL	15 Dec 2015	0.3
EPS Theme 1 Symposium 'Developmental Biology of Plants', Wageningen, NL	21 Jan 2016	0.3
EPS Theme 4 Symposium 'Genome Biology', Wageningen, NL	16 Dec 2016	0.3
EPS Theme 3 Symposium 'Metabolism and Adaptation', Wageningen, NL	14 Mar 2017	0.3
EPS Theme 1 Symposium 'Developmental Biology of Plants', Wageningen, NL	30 Jan 2018	0.3
► **Lunteren Days and other national platforms**		
Annual Meeting 'Experimental Plant Sciences', Lunteren, NL	14-15 Apr 2014	0.6
Annual Meeting 'Experimental Plant Sciences', Lunteren, NL	13-14 Apr 2015	0.6
Annual Meeting 'Experimental Plant Sciences', Lunteren, NL	11-12 Apr 2016	0.6
Annual Meeting 'Experimental Plant Sciences', Lunteren, NL	11-12 Apr 2017	0.6
Annual Meeting 'Experimental Plant Sciences', Lunteren, NL	9-10 Apr 2018	0.6
PhD Consortium Meeting, Wageningen, NL	11 Jun 2014	0.2
PhD Consortium Meeting, Valthermond, NL	22 Sep 2014	0.2
PhD Consortium Meeting, Wageningen, NL	03 Dec 2014	0.2
PhD Consortium Meeting, Valthermond, NL	23 Apr 2015	0.2
PhD Consortium Meeting, Wageningen, NL	02 Oct 2015	0.2
PhD Consortium Meeting, Ter Apelkanaal, NL	15 Mar 2016	0.2
PhD Consortium Meeting, Wageningen, NL	21 Sep 2016	0.2
► **Seminars (series), workshops and symposia**		
Plant Breeding Research Day, Wageningen, NL	30 Sep 2014	0.3
Plant Breeding Research Day, Wageningen, NL	29 Sep 2015	0.3
Symposium: 'Omics Advances for Academia & Industry', Wageningen, NL	11 Dec 2014	0.3
Symposium: 'From Big Data to Biological Solutions', Wageningen, NL	18 Jun 2015	0.3
Symposium: WURomics 'Technology-driven Innovation for Plant Breeding', Wageningen, NL	15 Dec 2016	0.3
► **Seminar plus**		
► **International symposia and congresses**		
International Plant & Animal Genome Conference XXVI, San Diego, CA, USA	13-17 Jan 2018	1.5
International Cell Wall Research Conference VIII, Monterey, CA, USA	18-22 Jun 2018	1.3
► **Presentations**		
Oral		
First PhD project presentation at PhD Consortium Meeting, Valthermond, NL	05 Mar 2014	1.0
Final PhD project presentation at PhD Consortium Meeting, Ter Apelkanaal, NL	20 Nov 2017	1.0
International Cell Wall Research Conference VIII, Monterey, CA, USA	19 Jun 2018	1.0
Poster		
Annual meeting 'Experimental Plant Sciences', Lunteren, NL	9-10 Apr 2018	1.0
► **IAB interview**		
► **Excursions**		
Subtotal Scientific Exposure		15.7

3) In-Depth Studies	date	cp
► **Advanced scientific courses & workshops**		
Microscopy and Spectroscopy in Food and Plant Sciences, Wageningen, NL	6-9 May 2014	0.6
Polyploids QTL Mapping Training I, Wageningen, NL	13 Jun 2014	0.3
Polyploids QTL Mapping Training II, Wageningen, NL	24 Jun 2015	0.3
Polyploids QTL Mapping Training III, Wageningen, NL	12-13 Dec 2016	0.5
QTL Mapping within the Genstat Environment, Wageningen, NL	08 Sep 2015	0.2
► **Journal club**		
Participation in Journal Club 'Quantitative Genetics', Plant Breeding, Wageningen, NL	2015-2018	3.0
► **Individual research training**		
Subtotal In-Depth Studies		4.9

CONTINUED ON NEXT PAGE

4) Personal Development	*date*	*cp*
► **General skill training courses**		
Competence Assessment, Wageningen, NL	19 Nov 2014	0.3
Scientific Writing, Wageningen, NL	13 Sep - 15 Nov 2016	1.8
ICM Time Management, Utrecht, NL	17 Feb 2017	0.6
Efficient Writing Strategies, Wageningen, NL	Apr - Jun 2017	1.3
► **Organisation of meetings, PhD courses or outreach activities**		
► **Membership of EPS PhD Council**		
Subtotal Personal Development		4.0

TOTAL NUMBER OF CREDIT POINTS*	**32.1**

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

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

The research described in this thesis was financially supported by *Aeres University of Applied Sciences*, the *Centre for Biobased Economy* (CBBE), *AVEBE* and *Averis Seeds*. These funds are gratefully acknowledged.

Financial support for printing this thesis from the *Laboratory of Plant Breeding* of *Wageningen University*, *Aeres University of Applied Sciences* and the *Centre for Biobased Economy* (CBBE) is gratefully acknowledged.

Colophon

Cover painting and design by *Nina Klaassen* and *Renée de Vink*.

Thesis layout by *Michiel T. Klaassen*.

Printed by *GVO Printers & Designers* (Ede) on FSC-certified paper.

Propositions

1. To enhance protein content in potato, the rates of nitrogen uptake, biosynthesis of amino acids and accumulation of storage proteins in tuber vacuoles should be increased simultaneously.
   (this thesis)

2. Dissecting the complex genetics of protein content in potato will be more effective using inbred diploids than heterozygous tetraploids.
   (this thesis)

3. To curb conflicts between studies, conclusions should not rest on dichotomization based on significance thresholds alone.

4. Only with a better understanding of the scientific process will journalists be able to convey findings from science correctly.

5. Without knowledge on consumer trends and behaviour, key governmental decision makers will not be able to effectively influence the consumption of plant proteins in society.

6. Economic growth in the Netherlands can be deduced from the level of traffic congestion on the A1 automotive highway from Wageningen to Dronten.

Propositions belonging to the thesis:

Genetic and molecular approaches to valorise protein and fibre in potato

Michiel T. Klaassen

Wageningen, 31 January 2020

INVITATION

You are cordially invited to attend the public defence of my thesis entitled:

GENETIC AND MOLECULAR APPROACHES TO VALORISE PROTEIN AND FIBRE IN POTATO

FRI 31/01/2020
AT 16.00 PM

In the Aula of
Wageningen University
(Generaal Foulkesweg 1A,
Wageningen)

Michiel Klaassen
mtklaassen@gmail.com

Paranymphs:

• Bart Waalewijn
bartwaalewijn@gmail.com

• Hiskia Gerestein
hiskiagerestein@hotmail.com