

QTL discovery for agronomic and quality traits in diploid potato clones using PotatoMASH amplicon sequencing

Lea Vexler ^{1,2,3} Maria de la O Leyva-Perez ¹ Agnieszka Konkolewska ¹ Corentin R. Clot ^{2,3} Stephen Byrne ¹ Denis Griffin ¹ Tom Ruttink ^{4,5} Ronald C.B. Hutten,² Christel Engelen,² Richard G.F. Visser ² Vanessa Prigge,⁶ Silke Wagener,⁶ Gisele Lairy-Joly,⁷ Jan-David Driesprong,⁸ Ea Høegh Riis Sundmark,⁹ A. Nico O. Rookmaker,¹⁰ Herman J. van Eck ^{2,3,*} Dan Milbourne ^{1,*}

¹Teagasc, Crop Science Department, Oak Park, Carlow R93 XE12, Ireland

²Plant Breeding, Wageningen University & Research, P.O. Box 386, Wageningen 6700 AJ, The Netherlands

³The Graduate School Experimental Plant Sciences, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands

⁴Flanders Research Institute for Agriculture, Fisheries and Food (ILVO), Plant Sciences Unit, Caritasstraat 39, Melle 9090, Belgium

⁵Department of Plant Biotechnology and Bioinformatics, Faculty of Sciences, Ghent University, Technologiepark 71, Ghent 9052, Belgium

⁶SaKa Pflanzenzucht GmbH & Co. KG, Eichenallee 9, Windeby 24340, Germany

⁷Germicopa Breeding, 1 Allée Loeiz, Quimper 29000, France

⁸Meijer Potato, Bathseweg 47, Rilland 4411 RK, The Netherlands

⁹Danespo A/S, Dyrskuevej 15, Give DK-7323, Denmark

¹⁰AVERIS Seeds, Valtherblokken zuid 40, Valthermond 7876 TC, The Netherlands

*Corresponding author: Teagasc, Crop Science Department, Oak Park, Carlow R93 XE12, Ireland. Email: dan.milbourne@teagasc.ie; *Corresponding author: Plant Breeding, Wageningen University & Research, P.O. Box 386, Wageningen 6700 AJ, The Netherlands. Email: Herman.vanEck@wur.nl

We genotyped a population of 618 diploid potato clones derived from six independent potato-breeding programmes from NW-Europe. The diploids were phenotyped for 23 traits, using standardized protocols and common check varieties, enabling us to derive whole population estimators for most traits. We subsequently performed a genome-wide association study (GWAS) to identify quantitative trait loci (QTL) for all traits with SNPs and short-read haplotypes derived from read-backed phasing. In this study, we used a marker platform called PotatoMASH (Potato Multi-Allele Scanning Haplotags); a pooled multiplex amplicon sequencing based approach. Through this method, neighboring SNPs within an amplicon can be combined to generate multiallelic short-read haplotypes (haplotags) that capture recombination history between the constituent SNPs and reflect the allelic diversity of a given locus in a different way than single bi-allelic SNPs. We found a total of 37 unique QTL across both marker types. A core of 10 QTL was detected with SNPs as well as with haplotags. Haplotags allowed to detect an additional 14 QTL not found based on the SNP set. Conversely, the bi-allelic SNP set also found 13 QTL not detectable using the haplotag set. We conclude that both marker types should routinely be used in parallel to maximize the QTL detection power. We report 19 novel QTL for nine traits: Skin Smoothness, Sprout Dormancy, Total Tuber Number, Tuber Length, Yield, Chipping Color, After-cooking Blackening, Cooking Type, and Eye depth.

2008; Li *et al.* 2010; Baldwin *et al.* 2011; Lindhout *et al.* 2011; Urbany *et al.* 2011; D'hoop *et al.* 2014; Rosyara *et al.* 2016; Schönhals *et al.* 2017; Sharma *et al.* 2018; Klaassen *et al.* 2019; Byrne *et al.* 2020; Prodhomme *et al.* 2020; Vos *et al.* 2022; Zhang, Qu, *et al.* 2022). Genetic studies in diploid potato have largely been based on mapping specific traits using biparental crosses, with relatively few, and generally smaller scale GWAS studies in diploid germplasm sets (Díaz *et al.* 2021; Parra-Galindo *et al.* 2021; Yang *et al.* 2021), and, to our knowledge, this study is the most extensive diploid potato panel, multitrait GWAS so far.

In potato, genotyping-by-sequencing (GBS) as marker discovery and screening strategy usually yields tens to hundreds of thousands of SNPs: e.g. 186k (Sverrisdóttir *et al.* 2017), 40k (Byrne *et al.* 2020), 22.5k markers (Wang *et al.* 2021); and SNP arrays in potato typically contain up to tens of thousands of markers: e.g. 20k (Vos *et al.* 2022) and 8.3k markers (Mosquera *et al.* 2016; Rosyara *et al.* 2016). In a previous study, we developed a marker system called PotatoMASH (Leyva-Pérez *et al.* 2022), with the specific ambition of exploring the potential of low cost, genome-wide genotyping for application in potato breeding and genetics. PotatoMASH surveys 339 loci using a multiplex amplicon sequencing approach followed by deep NGS sequencing (2 × 150 bp Illumina sequencing). The question of what is the minimum number of loci that would provide reasonable genome coverage for effective downstream analysis such as GWAS is in the basis of the development of PotatoMASH. It was previously found that “useful” levels of linkage disequilibrium (LD) extended between 0.6 and 1.5 Mb depending on the population under examination and the LD criterion used. In addition, almost no LD decay was observed across the pericentromeric heterochromatin (Vos *et al.* 2015; Sharma *et al.* 2018). This is why PotatoMASH was designed to detect variation at 339 loci evenly distributed every 1 Mb across the euchromatic portion of the genome (Leyva-Pérez *et al.* 2022), so no site could be more than 0.5 Mb from at least one locus. On the other hand, SNPs are almost entirely bi-allelic, and surveying a single SNP locus per megabase will not efficiently survey the diversity of real haplotypes at any one locus. Because of the high SNP density in potato germplasm, PotatoMASH actually yields >2,000 SNPs, and additional tools can be used for read-backed phasing (Schauumont *et al.* 2022), to create short haplotypes (165–180 bp) that can be used as a multiallelic marker system. These multiallelic haplotags better represent the real allelic composition at a locus and may have better discriminatory power than SNPs for quantitative trait loci (QTL) detection in genome-wide association analysis. Proof of concept of the detection power of PotatMASH was provided by detecting the same QTL associated with fry color that was originally detected in a GWAS involving 40 K GBS-derived SNP markers (Byrne *et al.* 2020). In addition, we observed that the multiallelic haplotags potentially had better discriminatory power than SNPs in GWAS, since the QTL was only detected when using multiallelic haplotags and not SNPs (Leyva-Pérez *et al.* 2022).

In this study, we describe a set of 618 diploid potato genotypes, assembled by a consortium of six breeding programmes (DIFFUGAT project <https://diffugat.eu/>). This material will form the basis of the diploid breeding approaches described above. Phenotypic data were collected on 23 traits over 3 years (2019–2021) This collaborative project aims to improve commercially relevant traits in a diploid genetic background with several essential reproductive traits such as (1) self-compatibility to allow fixation of genetic gains, (2) 2n gametes to allow sexual polyploidization and hybridization with varieties, and (3) a high level of male and female fertility.

The objectives of this study were (1) to characterize this panel for a set of traits that are routinely phenotyped during the selection process in these breeding programmes; (2) to map loci underlying the control of these traits using GWAS; (3) to test haplotags based QTL detection in a wide variety of traits. A longer-term goal is to utilize this information to develop marker-based tools to facilitate selection in this germplasm and extended sets of breeding clones related to it within individual programmes.

Materials and methods

Plant materials and phenotypic evaluation

We used a panel of 618 diploid potato clones provided by a consortium composed of commercial breeders and research institutes. The panel represents clones from diploid breeding programs, where commercially relevant traits are combined with traits important for diploid breeding, such as fertility, self-compatibility, and 2n gamete production. Contributions were made by C. Meijer B.V., The Netherlands—225 individuals; Wageningen University, The Netherlands—134; Danespo A/S, Denmark—101; SaKa Pflanzenzucht GmbH & Co. KG, Germany—93; Germicopa Breeding, France—60; and Averis Seeds B.V., The Netherlands—17 individuals. Because of the commercial nature of the material, pedigree information could not always be provided. In general, the panel is composed of elite diploid breeding clones, primary dihaploids extracted from tetraploid varieties and donors of resistance and fertility traits.

For intellectual property reasons, the breeding material was not shared between companies. Instead, consortium members evaluated their own material using an augmented design, with replicated checks shared across the 6 locations over 3 years (2019–2021). Each company implemented field trial design according to their own system, but the check varieties were included across programmes: Two control varieties were used in 2019 (Lady Claire and Fontane), and 2 additional control varieties were introduced to the experiments in years 2020 and 2021 (Darling and Laperla). Those 4 controls were used to estimate the environmental variance across the sites. Some additional controls were introduced locally within each company in accordance with their local protocols for field trials. The size of each experimental unit was 8 plants per plot, except for Averis, who planted 14 plants per plot, and we accounted for this in measurements that are influenced by number of plants: Yield, Total Tuber Number and Dry Matter Content, by rescaling the measurement proportionally to 8 plants. All companies used a planting distance of 75 cm between the ridges and 30 to 35 cm between plants. Fungicide treatment, fertilizer, and irrigation were applied according to each company's own growing protocol and according to needs each season. More experimental information is provided in [Supplementary File 1](#).

An overview of all 23 morphological, agronomic, and quality traits examined in this study is shown in [Table 1](#). All consortium members used an agreed standardized set of protocols for scoring each trait. While most phenotyping efforts need no further clarification, some observation methods are briefly outlined below. Tuber length (TPM) was measured by counting how many randomly picked tubers are required to fill a PVC gutter of 1 m length ([Supplementary File 2](#)). This means that higher scores are given to shorter tubers. Enzymatic Browning (EnzB) was scored on strings of tuber tissue 2 h after being scraped from peeled raw potatoes using a coarse kitchen grater. Presentability of tubers (PTY) is a holistic trait as defined by breeders' experience and includes regularity and goodness of shape, size, eyes, and skin phenotypes. Skin Smoothness (SkinS) relates to the feel and washability of tubers.

Table 1. Overview of phenotypic traits, scales and numbers of genotypes (including controls) that were assessed for each trait and number of observations over three years.

Trait	Abbr.	Scale	Number of genotypes (including controls)	Number of observations over 3 yr
Yield	YLD	In kg per plant, fresh weight at harvest	567	1650
Canopy stage 1 6 wk after planting	Can1	1 = plants have not yet emerged to 9 = largest canopy in the trial	475	1347
Canopy stage 2 10 wk after planting	Can2	1 = plants have not yet emerged to 9 = largest canopy in the trial	550	1295
Tuber Length	TPM	Tubers per meter count was used with correction table	307	907
Total Tuber Number	TTN	Count of tubers	523	1452
Tuber Shape	TSH	1 = very round, 2 = round, 3 = round-oval, 4 = round-oval to oval, 5 = oval, 6 = oval to long-oval, 7 = long-oval, 8 = long, 9 = very long	569	1646
Yellow Skin Color	YSC	1 = white, 2 = cream, 3 = light yellow, 4 = yellow, 5 = dark yellow, 6 = brown	536	1542
Yellow Flesh Color	FC	1 = clear white, 2 = white, 3 = cream, 4 = light yellow, 5 = yellow, 6 = dark yellow, 7 = very dark yellow	549	1575
Eye Depth	EYE	1 = very deep to 9 = very shallow	567	1648
Presentability of Tubers	PTY	1 = very bad to 9 = very good	565	1644
Skin Smoothness	SkinS	1 = rough to 9 = very smooth	567	1488
Skin Brightness	Gloss	1 = dull to 9 = clear	554	1411
Sensitivity to Common Scab	Scab	1 = heavy symptoms to 9 = no symptoms	424	1278
Enzymatic Browning	EnzB	1 = ink black, 2 = uniformly black, 3 = discoloration to black, 4 = darkening of red and gray discoloration, 5 = bright red and dark gray discoloration, 6 = start of red/gray discoloration, 7 = clear start of discoloration, 8 = very slight discoloration, 9 = no discoloration	552	1467
Cooking Type	CT	2 = very floury, loose boiling, sloughing, 4 = floury, crumbly and fairly loose, 6 = slightly floury and fairly firm, 8 = not floury, firm cooking, 9 = extreme firmness	555	1393
After-cooking blackening	ACB	1 = very dark to 9 = pure color (no darkening at all)	558	1498
Chipping Color 1st time point stored at 8°C	QDC_1-8	1 = very dark to 9 = pure color (no darkening at all)	559	1593
Chipping Color 2nd time point stored at 8°C	QDC_2-8	1 = very dark to 9 = pure color (no darkening at all)	367	857
Chipping Color 2nd time point stored at 4°C	QDC_2-4	1 = very dark to 9 = pure color (no darkening at all)	505	1353
Dry Matter Content	DM	% relative to fresh weight	566	1626
Sprout Dormancy	SD	1 = heavy sprouting (early) to 9 = no sprouting	536	1416
Tuber Regularity	REG	1 = bad to 9 = good	565	1646
Maturity	MAT	1 = plants still green and flowering to 9 = plants reached senescence	427	857

Skin Brightness (Gloss) is a visual assessment referring to a glossy or shiny skin finish. Cooking Type (CT) was evaluated by boiling samples of two tubers per plot for 25 min. After-cooking blackening (ACB) was assessed on the cooled-down potato 1 day after cooking. Processing quality was assessed using Chipping Color data from three treatments: (1) tubers stored at 8°C for 4 months before crisping (QDC1-8), tubers stored at 8°C for 6 months before crisping (QDC2-8) and tubers stored at 4°C for 6 months before crisping (QDC2-4). The color is assessed for three potato tubers, cut into slices of 1 mm and fried at 180°C until water (“bubbles”) has disappeared from the crisps.

Statistical analysis of phenotypic data

All statistical analyses and data visualizations were performed using R version 4.2.1 unless otherwise specified in results. Visual inspection of the distribution of the data and quantile–quantile (QQ) plots of residuals vs quantiles revealed some obvious deviations from homoscedasticity or normality in the continuous traits: Yield, Tuber Length, and Total tuber number per plant (Fig. 1). The data of those traits were transformed with Yeo-Johnson transformation using the R package “car” (Fox and Weisberg 2019). Although the majority of the traits were measured on an ordinal scale, an inspection of diagnostic plots for residuals indicated no strong violations of the assumption of normal error distributions (Supplementary File 3) and were all analyzed as quantitative

traits, as previously performed (D’hoop et al. 2008), assuming the error variation to be normally distributed with constant variance.

Check varieties were used in the estimates of the Best Linear Unbiased Estimators (BLUEs) of phenotypic means of all 23 traits across years and locations but were excluded from the GWAS. We used a multiple linear regression package lme4, using the lm function (Bates et al. 2015) to calculate the BLUEs with the following equation:

$$\text{Trait} = \text{Genotype} + \text{Year} + \text{Location} + \text{Location} * \text{Year} + \text{Error},$$

where genotype is the clone name and does not include any genetic information such as pedigree due to intellectual property rights and Location is the site of each company. Location*Year effect was applied when analyzing data from more than one company. All independent variables: Genotype, Year, Location and Location*Year, were considered as fixed effects due to the low number of levels.

Least square means, calculated for the BLUEs with the R package “emmeans” (Lenth et al. 2021), served as the final phenotypic data used in the association analysis.

A Pearson’s correlation matrix between the vegetation indices and the vegetative growth parameters was generated using the package corplot for R. Correlation coefficients were tested at $P = 0.05$.

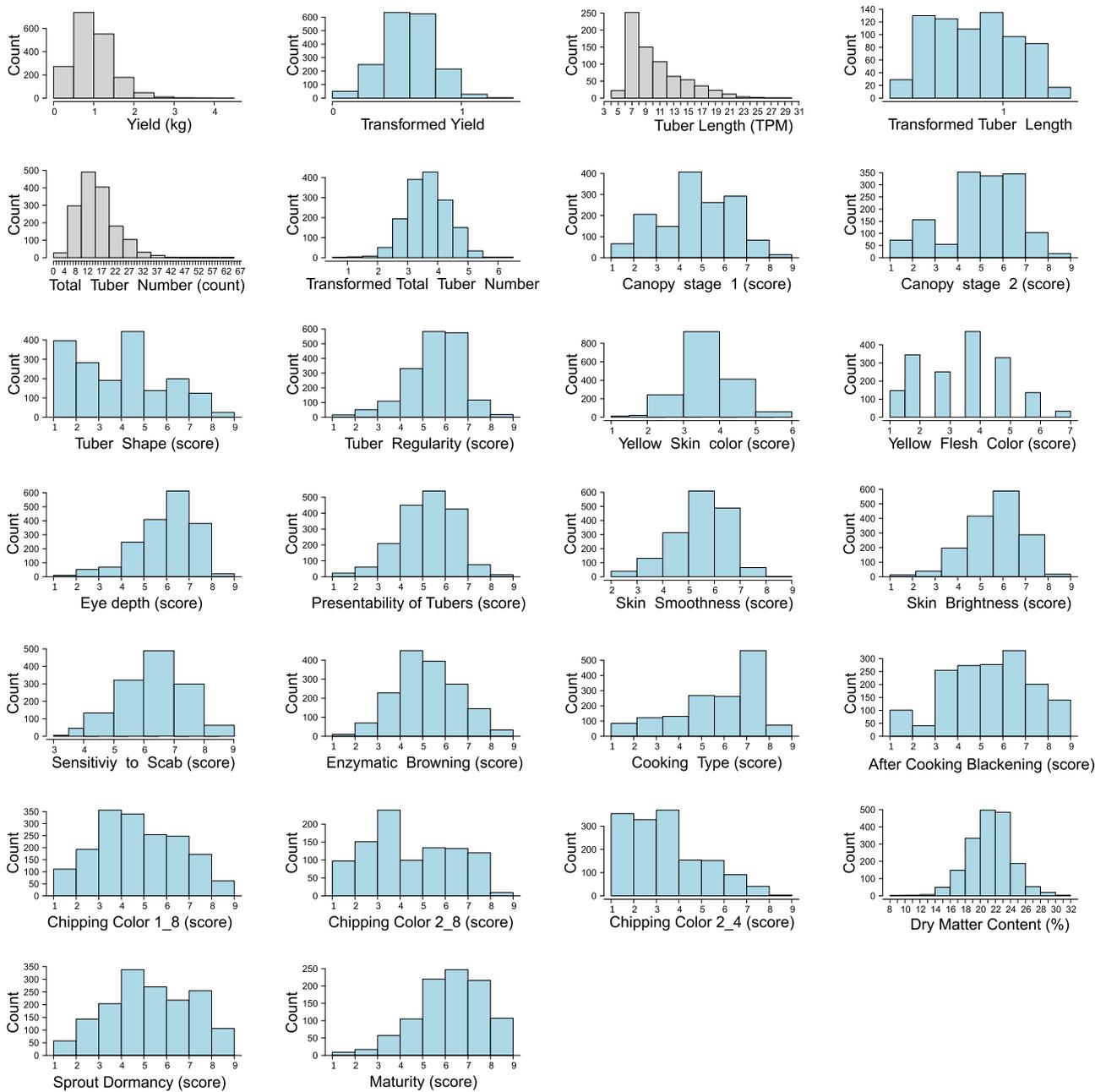


Fig. 1. Frequency distribution histograms for all traits of this study, across all companies. The horizontal axis indicates the data range of traits, and the vertical axis indicates the frequency of individuals. For traits Yield, Tubers Length and Total Tuber Number per plant: the histograms for Yield, Tubers Length and Total Tuber Number (in gray) represent the raw data, and the blue histograms represent the transformed data that we used in the downstream analysis.

Broad-sense heritabilities (H^2) were calculated for each breeding population separately on an entry-mean basis according to the formula:

$$H^2 = \sigma_g^2 / (\sigma_g^2 + \sigma_{gy}^2 / n_{year} + \sigma_e^2 / n_{year}),$$

where σ_g^2 is the genotypic variance, σ_{gy}^2 is the genotype-by-year variance, σ_e^2 the error variance, and n_{year} is the number of years.

The k-matrix of the genomic data was calculated with the R package GWASpoly (Rosyara et al. 2016) and the modeled least square means were used to calculate the Marker-based heritabilities, (h_{NPs}^2 and $h_{haplotags}^2$) with the R package “heritability” with the *marker_h2_means* function (Kruijer et al. 2014; Kruijer and White 2023).

Genotypic data

Data collection with PotatoMASH

Leaf material was sampled in 2019, the first year of the field trials, freeze-dried and stored with silica gel until use. Approximately 5 mg of dry tissue was used to extract DNA with Mag-BIND Plant DNA DS Kit (Omega-VWR M1130-00, Philadelphia, USA), using the KingFisher Flex automated extraction & purification system (Thermo Scientific, Austin, TX, USA).

PotatoMASH libraries were obtained and haplotyping was performed as in Leyva-Pérez et al. (2022) (<https://www.protocols.io/view/potatomash-library-construction-e6nvw53zdvmk/v2>) with the following adjustments to the bioinformatics pipeline: Merged and filtered reads were mapped to the *S. tuberosum* genome v6.1 (Pham et al. 2020). Variant calling was then filtered

with a minimum allele frequency of 0.01 and a maximum of 0.99: `vcftools -bcf PotatoMASH.bcf -out PotatoMASH -min-alleles 2 -max-alleles 2 -recode -recode-INFO-all -minQ 30 -minDP 6 -maf 0.01 -max-maf 0.99 -remove-filtered-all -max-missing 0.5`.

Haplotypes nomenclature is given by the software SMAP (Schaumont et al. 2022) as a binary string code for the set of SNPs called in a specific locus, where the reference allele of each SNP is coded as “0” and the alternative allele is coded as “1” in a specific haplotype (Fig. 4c). The final haplotag name is the PotatoMASH locus name plus the binary string in which 0 means same base as reference genome, 1 is alternative base, and “-” is an indel at that SNP position (e.g. C1_1_000110-10).

For the 334 polymorphic loci, the average locus correctness score (number of samples with sum of discrete haplotag dosage calls equals 2, divided by total number of samples with sufficient read depth for that locus, expressed as percentage) was 92. SMAP also calculates the sample correctness score per sample (number of loci where the sum of discrete haplotag dosage calls equals 2, divided by the total number of loci with sufficient read depth, expressed as percentage). Since the average locus correctness score was high for the 334 loci, we assumed that individuals with low sample correctness score would be due to technical errors or putative cross contamination. Therefore, we removed 21 genotypes with a sample correctness lower than 40. A final panel of 558 genotyped individuals was used for the GWAS. SNPs and haplotags datasets are provided in Supplementary Files 4 and 5.

Population structure

Population structure was evaluated using a principal component analysis (PCA) calculated with Plink 1.9 using SNPs with a minimum allele frequency >0.01 (Purcell et al. 2007). The population genetic structure was assessed using the Bayesian clustering method implemented in STRUCTURE version 2.3.4 (Pritchard et al. 2000). An admixture model and correlated allele frequencies were chosen for estimating the proportion of ancestral contribution in each accession. We tested various K-values ranging from 1 to 10 with 3 independent replications at each K, 10,000 generations burn-in period and 10,000 Markov Chain Monte Carlo (MCMC) repetitions. Calculation of Delta K: (1) Mean $L(K)$ (\pm SD) was done over 3 independent runs for each K value. (2) Rate of change of the likelihood distribution (mean \pm SD) was calculated as $L'(K) = L(K) - L(K - 1)$. (3) Absolute values of the second order rate of change of the likelihood distribution (mean \pm SD) were calculated according to the formula: $|L''(K)| = |L'(K + 1) - L'(K)|$. (4) ΔK calculated as $\Delta K = \text{mean}[L''(K)] / \text{sd}[L(K)]$ (Evanno et al. 2005). Visualizing admixture plot was done with the fastSTRUCTURE software `distruct.py` function (Raj et al. 2014).

Genome-wide association studies

Two datasets were used for the GWAS. We first identified SNPs across the sequenced amplicons and used these as a data set in a GWAS. We then used this SNP set to construct short haplotypes with SMAP `haplotype-sites` tool (see “Data collection with PotatoMASH” section) combined with discrete genotype calling, that were then used simply as presence-absence markers for GWAS. The distinct haplotags were treated as “pseudoSNPs” for the purpose of the analysis.

Association analysis for both SNP and haplotag data was done with the R package GWASpoly (Rosyara et al. 2016). Population structure was controlled using the K model and QQ plots were used to assess if there was sufficient control of population structure. The function GWASpoly with additive and nonadditive models was used to test for association at each marker. Marker

curation was carried out using the maximum genotype frequency option with default parameter setting (`geno.freq = 1-5/N`, where N is the number of genotypes), so markers present in fewer than 5 individuals are removed. The genome-wide false discovery rate was controlled using the M.eff method (a Bonferroni-type correction but using an effective number of markers that accounts for linkage disequilibrium (LD) between markers) at level = 0.05. We did not use the leave-one-chromosome-out (LOCO) approach due to the inflation of the P-values as observed with the QQ plots.

Results

Phenotypic data

Taken together, phenotyping of the panel of 618 diploids resulted in 32,590 data points, collected over 3 years, across 6 locations, for a total of 23 agro-morphological and quality traits (Fig. 1, Table 2). These data were unbalanced given that some locations/breeders focused on a single niche market (e.g. starch). A strong year-by-location interaction was observed (Supplementary File 6) using control varieties planted across all sites. From these raw data best linear unbiased estimators (BLUES) were calculated while taking the year-by-location interaction into account with the regression models.

We could not calculate the broad sense heritability (H^2) across all companies, as only the control varieties were shared. The heritabilities presented in Table 2 are the average of the estimated trait heritability for each company and varied mostly between moderate to high values, ranging from 50 to 90%. Traits largely controlled by single loci such as Tuber shape (TSH), Yellow flesh color (FC), and Maturity (MAT), typically show H^2 between 82 and 90%, according to expectations. Some of the complex polygenic traits like Dry Matter Content (DM) and Yield, also show an exceptionally high H^2 of 85–89%. Furthermore, the majority of the processing and quality traits such as Enzymatic Browning (EnzB), Cooking Type (CT), After-cooking Blackening (ACB), and Chipping Color showed moderately high H^2 values (73–84%). Marker-based heritability, were also calculated using both marker types, SNPs and haplotags, and were lower than the broad sense heritability for all traits (Supplementary File 7).

Correlations between traits

The correlations between traits are shown in Fig. 2, and are based on the phenotype estimated means (Supplementary File 8). The highest positive correlations were observed between Skin Smoothness and Skin Brightness (while both had a negative correlation with Yellow Skin Color). High correlations were also observed for tuber visibility traits such as Tuber Regularity, Tuber Presentability, and Eye depth. Yield showed a negative correlation with Maturity and a positive correlation with Canopy development. Canopy stage 1 and Canopy stage 2 positively correlated with Total Tuber Number.

Genotyping, variant calling, and genetic diversity

After merging and filtering reads, we retained on average 275,855 reads per sample, which corresponds to an average of 813 reads per amplicon per sample, well exceeding the required minimum of 20x read depth recommended for SMAP haplotype calling for diploids. The amplification efficiency of the primer pairs (either low or high) was consistent across most samples (Fig. 3).

After filtering, 2,730 SNPs were identified across the panel (Supplementary File 4). Out of 339 PotatoMASH target loci, SMAP

Table 2. Mean, standard deviation (SD), minimum (Min), maximum (Max) values, and Broad sense (H^2) Heritability (%) for all traits.

Trait	Min	Max	Mean	SD	Average H^2 across all companies	Number of companies tested
Canopy stage 1	1	9	5.24	1.69	60.1	5
Canopy stage 2	1	9	5.53	1.73	75.2	6
Yield	0.003	4.26	0.98	0.5	85.5	6
Tuber Length	5.37	28.11	10.37	4.01	87.1	4
Total Tuber Number	0.75	61.25	15.42	6.72	69.2	5
Tuber Shape	1	9	4.4	2.07	90.4	6
Tuber Regularity	1	9	6.05	1.23	64.9	6
Yellow Skin Color	1	6	4.12	0.8	63.6	6
Yellow Flesh Color	1	7	3.6	1.5	88.8	6
Eye depth	1	9	6.47	1.31	82.6	6
Presentability of Tubers	1	9	5.65	1.24	74.5	6
Skin Smoothness	2	9	5.95	1.09	60.3	6
Skin Brightness	1	8	5.58	1.09	62.1	5
Sensitivity to Common Scab	3	9	6.76	1.16	50.6	6
Enzymatic Browning	1.5	9	5.51	1.36	81.7	6
Cooking Type	2	9	6.62	1.68	74.3	6
After-cooking Blackening	1	9	5.9	1.9	72.7	6
Chipping Color 1_8	1	9	5.25	1.86	77.8	6
Chipping Color 2_8	1	9	4.93	1.96	83.6	3
Chipping Color 2_4	1	9	3.82	1.77	81.8	4
Dry Matter Content	8.35	31.08	21.35	2.84	88.7	6
Sprout Dormancy	1	9	5.81	1.82	74.9	5
Maturity	1	9	6.6	1.44	82.0	4

haplotype-sites could identify 334 loci with polymorphic, multiallelic haplotypes. A total of 2,955 short multiallelic haplotags were identified across the panel (Supplementary File 5), ranging from 2–30 haplotags per locus, while most loci had 8–9 haplotags per locus (Table 3, Fig. 4a). This is higher than previously reported by Leyva-Pérez et al. (2022) in a tetraploid population where 2–14 haplotags per locus (on average 6 haplotags per locus) were reported. The higher haplotype diversity suggests higher genetic diversity in the used diploid panel.

As expected in our diploid panel, 2 haplotags (either homozygous or heterozygous) were successfully called at each locus, for each individual, in 91% of cases. SMAP analyses the relative read depth per haplotag per locus per individual, and outputs the distribution across all loci to check, if that fits the typical frequency spectrum expected for diploids (Fig. 4b). Thirty-nine of the individuals showed a tetraploid typical frequency spectrum and were excluded. As final output, we obtained a table with discrete dosage calls for each haplotag per locus, per sample (Fig. 4c), which was used for downstream analysis.

Population structure

We examined population structure by principal component analysis (PCA) using the SNP data and observed 2 main clusters, with separation mainly occurring on the 1st principal component which explains ~17% of the genetic variation, indicating that the diploid population of Meijer deviates from the gene pools of the other breeding programs (Fig. 5a). We also examined the underlying population structure of the panel through Bayesian-based approach using STRUCTURE v 2.3.4. and with the log mean probability and deltaK (change in log probability) per K (number of sub-populations) generated the highest peak at K = 2 (Fig. 5c), and this confirmed the conclusion of two sub-populations. We therefore decided to perform QTL discovery using three sets of potato genotypes: the “full panel”, the sub-populations “only Meijer” and the rest not including Meijer (referred as “no Meijer”). In this way, we were hoping to capture QTL that were robust across all subpopulations in addition to subpopulation-specific QTL.

GWAS of multiple traits

To capture all the potential QTL, we performed six GWAS (three genotype-sets described above with the two marker-sets, SNPs and haplotags).

We identified 37 QTL for 20 out of 23 traits. For three traits: Tuber regularity, Skin brightness and Presentability of tubers we did not detect QTL. Of the 37 QTL identified, only 10 QTL were detected with both SNPs and haplotags. Fourteen QTL were only detected by haplotags, and 13 QTLs were only detected by SNPs (Fig. 6; Table 4). The full information of all the significant markers, including their marker’s effects, are provided in Supplementary File 9.

Differences in QTL detected across populations

Differences in the sets of QTL were observed across populations. Only 5 QTL, for Tuber Shape, Eye depth, Yellow Flesh Color (2 QTL) and Maturity were detected across all 3 populations with at least 1 marker shared for each trait across all populations. Another 15 QTL were detected either in the “no Meijer” sub-population (6) or in the “only Meijer” sub-population (9). Eleven QTL were shared between the full panel and one of the other two sub-populations. Ten QTL were detected with the full panel but were not significant in either sub-population.

Identification of previously characterized QTL

Some of the traits evaluated here, were previously described in detail in the literature and enabled us to validate our approach (Fig. 7). Indeed, we identified a highly significant QTL ($-\log_{10}(P\text{-value}) = 14.36$) on chr10 across the 3 phenotypic datasets for Tuber Shape, which was detected both with SNP and haplotag markers (Fig. 7, Supplementary File 9). This QTL corresponds to the well described Ro locus (van Eck et al. 1994; van Eck et al. 2022).

The well-known Y (Yellow) locus and the causal gene involved in yellow flesh color beta-hydroxylase (*Chy2* or *BCH*) (Bonierbale et al. 1988; Thorup et al. 2000; Brown et al. 2006; Wolters et al. 2010), One isoform (PGSC0003DMG400009501) of the *Bch* gene was reported to be located at 44.1 Mb in DMv4.03 (Pandey et al.

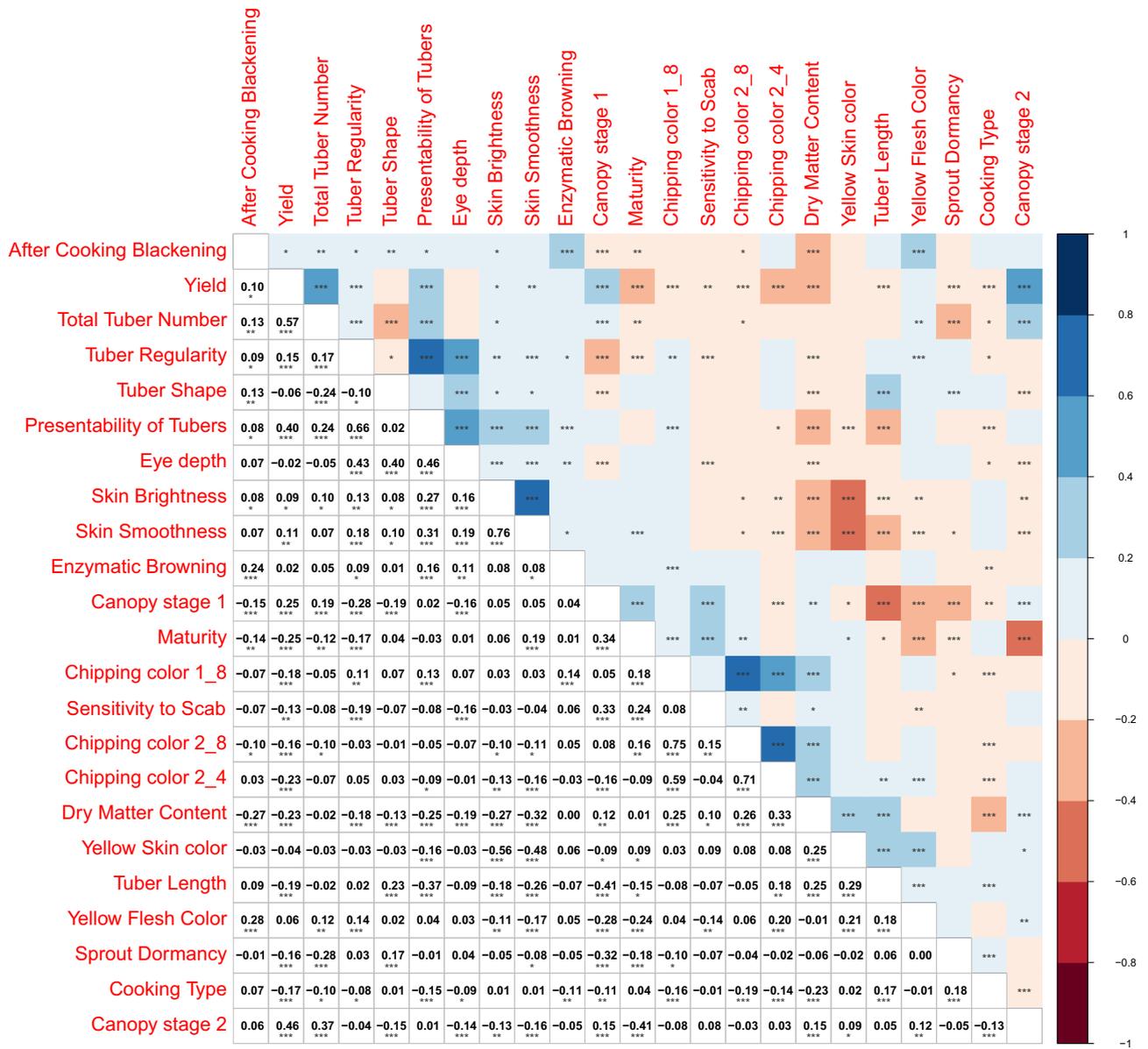


Fig. 2. Matrix of pairwise Pearson's correlation between all traits. Positive correlations are displayed in blue and negative correlations in red. Color intensity is proportional to the correlation coefficients according to the scale displayed on the right. Marking of significance level: ***0.001, **0.01, *0.05.

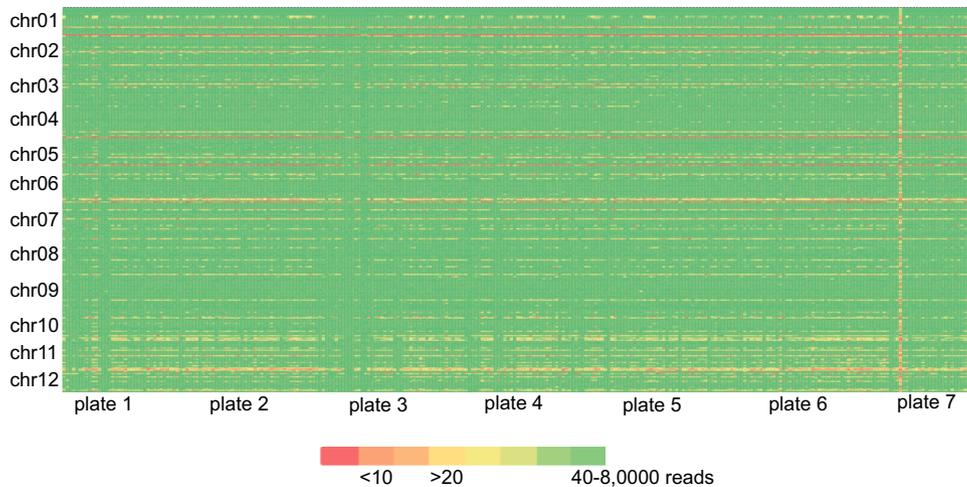


Fig. 3. Coverage of the 339 PotatoMASH core loci. Heat map of the number of merged and filtered reads of 618 samples (in columns, each plate is of ~96 samples) that mapped to each locus (in rows).

2022) and aligns with position 42.9 Mb on chr03 of the DMv6.1 reference genome sequence (Pham et al. 2020). We identified 1 QTL for Yellow flesh color on chr03 spanning the region from 44.04

Mb to 49.75 Mb (PotatoMASH loci C3_15 to C3_20), peaking at 47.03 Mb (PotatoMASH locus C3_18) with LOD score ($-\log_{10}(P\text{-value})$) of 44.66 for the significant marker chr03_47024967 (Supplementary File 9). Two additional PotatoMASH loci on chr03 (C3_11 at 40.41 Mb; and C3_30 at 59.35 Mb) also showed significant associations with Yellow Flesh Color.

Furthermore, we also detected a QTL for Maturity on chr05, peaking at 4.94 Mb (PotatoMASH locus C5_6). The haplotag C5_6_0011010 and the 3 SNPs of this haplotag (chr05_4941391, chr05_4941406 and chr05_4941464) were associated with late maturity and the haplotag C5_6_0000000 with early maturity. This QTL is near to the region containing StCDF1 gene (Soltu.DM.05G005140.1, chr05:4485531.4488495 DMv6.1), which is well established as the gene largely responsible for the plant maturity in potato (Kloosterman et al. 2013).

Table 3. Summary of genotyping and variant calling with PotatoMASH.

Total samples	558
SNPs called	7503
SNPs filtered	2730
Polymorphic loci	334
Number of haplotypes	2955
Haplotypes per locus	2–30 (8.8 avg.)
2 haplotags called per locus per individual (either homozygous or heterozygous)	91%

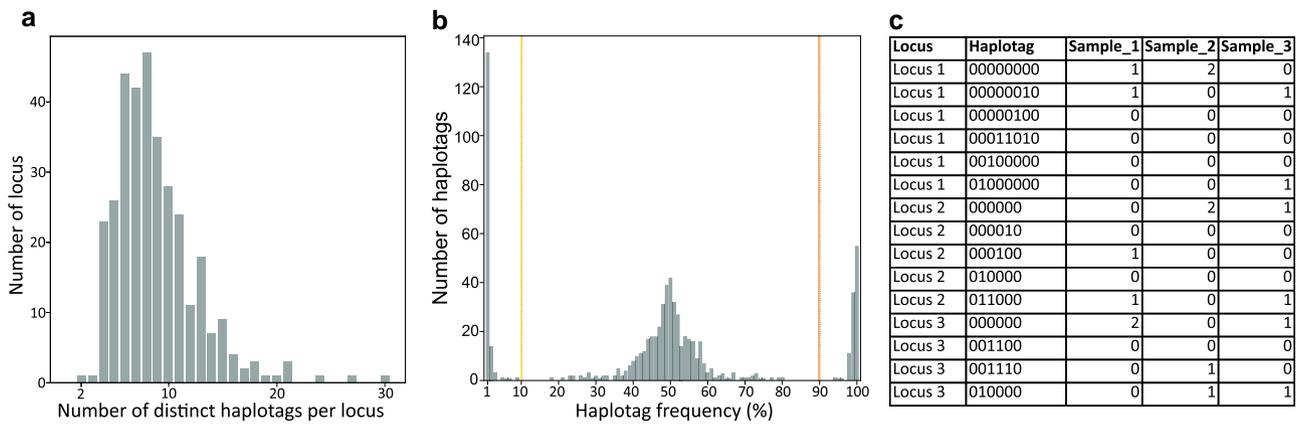


Fig. 4. a) Haplotag diversity distribution of 334 loci across the individuals in the panel. b) Haplotag frequency spectrum of one individual, the haplotag frequency is calculated by the relative read depth (%) for each haplotag within its locus. c) Example of tabular data generated by SMAP haplotype-sites with 3 genotypes (samples), 3 loci, 15 haplotags, and diploid discrete dosage calls for each locus/sample. Loci 1 and 3 include haplotags not detected in samples 1–3 but in other genotypes not shown (samples 4–558).

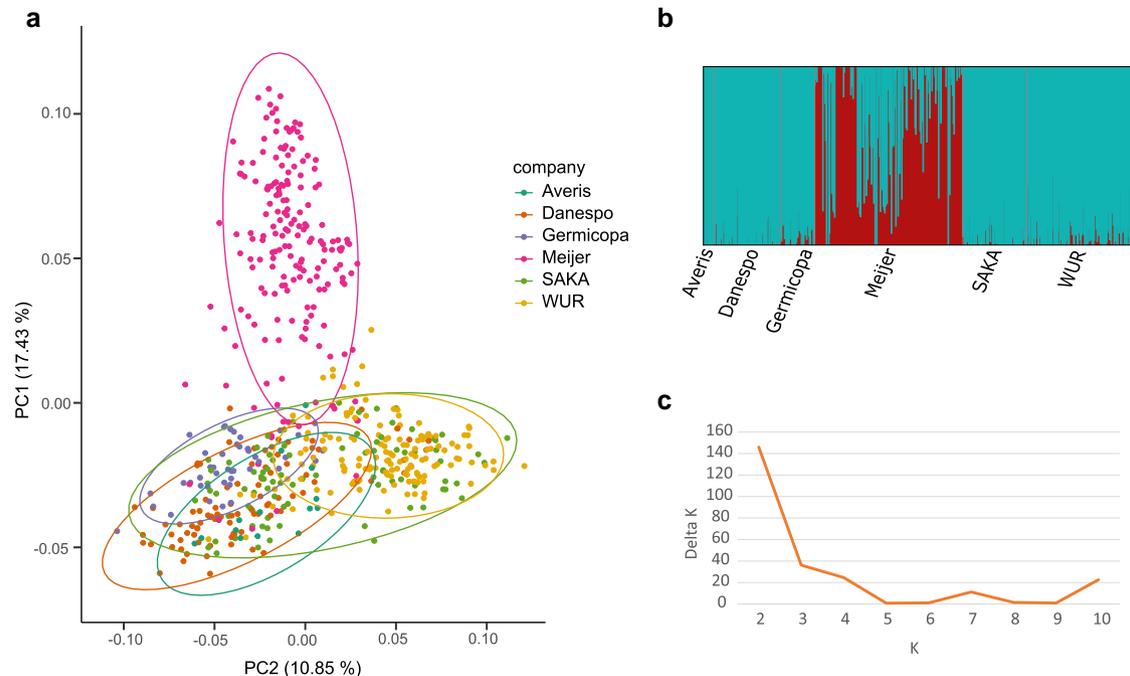


Fig. 5. a) Principle component analysis (PCA) with SNP data of all six companies. b) Estimation of hypothetical sub-populations using K-values. c) The number of identified sub-populations (K) vs DeltaK estimated based on Evanno method.

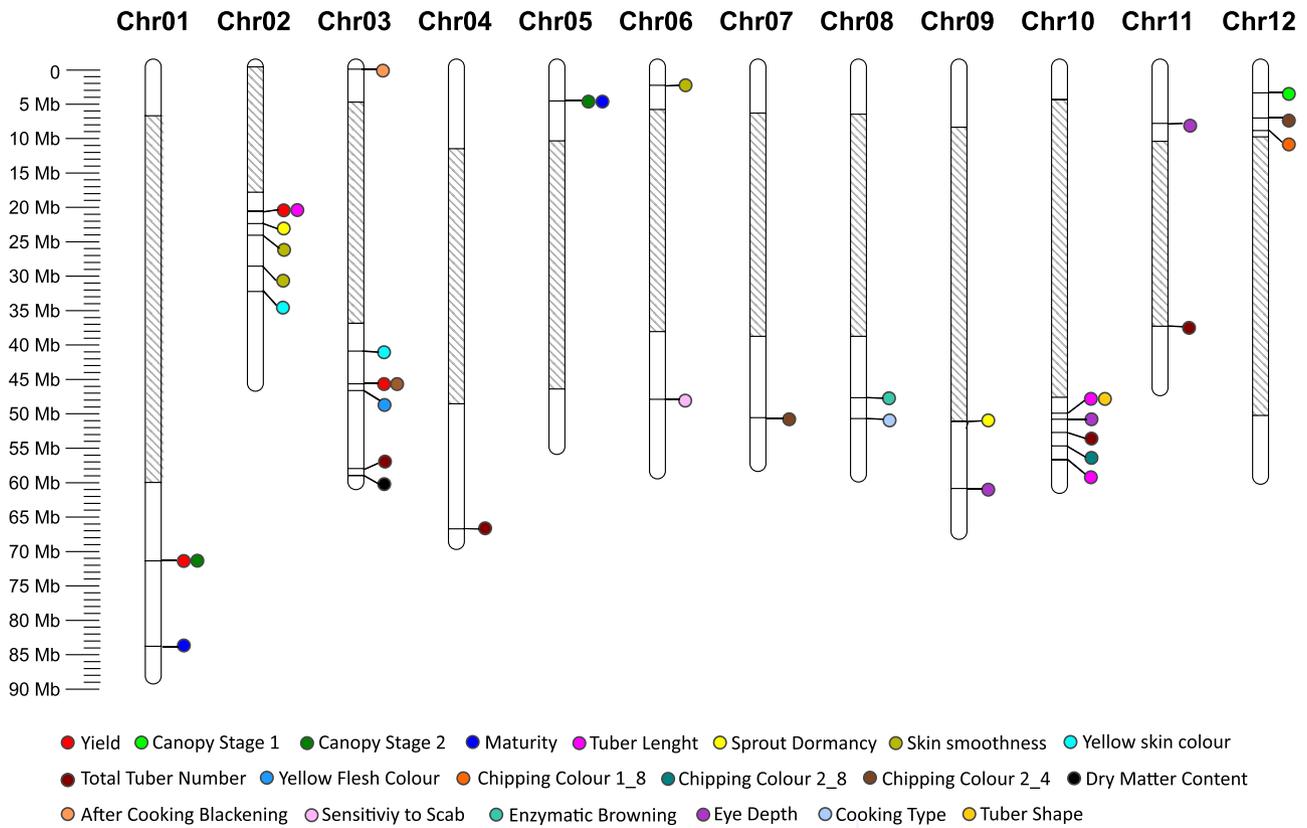


Fig. 6. Physical map with the positions (Mb) of all QTL. Gray regions on the chromosome indicate pericentromeric heterochromatin, without PotatoMASH amplicons.

QTL for agronomic and morphological traits

Eighteen of the QTL identified in this work were confirmed with previous QTL studies in potato at the diploid and tetraploid level (Table 4). We also detected new QTL for complex traits not yet reported before. In total, we discovered 19 novel QTL on 8 chromosomes: 5 QTL on chr02—two QTL for Skin Smoothness, 1 for Sprout Dormancy, 1 for Total Tuber Number, and 1 for Tuber Length. Four QTL on chr03—1 for Yield and 1 for Total Tuber Number. One QTL on chr06 for Skin Smoothness. One QTL on chr07 for Chipping Color. Two QTL on chr08—1 for After-cooking Blackening and 1 for Cooking Type. Two QTL on chr09—1 for Eye depth and 1 for Sprout Dormancy. One QTL on chr10 for Total Tuber Number. Two QTL on chr11—1 for Eye depth and 1 for Total Tuber Number. Three QTL on chr12—1 for Canopy stage 2 and 2 for Chipping Color.

We detected a QTL for Canopy stage 2 (canopy coverage 10 weeks after planting) peaking at the same PotatoMASH locus where the well-known QTL for Maturity was detected, C5_6 (4.94 Mb). The significant haplotag, C5_6_0000000 was associated with earliness and lower canopy cover while SNPs chr05_4941391 and chr05_4941406 were associated with lateness and higher canopy cover. This association between maturity and canopy type is also confirmed by the significant correlation between the phenotypic values of these two traits ($r^2 = 0.41$). The two additional QTL detected on chr01 and chr12 for early-stage canopy development (6 weeks after planting), could not be associated with plant maturity and seem to be caused by genetically independent loci affecting canopy vigor.

The novel QTL for Yield was detected in chr03, locus C3_17 (46.06 Mb). It was identified in both the full panel and in the sub-

population “no Meijer” with a significant SNP chr03_46058754 and with the haplotag specific to this SNP, C3_17_011000, both associated with low yield. A QTL in this region was also detected for Total Tuber Number with the same significant SNP and haplotag both associated with low Total Tuber Number. This could be a new region associated with Yield and yield-related traits and is also supported by the significant correlation between Yield and Total Tuber Number ($r^2 = 0.57$). Two additional PotatoMASH loci on chr03, C3_7 and C3_29 were associated with low Total Tuber Number although we had not considered them separate loci in the QTL count.

Two additional novel QTL were detected for Total tuber number. One QTL on chr10 at locus C10_12 (53.13 Mb) was identified only in the full panel but not in any of the sub-populations, with a significant haplotag C10_12_00000100000 associated with low Total Tuber Number. No SNP was significant. One QTL was detected on chr11 at locus C11_12 (37.67 Mb), for the “no Meijer” sub-population only, with a significant haplotag C11_12_000100000 associated with low Total Tuber Number. No SNP was significant.

For Tuber Length, we detected one new QTL on chr02 at locus C2_4 (20.96 Mb), with the significant SNP chr02_20959691 with a small positive effect associated with shorter tubers (higher number of tubers per meter, TPM). This association is based on only 127 individuals of the “no Meijer” sub-population. This could be another new region associated with Yield and yield-related traits and is also supported by the significant correlation between Yield and Tuber Length ($r^2 = -0.19$, high Yield correlates negatively with shorter tubers). A QTL for higher Yield was also detected in the same locus, C2_4, but in the sub-population “only Meijer” and with different SNPs/haplotags suggesting different origins of this locus.

Table 4. Overview of the 37 QTLs for the 20 trait, with the significant markers for each population: underlined—QTL only with SNPs, italics—QTL only with haplotags, bold—QTL with both. Columns from left to right: Trait, given name of QTL, QTL location in potato genome DMv6, name of haplotag and SNP for each population, previously reported QTL and positions in Mb when available, in potato genome DMv4 and Literature column citing the previous works in which these QTLswere found.

Trait	QTL			Full panel		No Meijer		Only Meijer		Literature
	Name	chr (Mb in DMv6.1)	chr (Mb in DMv4.3 when available)	Haplotags	SNPs	Haplotags	SNPs	Haplotags	SNPs	
Yield	YLD_C1_19	<u>chr01 (71.77)</u>				C1_19_00110			chr02_20959684_12	Bradshaw et al. (2008), McCord et al. (2011), Hurrado-Lopez et al. (2015), Manrique-Carpintero et al. (2015), Rak et al.
	YLD_C2_4	<u>chr02 (20.99)</u>					chr03_46058754		<u>chr02_20959746</u>	
	YLD_C3_17	chr03 (46.06)		C3_17_011000	chr03_46058754	C3_17_011000		chr03_46058754		
Skin Smoothness	Skins_C2_8	chr02 (24.47)						chr02_24470953		
	Skins_C2_13	<u>chr02 (28.96)</u>		C2_13_100010000						
	Skins_C6_3	<u>chr06 (2.68)</u>		C6_3_0000101-00		C2_8_00000010000 C2_13_100010000 C6_3_0000101-00				
	CT_C8_21	<u>chr08 (51.08)</u>								
After-cooking Blackening	ACB_C3_1	<u>chr03 (3.38)</u>								
	DM_C3_30	<u>chr03 (59.35)</u>								
Canopy Stage 1 Canopy Stage 2	Can1_C12_0	<u>chr12 (3.78)</u>								
	Can2_C1_19	<u>chr01 (71.77)</u>								
	Can2_C5_6-C5_8	chr05 (4.94-7.25)		C5_6_0000000	chr05_4941391, chr05_4941406, chr05_7251491, chr05_7251555			chr12_3783754	chr05_4941391, chr05_4941406, chr05_6204154, chr05_7251491, chr05_7251555	Kloosterman (2006), D'hoop et al. (2008), D'hoop et al. (2014), Bradshaw et al. (2008), D'hoop et al. (2014)
Enzymatic Browning	EnzB_C8_18	<u>chr08 (48.05)</u>								
	TSH_C10_7-C10_12	chr10 (48-53.13)		C10_7_01111000010, C10_8_0001111100, C10_8_0001111110, C10_8_0001111110, C10_8_011000000, C10_9_000101000100000, C10_9_00010100010110	chr10_49148246, chr10_49148249, chr10_49148305, chr10_49148316, chr10_50323192, chr10_50323244	C10_8_0001111100, C10_9_000101000100000, C10_12_0001010000000	chr10_49148293, chr10_49148305, chr10_49148316, chr10_50323192, chr10_50323244	C10_8_0001111110, C10_8_011000000, C10_9_000101000100000, C10_9_0111100000110110, C8_18_0010011001001010100	chr10_49148246, chr10_49148249, chr10_49148305, chr10_49148316, chr10_50323192, chr10_50323244	Wenji et al. (2007), D'hoop et al. (2008), Urbany et al. (2011), D'hoop et al. (2014), van Eck et al. (1994), Rosyara et al. (2016), Sharma et al. (2018), Pandey et al. (2022)
Tuber Shape										

(continued)

Table 4. (continued)

Trait	Name	chr (Mb in DMV6.1)	Full panel			No Meijer			Only Meijer			Literature	
			Haplotags	SNPs	Haplotags	SNPs	Haplotags	SNPs	Haplotags	SNPs	chr (Mb in DMV4.3 when available)	Reference	
Eye depth	EYE_C9_21-C9_23	chr09 (61.25-62.91)						C9_21_000100010100, C9_23_000001	chr09_61254396, chr09_62045322	3 (53.1), 5 (43.9), 10 (48.6)	Li et al. (2005)		
	EYE_C10_6-C10_10	chr10 (4.74-51.2)	C10_9_0001010001000000	chr10_4741749, chr10_49148293, chr10_49148305, chr10_49148316, chr10_50323244,		chr10_4741749, chr10_49148293, chr10_49148305, chr10_49148316		C10_9_011100000110110, C10_9_000101000100000, C10_10_01010000	chr10_49148246, chr10_49148293, chr10_49148305, chr10_49148316, chr10_50323151, chr10_50323153, chr10_50323225, chr10_50323244, chr10_50323263,		Slavicka et al. (2008), Kosyara et al. (2016), Sharma et al. (2018), Pandey et al. (2022)		
Maturity	EYE_C11_8	chr11 (8.19)		chr11_8190769		chr10_4941391, chr05_4941406, chr05_4941464			chr05_4941391, chr05_4941406, chr05_6204154		(McCord et al. 2011), Kloosterman et al. (2013), da Silva Pereira et al. (2021b)		
	MAT_C1_3	chr01 (84.18)		chr01_8417600									
Tuber Length	MAT_C5_6-C5_7	chr05 (4.94)	C5_6_0000000, C5_6_0011010, C5_7_0001111110	chr05_4941391, chr05_4941406, chr05_4941464	C5_6_0000000,	chr05_4941391, chr05_4941406	C5_7_0001111110						
	TPM_C2_4	chr02 (20.96)											
Sensitivity to Common Scab	TPM_C10_8-C10_10	chr10 (49.15-51.02)	C10_8_0001111110, C10_10_00100100					C10_8_0001111110, C10_9_000101000100000, C10_9_011100000110110, C10_10_01010000,	chr10_49148246, chr10_49148249, chr10_49148293, chr10_49148305, chr10_49148316, chr10_49148341, chr10_50323151, chr10_50323153, chr10_50323192, chr10_50323225, chr10_50323244, chr10_50323263	1, 2, 3, 5, 7	Zhang, Qu, et al. (2022)		
	TPM_C10_16	chr10 (67.02)		C10_16_000000000000									
Total Tuber Number	TTN_C3_29	chr03 (58.36)		chr03_58369063, chr03_58369113					chr03_58369063, chr03_58369113				Manrique-Carpintero et al. (2015), Zhang et al. (2022)
	TTN_C4_33	chr4 (67.1)	C4_33_001000010		C4_33_001000010								
Sensitivity to Common Scab	TTN_C10_12	chr10 (62.17)	C10_12_000001000000		C11_12_000100000								
	TTN_C11_12	chr11 (67.67)			C6_18_011111111110								
Chipping Color 1-8	Scab_C6_18	chr06 (48.3)	C6_18_011111111110										
	QDC1_8-C12_11	chr12 (9.26)		chr12_9258876									
Chipping Color 2-8	QDC2_8-C10_14	chr10 (55.08)											
	QDC2_4-C3_17	chr03 (46.06)											
Chipping Color 2_4	QDC2_4-C7_23	chr07 (50.94)			C3_17_000010				chr10_55081844				
	QDC2_4-C12_9	chr12 (7.44)	C12_9_001001000100										
Yellow Flesh Color	FC_C3_15-C3_20	chr03 (44.04-49.75)	C3_15_0000000, C3_17_000000, C3_18_0000000, C3_18_00011000, C3_20_000000	chr03_44040545, chr03_46058758, chr03_47024967, chr03_47024968, chr03_49750222	C3_15_0000000, C3_17_000000, C3_18_0000000, C3_18_00011000	chr03_45076941, chr03_47024967, chr03_47024968, chr03_49750222	C3_17_000000, C3_18_0000000, C3_18_00011000	chr03_47024967, chr03_47024968	1 (63.8), 3 (44.1), 3 (48.5), 3 (49.3), 3 (50.8)	Sharma et al. (2018), Pham et al. (2020), Pandey et al. (2022)			

(continued)

Table 4. (continued)

Trait	QTL		Full panel			No Meijer			Only Meijer			Literature	
	Name	chr (Mb in DMv6.1)	Haplotags	SNPs	Haplotags	SNPs	Haplotags	SNPs	Haplotags	SNPs	chr (Mb in DMv4.3 when available)	Reference	
Yellow Skin Color	YSC_C2_17 YSC_C3_12-C3_13	chr02 (62.62) chr03 (41.3-42.87)	C2_17_0010010	chr03_41301966. chr03_41302030		chr03_41301966. chr03_41302030. chr03_42866260							
Sprout Dormancy	SD_C2_6 SD_C9_11	chr02 (22.77) chr09 (51.47)		chr09_51478036. chr09_51478037. chr09_51478039	C2_6_01101	chr02_22766739, chr02_22766770				1, 3, 4, 5, 6, 7, 8, 11	Bradshaw et al. (2008), D'hoop et al. (2008), Urbany et al. (2011), D'hoop et al. (2014)		

Eye depth is a well-characterized trait and indeed we detected the well-known, large-effect QTL on chr10 in our full panel, spanning across the PotatoMASH loci C10_6 to C10_10 (4.74 to 51.2 Mb) peaking at 50.32 Mb. The deep eye (Eyd) phenotype was found to be associated with round tubers (Ro) (Li et al. 2005). The Eyd/eyd locus is located on chr10 and is closely linked with the major locus for Tuber Shape (Ro/ro). In the QTL detected here, the significant haplotags C10_8_0001111110, C10_9_011100000110110, and C10_10_01010000 and SNP alleles chr10_49148246, chr10_49148293, chr10_49148305, chr10_49148316, chr10_50323151, chr10_50323153, chr10_50323225, chr10_50323244, and chr10_50323263, were all associated both with deep eyes and round tubers, being their effects consistent with the genetics known (Li et al. 2005). In the opposite direction of effect, we found C10_9_000101000100000 associated with flat eyes and long tubers. We also detected a novel QTL for Eye depth on chr11 at C11_8 (8.19 Mb), with a significant SNP chr11_8190769, associated with deep eyes. No specific haplotag was detected.

Skin Smoothness is a complex trait, and many complementary factors influence tubers' skin texture, such as soil and climate. Earlier genetic studies by De Jong (1981) involved skin russeting as a phenotypic category, but in our panel no russeting phenotype was observed. In our study, only the skin texture was phenotyped, using a scoring scheme ranging from rough skin to smooth skin. Therefore, our study is the first to identify QTL for Skin Smoothness with no russeting. We detected three QTL for Skin Smoothness: Two QTL were detected on chr02: in PotatoMASH locus C2_8 (position 24.47 Mb) with the significant SNP chr02_24470953 of the haplotag C2_8_0000001000, and in locus C2_13 (position 28.96 Mb) with the significant haplotag C2_13_100010000. Both QTL were associated with smoother skin. The third QTL was found on chr06 at locus C6_3 (position 2.86 Mb), where the haplotag C6_3_0000101-00 was associated with rough skin, but no specific SNP allele underlying this haplotag was significant.

We detected a QTL for Sensitivity to Common Scab on chr06 at locus C6_18 (position 48.3 Mb). The significant haplotag C6_18_01111111110 was associated with susceptibility to Common Scab. This haplotag allele was present in only three individuals of the sub-population "no Meijer". The specific SNP for this haplotag (the only SNP not shared by the other haplotags in C6_18) was chr06_48297069 but was not statistically significant, most likely due to high missing data in this position (70%). We present this marker allele here, as a potential source for negative selection in future breeding, but further investigation needs to be done for validation.

Cooking Type is a complex trait. Previous studies revealed multiple QTL on multiple chromosomes: ch01, ch02, ch06, ch09, ch10, and ch011 (Kloosterman 2006; D'hoop et al. 2008; D'hoop et al. 2014). Our study is the first to report a QTL on chr08, at locus C8_21 (position 51.08 Mb), which was detected only for the "only Meijer" sub-population, with the significant haplotag C8_21_0000000000 associated with flouriness. No SNP was significant.

We detected 5 QTL for Chipping Color measured after three storage conditions (Table 1). For Chipping Color after storage at 8°C for 4 months before crisping, we identified 1 novel QTL on chr12 at locus C12_11 (position 9.26 Mb) with the full panel. The significant SNP allele, chr12_9258876, was associated with the dark color of crisps. This SNP allele is shared by a few haplotags, and none of these haplotags was significant.

For Chipping Color after storage at 8°C for 6 months before crisping, we detected 1 QTL on chr10 at locus C10_14 (position 55.08 Mb) with the sub-population "no Meijer". The significant

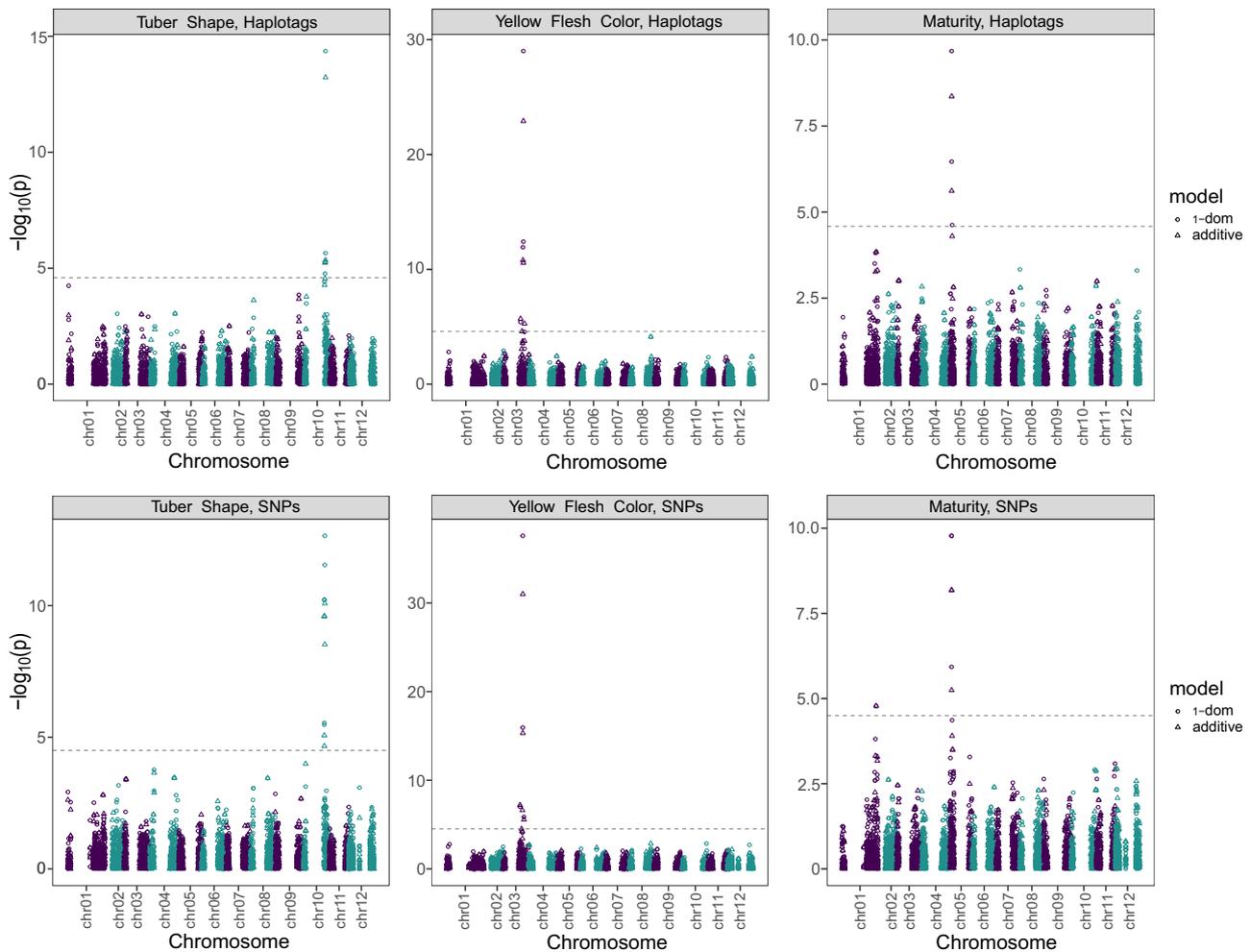


Fig. 7. Manhattan plots for the reference traits: Tuber Shape, Yellow flesh color, and Maturity. Top: analysis with Haplotype data. Bottom: analysis with SNP data.

SNP allele, chr10_55081844, with a positive effect was associated with a light, pure color of crisps. This SNP allele is shared by two haplotags, but none of the haplotags was significant. A previous work with tetraploid clones collected from the breeding program in Teagasc (Ireland) for “off the field” fry color, detected a large effect QTL on chr10 peaking at 56.16 Mb in DMv6.1 (55.28 Mb in DMv4.3) (Byrne et al. 2020). Our significant SNP allele at chr10 is at 55,081,844 bp is approximately 1 Mb distance from the one identified by Byrne et al (2020).

Three additional novel QTLs for Chipping Color were detected for storage at lower temperature (4°C) for 6 months. One QTL mapped on chr03 at locus C3_17 (position 46.06 Mb) with the sub-population “no Meijer” with the significant haplotag C3_17_000010 associated with dark color of crisps. Two other QTL were detected on this same locus for low Yield and Total Tuber Number, but with a different haplotag, C3_17_011000. Related to this, we found a negative correlation between Yield and Chipping Color 2.4 ($r^2 = -0.23$). We observed that only a small portion of the population (~4%) was heterozygous for those 2 alleles, possibly affecting this correlation, but we did not find a significant correlation between Chipping Color and Tuber Number. The second QTL associated with a light, pure color of crisps was detected on chr07 at locus C7_23 (position 50.94 Mb) with the sub-population “only Meijer”, specifically with the significant SNPs alleles chr07_50949343 and chr07_50949356. Those two SNP alleles are

in complete LD but are dispersed in many haplotags and no haplotag resulted statistically significant. The third QTL was detected on chr12 at locus C12_9 (position 7.44 Mb) using the full panel, with the significant haplotag C12_9_0010010001100 associated with a darker color of crisps. No SNP allele was significant.

It is useful to remember that the germplasm panel is derived from several independent commercial potato breeding programmes. We did not do an extensive analysis of each population source separately, but the fact that some QTL were exclusively discovered in one or the other of the “sub-populations”, suggests that beneficial alleles in one population may have the potential to augment genetic gain in populations lacking those alleles (alternatively, in some populations, during the breeding efforts, those alleles were successfully purged, or simply never possessed some undesired effect alleles). For example, the 4 significant SNP alleles and haplotags, chr09_61254396, chr09_62045322, C9_21_000100010100, and C9_23_000001, all associated with the undesirable trait of deep eyes in the QTL EYE_C9_21-C9_23, were only found in the “only Meijer” sub-population. Those markers co-segregate in the same 25 individuals (~13%), and it is possible that they all originated from the same source with deep eyes. On the other hand, in the “no Meijer” population, the SNP chr09_61254396 is present in 11 (~3%) individuals, the SNP chr09_62045322 is only present in 7 individuals (~2%), the haplotag C9_21_000100010100 is present in 9 individuals (2.4%) and

C9_23_000001 is not present at all. This suggests that the 2 sub-populations do not share the same ancestor or that the sub-population “no Meijer” have successfully selected against this negative allele. Another example is in the significant haplotag C6_18_01111111110 associated with susceptibility to common Scab that is only present in three individuals in the “no Meijer” sub-population.

Differences in QTL detected with SNPs vs haplotags

Fourteen of the QTL were detected by haplotags only, 13 QTL were identified with SNP data only, and 10 QTL were discovered with both SNPs and haplotags. To gain a better understanding of the ability to detect QTL with either SNPs or haplotags, we manually re-examined all individual QTL. We observed that in most cases of QTL detected with the haplotags only, the significant haplotag presents a specific composition of SNPs, but each individual SNP is dispersed across multiple haplotags of different SNP compositions. In Table 5, we present 4 examples of this phenomenon. This is also visible when looking at the dosage effect of the markers. One example for this is the new QTL discovered for Skin Smoothness, where the significant haplotag C6_3_0000101-00 has negative effect, while none of the underlying SNPs have a significant effect nor the other haplotags composed by the same SNPs (Fig. 8). In the opposite scenario of the QTL detected only with the SNP data, we observed that the significant SNP was shared in many haplotags (Table 6), which have a lower frequency in the population than the frequency of the significant SNP. To understand this phenomenon, we looked at the minor allele frequency of both SNPs and haplotags and observed that the minor allele frequency in the case of the SNPs is greater than 1% for most SNPs. When looking at haplotags, the frequency of individual haplotypes is much lower, with approximately 1,200 of the haplotags have a frequency below 1% (Fig. 9).

Discussion

Genetic improvement of potato at the diploid level is experiencing a resurgence, largely driven by the use of alleles that can

overcome the gametophytic self-incompatibility system in diploid material, allowing the development of strategies to rapidly accumulate and fix traits in a manner not possible at the tetraploid level. The primary goal of this study was to genetically characterize a large pool of diploid potato breeding material that is at the foundation of the diploid breeding efforts of several commercial breeding programmes that are engaged in a collaborative initiative towards innovative potato breeding schemes, combining the analytical breeding strategy (Chase 1963), which makes use of diploids to facilitate genetic studies and selection before returning to the tetraploid level through interploidy crosses, with self-compatibility.

Phenotypic data

The diploid clones used in this study represent a very diverse collection and the commercial traits display a wide range of phenotypic trait values. In Table 2, we show that for each trait the full scale of trait values was observed, indicative of primitive material, primary dihaploids with compromised vigor, as well as elite material. On average poor trait values were observed for quality traits such as discoloration due to Enzymatic Browning and Chipping Color; notoriously difficult traits to improve. Most clones displayed the firm Cooking Type, which is negatively correlated with (and largely due to) low values for Dry Matter Content. Canopy development, Tuber Shape and Sprout Dormancy are among the most diverse traits. Most clones had relatively round and uniform Tuber Shape and late Maturity. Late maturity is considered beneficial to obtain an extended period of flowering, which facilitates making crosses during the breeding program, which may explain its prevalence in early-stage prebreeding material. However, early maturity is desirable for several market classes of potato. Since early maturity is largely controlled by a single large-effect quantitative trait locus, the effort to regain early maturity should be relatively easy as material advances to more commercial status over cycles of selection (Song and Endelman 2023).

Broad sense heritabilities varied mostly between moderate to high values, ranging from 50 to 90% across all traits. Sensitivity to Common Scab had, on average, the lowest H^2 across companies (50%), suggesting either low reproducibility of disease

Table 5. Four examples of QTL detected by one unique haplotag and not with any of its constituent SNPs. The number of individuals carrying each SNP/haplotag allele is indicated within brackets.

Trait	Significant haplotag (number of individuals)	Underlying SNPs (number of individuals)	Nontrait-associated haplotags sharing the same underlying SNPs (number of individuals)
Total tuber number	C11_12_000100000 (63)	chr11_37673981 (189)	C11_12_000101000 (5) C11_12_001100000 (11) C11_12_010100000 (1) C11_12_010101000 (138)
Skin Smoothness	C6_3_0000101-00 (9)	chr06_2681972 (549) chr06_2681990 (163) deletion_chr06_2681999 (489)	C6_3_0000100-00 (329) C6_3_000010000- (1) C6_3_0000100000 (182) C6_3_0000100001 (38) C6_3_0000110100 (29) 6_3_0011101-10 (148) C6_3_0100100000 (42)
Skin Smoothness	C2_13_100010000 (15)	chr02_28958095 (310)	C2_13_000010000 (164) C2_13_000011000 (11) C2_13_001011000 (126)
Chipping Color 2_4	C12_9_0010010001100 (214)	chr12_7435710 (464) chr12_7435746 (531) chr12_7435801 (423) chr12_7435802 (464)	C12_9_0010010011100 (7) C12_9_0010011001110 (38) C12_9_0010011001110 (108) C12_9_0011010011100 (11) C12_9_0000010000000 (64) C12_9_0001010000000 (8) C12_9_0001010100000 (97)

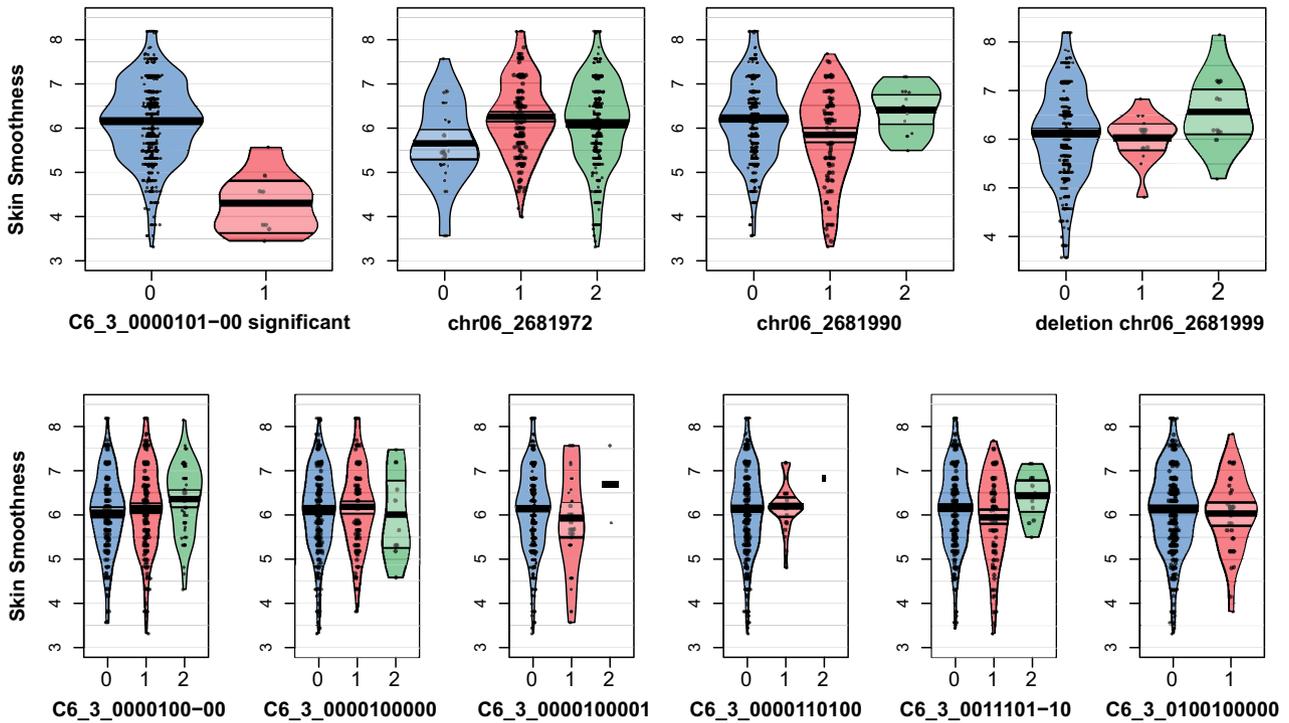


Fig. 8. Allele dosage in QTL C6_3 vs the effect on skin smoothness. Top: Allele dosage of the significant haplotag and the nonsignificant SNPs underlying this haplotag. Bottom: other six haplotags of this region that possess those SNPs but their combination in the haplotags was not associated with the trait.

Table 6. Six QTL detected with SNPs but not with the haplotag dataset, and all the haplotags composed by those SNPs. The number of individuals for each SNP/haplotag is given within brackets.

Trait	QTL	Significant SNPs (number of individuals)	Haplotags sharing this SNPs (number of individuals)
Dry matter content	DMC C3_30	chr03_59353418 (40)	C3_30_0000100 (12), C3_30_0110100 (20), C3_30_0110110 (2), C3_30_0111100 (3)
Chipping color 1_8	QDC1_8 C12_11	chr12_9258876 (281),	C12_13_0100000 (255), C12_13_0110000 (32)
Total tuber number	TTN C3_29	chr03_58369063 (458), chr03_58369113 (550)	C3_29_00100010 (409), C3_29_00101010 (5), C3_29_00110010 (69), C3_29_01110010 (29), C3_29_01110110 (2), C3_29_01111110 (293)
Total tuber number	TTN C3_7	chr03_37250410 (191)	C3_7_0001110 (21), C3_7_0011110 (143), C3_7_0101010 (29)
Eye depth	EYE C11_8	chr11_8190769 (239)	C11_8_00100111000 (84), C11_8_00100111100 (82), C11_8_01101111000 (73)
Chipping Color 2_4	QDC2_4_C7_23	chr07_50949343 (155), chr07_50949356 (156)	C7_23_000111000 (71), C7_23_010010000 (2), C7_23_010110000 (30), C7_23_010110010 (45), C7_23_010111000 (3)

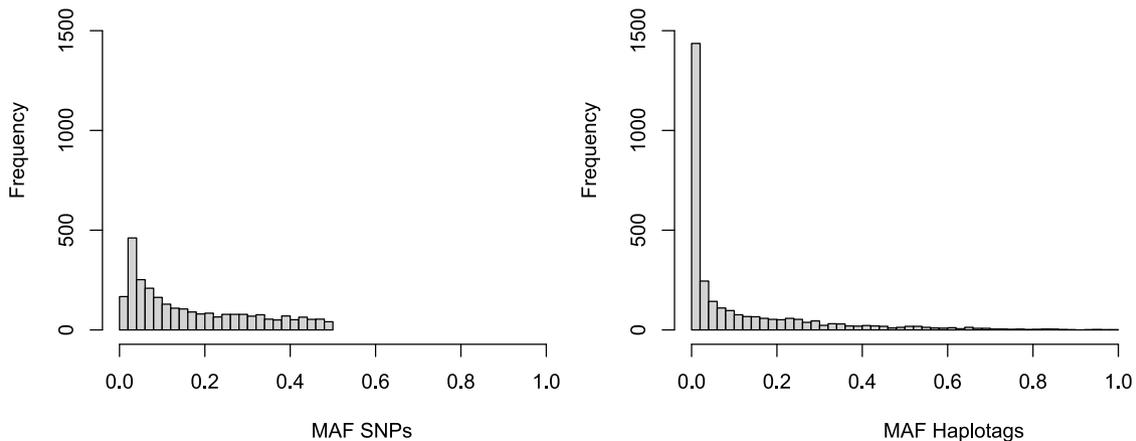


Fig. 9. Minor Allele frequency (MAF) distribution of 2,730 SNPs (left) and 2,995 haplotags (right).

development across years due to lack of exposure to the pathogen, environmental factors, or that different trial fields used across years are infested by different isolates. This is also visible with the performance of the controls over the years and between companies, where we can see a large variance in the scoring of Sensitivity to scab even in the same year and the same company (Supplementary File 6). Moderately low average H^2 values, ranging between 60 and 75%, are typical for traits such as Presentability of Tubers, Skin Smoothness, Skin Brightness, and Tuber Regularity, which have a somewhat ambiguous trait definition and a scale that is not objectively measurable, but rather a result of the so-called “breeders’ eye”. Despite the potential subjectivity of these scores, the H^2 values obtained suggest high repeatability. Genomic heritabilities were also calculated (Supplementary File 7) to catch a more accurate estimate than the average presented in Table 2. However, the results were much lower than expected, with, for example, heritability for tuber shape of 0.41 and 0.4. It could be that the tools available to calculate genomic heritability are not suited to the low number of marker sets we use in this study. This should be explored further in future analyses.

Correlation analysis of all the trait pairs was performed to examine associations between traits (Fig. 2). The traits of Skin Smoothness and Skin Brightness, albeit representing a subjective breeder’s score, show the highest correlation of 0.76, suggesting that the trait definitions are somewhat arbitrarily different, or share similar underlying aspects. Both traits also correlate with Yellow Skin Color (0.48 and 0.56), where darker skins imply thicker skin. Another pair of traits: Tuber Regularity and Presentability show a Pearsons correlation of 0.66. This is not unexpected because Regularity is an aspect within Presentability, along with Eye depth ($r^2 = 0.43$ and 0.46). Three processing traits related to Chipping Color also show high pairwise correlations (0.59, 0.71, 0.75), suggesting that the storage regime of tubers, causing cold sweetening, is less important than the initial Chipping quality at harvest. Byrne et al. (2020) made a similar observation on a similarly sized population of tetraploid breeding clones from a single commercial breeding programme. In this diploid gene pool, an unexpected positive correlation of 0.57 was found between Yield and Total Tuber Number. Such a correlation would be rather unexpected for a panel of varieties, selected for yield above a certain threshold. Maturity and Canopy development also show expected correlations where late maturity leads to bigger canopies at both stages (0.34, 0.41) and the Canopy-Yield correlation was 0.25 and 0.46. However, the negative correlation between Yield and Maturity is unexpected (-0.25). Canopy stage 1 and Canopy stage 2 correlates with Total Tuber Number (0.18 and 0.37), which could be due to a common plant architecture, where stronger above- and below-ground branching patterns or stems and stolons may contribute to larger canopy cover and tuber number.

In general, this “snapshot” of the extent of phenotypic diversity in this genepool suggests that variability exists for most important agronomic and quality traits, and further selection in all or individual parts of the panel is expected to allow improvement. In terms of the use of this material in strategies involving inbreeding, we also surveyed the material for the presence of diagnostic KASP markers for the *Sli* locus as described by Clot et al. (2020), and found that *Sli* is relatively common in the material, present in 17.5% of the individuals (data not shown). The presence of this locus throughout the material means that efforts to introgress it from exotic sources, with the accompanying issues such as increased timescales of the breeding process and potential linkage drag of unfavorable loci, are unnecessary.

The germplasm panel is composed of material from six different breeding programmes from The Netherlands, Germany, Denmark, and France. These programmes have a mixture of market class targets, including starch, table (domestic and export), processing (crisps and French fries), and specialty (e.g. salad) potatoes. Genome wide marker studies in the European cultivated potato genepool have previously shown some stratification for geographic origin of breeding programme and utility class (Uitdewilligen et al. 2015). However, this assessment is not strong, probably due to the relatively recently shared pool of progenitors of the material. When we examined the population structure of our panel, we found two highly distinct groups, one characterized by material from the breeding company Meijer and the other comprising all other material. This is interesting given the fact that diploid material from many Dutch breeding companies, including Meijer, has often originated from the diploid prebreeding programme at WUR, whereas these groups were quite distinct in our analysis. Whilst we treated the panel as two subpopulations for some analyses on this basis, there was also some visible stratification (along the second principal coordinate in Fig. 5) between the other companies. This did not correspond to either market class or geographical location of the programme.

In this experiment, different sets of diploid potato clones were grown at different locations, and trait values were evaluated by different observers. The same 2 or 4 control varieties were included in each trail to allow a fair comparison of phenotypic values. D’hoop et al. (2011) already compared phenotypic means from different experimental designs. One of their approaches made use of historical observations, on company specific candidate varieties, retrieved from breeder’s field books. In their other approach, all clones (now released as variety) were grown together in a balanced trial with two locations (sandy and clay soil), both with two replications. That study showed that either a single-year balanced field trial, or multiyear–multilocation breeders’ records yield robust phenotypic information that can be used in a genome-wide association study (D’hoop et al. 2011). In this study, the differences between company specific panels of diploids were controlled with structure. In particular, the material offered by Meijer was also treated separately. Regarding location differences, in terms of plant material and environmental conditions, the locations may have added variance to the error term, and we may have lost some power but also false negatives.

GWAS with an amplicon sequencing technique and short read haplotypes

As mentioned earlier, the primary goal of the study was to characterize the foundation breeding material that will contribute to future diploid breeding approaches focused on trait fixation through inbreeding in potato. We focused on agronomic and quality traits that breeders routinely monitor during selection in order to develop a capacity to increase the effectiveness of this process using genome-based methodologies in future. Based on the marker-inferred population structure, we performed 6 QTL discovery analysis: for the entire population, the 2 sub-populations and with the 2 marker sets, SNPs and haplotags. We identified a total of 37, nonredundant QTL. Discovery of these QTL in the population gives us the potential to manipulate their configuration in future material, for instance, to accumulate and fix beneficial alleles or eliminate detrimental alleles. One potential problem with this approach in previous GWAS studies is that associated SNP markers, whilst associated with traits, may still be dispersed amongst different haplotype blocks in which the effective allele underlying the trait is also variably present. In addition to QTL discovery within this prebreeding panel, the study also allowed

us to further explore whether the multiallelic discrimination power of PotatoMASH short read haplotypes (haplotags) can resolve this. The general approach certainly seems promising.

An average of approximately 9 haplotags was observed per PotatoMASH locus (range 2–30). This exceeds the average number we previously detected in a panel of tetraploid breeding clones (average of 6, range 2 to 14) (Leyva-Pérez et al. 2022). This greater diversity may result from the wider set of utility classes being surveyed, and because at least some of the diploid material is the result of introgression breeding for resistance loci from wild species. In our previous study on the tetraploid panel, we empirically illustrated the hypothesis that haplotags better represent the actual underlying allelic variation at a locus and may offer advantages over bi-allelic SNPs for QTL detection. We posited that this was due to better representation of regions of identity by descent harboring the causal allelic variant of the QTL, and that haplotags are more likely to be in LD with allelic variants of genes with an effect on trait values. Conversely, some or all of the component SNPs may be dispersed across multiple haplotypes, some of which are not in LD with the effective QTL allele.

In our study, 14 of the QTL were detected by haplotags only, 13 QTL were detected with SNP data only, and 10 QTL were detected with both datasets. We found that in most cases, these differences were due to the genetic architecture: the first situation of QTL detected with the haplotags data only, which was our original expectation, occurs when more unique haplotags are in greater LD with QTL causal alleles, whereas the bi-allelic SNPs were dispersed across multiple haplotags, some of which were not in LD with the effective QTL allele (Table 5, Fig. 8).

A marker allele has to be present at sufficient frequency to support association via a statistical test. We observed much higher number of rare haplotags in our population than low-frequency SNPs (Fig. 9). Therefore, for some traits, we will fail to identify statistically significant associations with haplotags with very low frequencies, which can result in the opposite phenomenon, where a QTL is only identified with SNP data. Only when a haplotag coincides with a haplotype specific SNPs their power to detect a QTL is equal.

Technical limitations affecting the power of QTL detection

The ability to detect QTL with one marker type over the other also depends on the analytical tools we use. In specific cases, we encountered that some features of our genotyping platform limited the QTL detection either with the haplotag dataset or the SNP dataset

- 1) PotatoMASH regions are designed to be single copy based on the DM reference genome. However, if some potato clones possess duplications with allelic variants of these regions, the reads of both copies may map back to the single copy reference genome sequence region during read mapping (since DM is also used as the reference sequence). This may be difficult to see in the SNP data, but can cause more than the expected two haplotypes in individuals with the duplication. SMAP will reject calling haplotags in these cases, assigning a missing value to that individual, and effectively filtering out instances in which this occurs. For example, in region C8_18, we detected a QTL for Enzymatic Browning in the Meijer population with the haplotag data, but not with SNP data. The SMAP locus correctness score of C8_18 was 60, suggesting the existence of additional read variation mapping to that locus (more than 2 haplotypes) that could not be explained as a single bi-allelic locus for 30% of the individuals.

Thus, the SNP calling in this locus would be wrong, but SMAP correctly rejected calling haplotags for this locus in those individuals and the association for Enzymatic Browning in this locus is based on the remaining 60% of the population.

- 2) A SNP allele is significant and is in LD with 1 haplotag but the haplotag is not significant. Close inspection of the two QTL where this occurred showed that the significant SNP allele was present in a low number of individuals and associated as a minor effect QTL allele, for which the logarithm of the odds-score (LOD-score) was just above the threshold (Supplementary File 10, doi:10.6084/m9.figshare.26163616.v2). In this case, if some individuals have missing data for that locus, the haplotag cannot be called and results in a missing value. Therefore, the LOD-score of the specific haplotag may not pass the significance threshold. This could affect in two ways, either that asymmetric missing data play a role, where one allele suffers more missing data or in a symmetric way that both alleles suffer the missing data, and the amount of data is simply too low to form a strong statistical test. One example for that is the SNP chr02_20959691 that was significant for Tuber Length at locus C2_4, but there was a lot of missing data for both markers sets in this position, so the association is based on 127 individuals. The haplotag C2_4_0001000000 is specific to this SNP but was not significant; we observed that out of those 127 individuals, SMAP failed to call haplotags for 11 individuals in this locus and this probably affected the mean phenotypic score of the allelic categories, and the association that was already weak in the first place, was lost when using haplotag data.

Concluding remarks

Although we do not view the number of SNPs we used as “optimal” for GWAS, we were, in fact, testing the hypothesis that haplotags would detect loci not detected by the component SNPs that were used to derive them. We conclude that short read haplotags can detect additional QTL not detectable by individual SNPs, but that, for the various reasons outlined, the opposite is also true. Thus, the approach we adopted, utilizing both sets of data (even though 1 is derived from the other) is the most optimal for QTL detection. One obstacle we faced when using haplotags for the GWAS is that we had to use them as “pseudoSNPs” to employ standard analysis software, and in this study, we have not explored the full potential of the multiallelic nature of the haplotags. From both a genetic and practical breeding point of view, it would be interesting to gain a better understanding of the nature of allelic interactions within and between loci. Recently, Thérèse Navarro et al. (2022) developed a software, mpQTL, for QTL analysis at any ploidy level under biallelic and multiallelic models, but for multiparental populations. Their approach was demonstrated with simulated data of short-range haplotypes of autotetraploid multiparental populations. Combining approaches like this with real-world data of higher genetic diversity panel like the current study, will give insights into the genetic control of traits in highly heterozygous systems.

The increased precision offered by the new paradigm in potato breeding means that genome-based tools will become more effective in augmenting selection, allowing the “shepherding” and subsequent fixation of multiple desirable alleles into single genotypes (or the elimination of detrimental alleles). To this end, we have characterized a panel partially representative of the foundational genepool of the future of diploid breeding across several potato-breeding programmes involved in collaborative efforts in this area.

The availability of low-cost, medium-density genotyping approaches capable of generating genome-wide multiallelic marker data in potato (e.g. PotatoMASH in this study, or the Potato DArTag EiB 1.0 generated by CGIAR <https://excellenceinbreeding.org/>) demonstrate that it is becoming feasible to implement such systems in breeding selection, implying the routine application to thousands of individuals per annum. These marker panels are also amenable to the addition of trait-specific markers, such as those targeting disease and pest resistance loci. We envisage a future where such assays can be used for a combination of marker assisted selection, genomic selection and monitoring the genomic constitution of inbred lines in terms of global and local homo/heterozygosity in potato breeding.

Data availability statement

All data necessary for confirming the conclusions of the article are present within the article's text, figures, and tables and the supplementary files.

Scripts and intermediate files for the PotatoMASH bioinformatics pipeline are also available at <https://doi.org/10.6084/m9.figshare.c.6926560>. The code to reproduce the results and figures of this article are available at <https://doi.org/10.6084/m9.figshare.c.6937662>.

[Supplemental material](#) available at G3 online.

Acknowledgements

We thank the members of the public-private partnership “A new method for potato breeding: the ‘Fixation-Restitution’ approach” and SusCrop ERANET funded project “DIFFUGAT: Diploid Inbreds For Fixation, and Unreduced Gametes for Tetraploidy” (Averis Seeds B.V., Bejo Zaden B.V., Danespo A/S, Germicopa, Den Hartigh B.V., SaKa Pflanzenzucht GmbH & Co. KG, C. Meijer B.V., and Teagasc) for providing their support. This research was carried out using the Teagasc high-performance computing cluster and storage systems, and the support of Dr. Paul Cormican is greatly appreciated.

Funding

The research was financially supported by the Dutch Topsector Horticulture & Starting Materials (project “A new method for potato breeding: the ‘Fixation-Restitution’ approach”; grant number TU18075). Within the Topsector, private industry, knowledge institutes and the government are working together on innovations for sustainable production of safe and healthy food and the development of a healthy green environment. L.V. was supported by a Teagasc Walsh Scholarship and the DIFFUGAT project (“Diploid Inbreds For Fixation, and Unreduced Gametes for Tetraploidy”) and the Department of Agriculture Food and the Marine (DAFM) under the ERA-NET Cofund SusCrop (Grant No. 771134), being part of FACCE-JPI.

Conflicts of interest

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this publication.

Author contributions

D.M., H.v.E., V.P., D.G., G.L.-J., J.-D.D., A.N.O.R., E.H.R.S., and R.H. obtained funding, conceived and designed the study. S.W.,

G.L.-J., J.-D.D., C.E., A.N.O.R., E.H.R.S., and R.H. managed the field trials, collected tissue samples and phenotypic data. L.V. isolated DNA, constructed PotatoMASH libraries, analyzed genomic and phenotypic data, performed GWAS and drafted the initial manuscript. M.d.l.O.L.-P. developed and supervised PotatoMASH molecular and bioinformatics pipeline. D.M., S.B., and A.K. advised on statistical and GWAS analysis. S.B., C.C., and T.R. advised on bioinformatics processing of the genomic data. T.R. provided the SMAP software. D.M., H.v.E., S.B., C.C., T.R., and M.d.l.O.L.-P. edited the manuscript. D.M., H.v.E., and R.G.F.V. supervised the research. All authors have read and agreed to the published version of the manuscript.

Literature cited

- Baldwin S, Dodds K, Auvray B, Genet R, Macknight R, Jacobs J. 2011. Association mapping of cold-induced sweetening in potato using historical phenotypic data. *Ann Appl Biol.* 158(3):248–256. doi:10.1111/j.1744-7348.2011.00459.x.
- Bates D, Martin M, Ben B, Steve W. 2015. Fitting linear mixed-effects models using lme4. *J Stat Softw.* 67(1):1–48. doi:10.18637/jss.v067.i01.
- Bonierbale MW, Plaisted RL, Tanksley S. 1988. RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. *Genetics.* 120(4):1095–1103. doi:10.1093/genetics/120.4.1095.
- Bradshaw JE. 2022. Breeding diploid F1 hybrid potatoes for propagation from botanical seed (TPS): comparisons with theory and other crops. *Plants.* 11(9):1121. doi:10.3390/plants11091121.
- 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*). *Theor Appl Genet.* 116(2):193–211. doi:10.1007/s00122-007-0659-1.
- Braun SR, Endelman JB, Haynes KG, Jansky SH. 2017. Quantitative trait loci for resistance to common scab and cold-induced sweetening in diploid potato. *Plant Genome.* 10(3):plantgenome2016.2010.0110. doi:10.3835/plantgenome2016.10.0110.
- Brown C, Kim T, Ganga Z, Haynes K, De Jong D, Jahn M, Paran I, De Jong W. 2006. Segregation of total carotenoid in high level potato germplasm and its relationship to beta-carotene hydroxylase polymorphism. *Am J Potato Res.* 83(5):365–372. doi:10.1007/BF02872013.
- Byrne S, Meade F, Mesiti F, Griffin D, Kennedy C, Milbourne D. 2020. Genome-wide association and genomic prediction for fry color in potato. *Agronomy.* 10(1):90. doi:10.3390/agronomy10010090.
- Chase SS. 1963. Analytic breeding in *Solanum tuberosum* L.: a scheme utilizing parthenotes and other diploid stocks. *Can J Genet Cytol.* 5(4):359–363. doi:10.1139/g63-049.
- Clot CR, Klein D, Koopman J, Schuit C, Engelen CJM, Hutten RCB, Brouwer M, Visser RGF, Juranić M, van Eck HJ. 2024. Crossover shortage in potato is caused by StMSH4 mutant alleles and leads to either highly uniform unreduced pollen or sterility. *Genetics.* 226(1). doi:10.1093/genetics/iyad194.
- Clot CR, Polzer C, Prodhomme C, Schuit C, Engelen CJ, Hutten RC, Van Eck HJ. 2020. The origin and widespread occurrence of Sli-based self-compatibility in potato. *Theor Appl Genet.* 133(9): 2713–2728.
- Clot CR, Vexler L, de LO, Leyva-Perez M, Bourke PM, Engelen CJM, Hutten RCB, van de Belt J, Wijnker E, Milbourne D, et al. 2024. Identification of two mutant JASON-RELATED genes associated with unreduced pollen production in potato. *Theor Appl Genet.* 137(4):79. doi:10.1007/s00122-024-04563-7.
- D'hoop BB, Keizer PLC, Paulo MJ, Visser RGF, van Eeuwijk FA, van Eck HJ. 2014. Identification of agronomically important QTL in

- tetraploid potato cultivars using a marker–trait association analysis. *Theor Appl Genet.* 127(3):731–748. doi:[10.1007/s00122-013-2254-y](https://doi.org/10.1007/s00122-013-2254-y).
- 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. doi:[10.1007/s10681-007-9565-5](https://doi.org/10.1007/s10681-007-9565-5).
- 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 Res.* 54(2):157–181. doi:[10.1007/s11540-011-9186-1](https://doi.org/10.1007/s11540-011-9186-1).
- da Silva Pereira G, Mollinari M, Qu X, Thill C, Zeng Z-B, Haynes K, Yencho GC. 2021. Quantitative trait locus mapping for common scab resistance in a tetraploid potato full-sib population. *Plant Dis.* 105(10):3048–3054. doi:[10.1094/PDIS-10-20-2270-RE](https://doi.org/10.1094/PDIS-10-20-2270-RE).
- da Silva Pereira G, Mollinari M, Schumann MJ, Clough ME, Zeng Z-B, Yencho GC. 2021. The recombination landscape and multiple QTL mapping in a *Solanum tuberosum* cv. ‘atlantic’-derived F1 population. *Heredity (Edinb).* 126(5):817–830. doi:[10.1038/s41437-021-00416-x](https://doi.org/10.1038/s41437-021-00416-x).
- De Jong H. 1981. Inheritance of russeting in cultivated diploid potatoes. *Potato Res.* 24:309–313.
- Díaz P, Sarmiento F, Mathew B, Ballvora A, Mosquera Vásquez T. 2021. Genomic regions associated with physiological, biochemical and yield-related responses under water deficit in diploid potato at the tuber initiation stage revealed by GWAS. *PLoS One.* 16(11):e0259690. doi:[10.1371/journal.pone.0259690](https://doi.org/10.1371/journal.pone.0259690).
- Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol.* 14(8):2611–2620. doi:[10.1111/j.1365-294X.2005.02553.x](https://doi.org/10.1111/j.1365-294X.2005.02553.x).
- Fox J, Weisberg S. 2019. *An R Companion to Applied Regression*. Thousand Oaks, CA: Sage.
- Hosaka K, Hanneman RE. 1998. Genetics of self-compatibility in a self-incompatible wild diploid potato species *Solanum chacoense*. 1. Detection of an S locus inhibitor (Sli) gene. *Euphytica.* 99(3):191–197. doi:[10.1023/A:1018353613431](https://doi.org/10.1023/A:1018353613431).
- Hurtado-Lopez PX, Tessema BB, Schnabel SK, Maliepaard C, Van der Linden CG, Eilers PHC, Jansen J, van Eeuwijk FA, Visser RGF. 2015. Understanding the genetic basis of potato development using a multi-trait QTL analysis. *Euphytica.* 204(1):229–241. doi:[10.1007/s10681-015-1431-2](https://doi.org/10.1007/s10681-015-1431-2).
- Jansky SH, Charkowski AO, Douches DS, Gusmini G, Richael C, Bethke PC, Spooner DM, Novy RG, De Jong H, De Jong WS, et al. 2016. Reinventing potato as a diploid inbred line-based crop. *Crop Sci.* 56(4):1412–1422. doi:[10.2135/cropsci2015.12.0740](https://doi.org/10.2135/cropsci2015.12.0740).
- Kao TH, McCubbin AG. 1996. How flowering plants discriminate between self and non-self-pollen to prevent inbreeding. *Proc Natl Acad Sci USA.* 93(22):12059–12065. doi:[10.1073/pnas.93.22.12059](https://doi.org/10.1073/pnas.93.22.12059).
- Klaassen MT, Willemsen JH, Vos PG, Visser RG, van Eck HJ, Maliepaard C, Trindade LM. 2019. Genome-wide association analysis in tetraploid potato reveals four QTLs for protein content. *Mol Breeding.* 39(10–11):1–12. doi:[10.1007/s11032-019-1070-8](https://doi.org/10.1007/s11032-019-1070-8).
- Kloosterman B. 2006. Transcriptomic analysis of potato tuber development and tuber quality traits using microarray technology. Dissertation. Wageningen University. <https://library.wur.nl/WebQuery/wda/1806317>.
- Kloosterman B, Abelenda JA, MdMC G, Oortwijn M, de Boer JM, Kowitwanich K, Horvath BM, van Eck HJ, Smaczniak C, Prat S, et al. 2013. Naturally occurring allele diversity allows potato cultivation in northern latitudes. *Nature.* 495(7440):246–250. doi:[10.1038/nature11912](https://doi.org/10.1038/nature11912).
- Kruijer W, Boer MP, Malosetti M, Flood PJ, Engel B, Kooke R, Keurentjes JJB, van Eeuwijk FA. 2014. Marker-based estimation of heritability in immortal populations. *Genetics.* 199(2):379–398. doi:[10.1534/genetics.114.167916](https://doi.org/10.1534/genetics.114.167916).
- Kruijer W, White I. 2023. Package ‘heritability’: marker-based estimation of heritability using individual plant or plot data. <https://cran.r-project.org/web/packages/heritability/heritability.pdf>.
- Lenth R, Singmann H, Love J, Buerkner P, Herve M. 2021. Emmeans: estimated marginal means, aka least-squares means. R Package. <https://github.com/rvleenth/emmeans>.
- Leyva-Pérez M de la O, Vexler L, Byrne S, Clot CR, Meade F, Griffin D, Ruttink T, Kang J, Milbourne D. 2022. PotatoMASH—a low cost, genome-scanning marker system for use in potato genomics and genetics applications. *Agronomy.* 12(10):2461. doi:[10.3390/agronomy12102461](https://doi.org/10.3390/agronomy12102461).
- Li L, Paulo M-J, van Eeuwijk F, Gebhardt C. 2010. Statistical epistasis between candidate gene alleles for complex tuber traits in an association mapping population of tetraploid potato. *Theor Appl Genet.* 121(7):1303–1310. doi:[10.1007/s00122-010-1389-3](https://doi.org/10.1007/s00122-010-1389-3).
- Li X-Q, De Jong H, De Jong DM, De Jong WS. 2005. Inheritance and genetic mapping of tuber eye depth in cultivated diploid potatoes. *Theor Appl Genet.* 110(6):1068–1073. doi:[10.1007/s00122-005-1927-6](https://doi.org/10.1007/s00122-005-1927-6).
- Lindhout P, Meijer D, Schotte T, Hutten RCB, Visser RGF, van Eck HJ. 2011. Towards F1 hybrid seed potato breeding. *Potato Res.* 54(4):301–312. doi:[10.1007/s11540-011-9196-z](https://doi.org/10.1007/s11540-011-9196-z).
- Malosetti M, Van der Linden C, Vosman B, Van Eeuwijk F. 2007. A mixed-model approach to association mapping using pedigree information with an illustration of resistance to *Phytophthora infestans* in potato. *Genetics.* 175(2):879–889. doi:[10.1534/genetics.105.054932](https://doi.org/10.1534/genetics.105.054932).
- Manrique-Carpintero NC, Coombs JJ, Cui Y, Veilleux RE, Buell CR, Douches D. 2015. Genetic map and QTL analysis of agronomic traits in a diploid potato population using single nucleotide polymorphism markers. *Crop Sci.* 55(6):2566–2579. doi:[10.2135/cropsci2014.10.0745](https://doi.org/10.2135/cropsci2014.10.0745).
- McCord PH, Sosinski BR, Haynes KG, Clough ME, Yencho GC. 2011. Linkage mapping and QTL analysis of agronomic traits in tetraploid potato (*Solanum tuberosum* subsp. *tuberosum*). *Crop Sci.* 51(2):771–785. doi:[10.2135/cropsci2010.02.0108](https://doi.org/10.2135/cropsci2010.02.0108).
- Mosquera T, Alvarez MF, Jiménez-Gómez JM, Mukhtar MS, Paulo MJ, Steinemann S, Li J, Draffehn A, Hofmann A, Lübeck J. 2016. Targeted and untargeted approaches unravel novel candidate genes and diagnostic SNPs for quantitative resistance of the potato (*Solanum tuberosum* L.) to *Phytophthora infestans* causing the late blight disease. *PLoS One.* 11(6):e0156254. doi:[10.1371/journal.pone.0156254](https://doi.org/10.1371/journal.pone.0156254).
- Pandey J, Scheuring DC, Koym JW, Vales MI. 2022. Genomic regions associated with tuber traits in tetraploid potatoes and identification of superior clones for breeding purposes. *Front Plant Sci.* 13:952263. doi:[10.3389/fpls.2022.952263](https://doi.org/10.3389/fpls.2022.952263).
- Parra-Galindo MA, Soto-Sedano JC, Mosquera-Vásquez T, Roda F. 2021. Pathway-based analysis of anthocyanin diversity in diploid potato. *PLoS One.* 16(4):e0250861. doi:[10.1371/journal.pone.0250861](https://doi.org/10.1371/journal.pone.0250861).
- Pham GM, Hamilton JP, Wood JC, Burke JT, Zhao H, Vaillancourt B, Ou S, Jiang J, Buell CR. 2020. Construction of a chromosome-scale long-read reference genome assembly for potato. *GigaScience.* 9(9):giaa100. doi:[10.1093/gigascience/giaa100](https://doi.org/10.1093/gigascience/giaa100).
- Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics.* 155(2):945–959. doi:[10.1093/genetics/155.2.945](https://doi.org/10.1093/genetics/155.2.945).
- Prodhomme C, Vos PG, Paulo MJ, Tammes JE, Visser RG, Vossen JH, van Eck HJ. 2020. Distribution of P1 (D1) wart disease resistance in potato germplasm and GWAS identification of haplotype-specific SNP markers. *Theor Appl Genet.* 133(6):1859–1871. doi:[10.1007/s00122-020-03559-3](https://doi.org/10.1007/s00122-020-03559-3).

- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M, Bender D, Maller J, Sklar P, Bakker P, Daly M, et al. 2007. Plink: a tool set for whole-genome association and population-based linkage analyses. *Am J Human Genet.* 81(3):559–575. doi:10.1086/519795.
- Raj A, Stephens M, Pritchard JK. 2014. fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics.* 197(2):573–589. doi:10.1534/genetics.114.164350.
- Rak K, Bethke PC, Palta JP. 2017. QTL mapping of potato chip color and tuber traits within an autotetraploid family. *Mol Breeding.* 37(2):15. doi:10.1007/s11032-017-0619-7.
- Rosyara UR, Jong D, Douches WS, Endelman DS, B J. 2016. Software for genome-wide association studies in autopolyploids and its application to potato. *Plant Genome.* 9(2):plantgenome2015.2008.0073. doi:10.3835/plantgenome2015.08.0073.
- Schaumont D, Veeckman E, Van der Jeugt F, Haegeman A, van Glabeke S, Bawin Y, Lukaszewicz J, Blugeon S, Barre P, Leyva-Pérez MDLO, et al. 2022. Stack Mapping Anchor Points (SMAP): a versatile suite of tools for read-backed haplotyping. bioRxiv:2022.2003.2010.483555. <https://doi.org/10.1101/2022.03.10.483555>, preprint: not peer reviewed.
- Schönhals EM, Ding J, Ritter E, Paulo MJ, Cara N, Tacke E, Hofferbert H-R, Lübeck J, Strahwald J, Gebhardt C. 2017. Physical mapping of QTL for tuber yield, starch content and starch yield in tetraploid potato (*Solanum tuberosum* L.) by means of genome wide genotyping by sequencing and the 8.3 K SolCAP SNP array. *BMC Genomics.* 18(1):642. doi:10.1186/s12864-017-3979-9.
- 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 (Bethesda).* 8(10):3185–3202. doi:10.1534/g3.118.200377.
- Śliwka J, Wasilewicz-Flis I, Jakuczun H, Gebhardt C. 2008. Tagging quantitative trait loci for dormancy, tuber shape, regularity of tuber shape, eye depth and flesh colour in diploid potato originated from six *Solanum* species. *Plant Breeding.* 127(1):49–55. doi:10.1111/j.1439-0523.2008.01420.x.
- Song L, Endelman JB. 2023. Using haplotype and QTL analysis to fix favorable alleles in diploid potato breeding. *Plant Genome.* 16(2):e20339. doi:10.1002/tpg2.20339.
- 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. *Theor Appl Genet.* 130(10):2091–2108. doi:10.1007/s00122-017-2944-y.
- Thérèse Navarro A, Tumino G, Voorrips RE, Arens P, Smulders MJ, Van De Weg E, Maliepaard C. 2022. Multiallelic models for QTL mapping in diverse polyploid populations. *BMC Bioinformatics.* 23(1):67.
- Thorup T, Tanyolac B, Livingstone K, Popovskiy S, Paran I, Jahn M. 2000. Candidate gene analysis of organ pigmentation loci in the Solanaceae. *Proc Natl Acad Sci USA.* 97(21):11192–11197. doi:10.1073/pnas.97.21.11192.
- Uitdewilligen JG, Wolters A-MA, D'hoop BB, Borm TJ, Visser RG, van Eck HJ. 2015. Correction: A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. *PLoS One.* 10(10):e0141940. doi:10.1371/journal.pone.0062355.
- Urbany C, Stich B, Schmidt L, Simon L, Berding H, Junghans H, Niehoff K-H, Braun A, Tacke E, Hofferbert H-R. 2011. Association genetics in *Solanum tuberosum* provides new insights into potato tuber bruising and enzymatic tissue discoloration. *BMC Genomics.* 12(1):1–14. doi:10.1186/1471-2164-12-7.
- van Eck HJ, Jacobs JME, Stam P, Ton J, Stiekema WJ, Jacobsen E. 1994. Multiple alleles for tuber shape in diploid potato detected by qualitative and quantitative genetic analysis using RFLPs. *Genetics.* 137(1):303–309. doi:10.1093/genetics/137.1.303.
- van Eck HJ, Oortwijn ME, Terpstra IR, van Lieshout NH, van der Knaap E, Willemsen JH, Bachem CW. 2022. Engineering of tuber shape in potato (*Solanum tuberosum*) with marker assisted breeding or genetic modification using StOFFP20. *Res Sq.* doi:10.21203/rs.3.rs-1807189/v1.
- van Lieshout N, van der Burgt A, de Vries ME, ter Maat M, Eickholt D, Esselink D, van Kaauwen MPW, Kodde LP, Visser RGF, Lindhout P, et al. 2020. Solyntus, the new highly contiguous reference genome for potato (*Solanum tuberosum*). *G3 (Bethesda).* 10(10):3489–3495. doi:10.1534/g3.120.401550.
- Vos PG, Paulo MJ, Bourke PM, Maliepaard CA, van Eeuwijk FA, Visser RGF, van Eck HJ. 2022. GWAS in tetraploid potato: identification and validation of SNP markers associated with glycoalkaloid content. *Mol Breeding.* 42(12):76. doi:10.1007/s11032-022-01344-2.
- Vos PG, Uitdewilligen JG, Voorrips RE, Visser RG, van Eck HJ. 2015. Development and analysis of a 20 K SNP array for potato (*Solanum tuberosum*): an insight into the breeding history. *Theor Appl Genet.* 128(12):2387–2401. doi:10.1007/s00122-015-2593-y.
- Wang F, Zou M, Zhao L, Xia Z, Wang J. 2021. Genome-Wide association mapping of late blight tolerance trait in potato (*Solanum tuberosum* L.). *Frontiers Genet.* 12:714575. doi:10.3389/fgene.2021.714575.
- Werij JS, Kloosterman B, Celis-Gamboa C, De Vos CR, America T, Visser RG, Bachem CW. 2007. Unravelling enzymatic discoloration in potato through a combined approach of candidate genes, QTL, and expression analysis. *Theor Appl Genet.* 115(2):245–252. doi:10.1007/s00122-007-0560-y.
- Wolters A-MA, Uitdewilligen JGAML, Kloosterman BA, Hutten RCB, Visser RGF, van Eck HJ. 2010. Identification of alleles of carotenoid pathway genes important for zeaxanthin accumulation in potato tubers. *Plant Mol Biol.* 73(6):659–671. doi:10.1007/s11103-010-9647-y.
- Yang H, Liao Q, Ma L, Luo W, Xiong X, Luo Y, Yang X, Du C, He Y, Li X, et al. 2021. Features and genetic basis of chlorogenic acid formation in diploid potatoes. *Food Chem.* 3:100039. doi:10.1016/j.fochms.2021.100039.
- Yuan J, Bizimungu B, De Koeber D, Rosyara U, Wen Z, Lagüe M. 2020. Genome-wide association study of resistance to potato common scab. *Potato Res.* 63(2):253–266. doi:10.1007/s11540-019-09437-w.
- Zhang C, Yang Z, Tang D, Zhu Y, Wang P, Li D, Zhu G, Xiong X, Shang Y, Li C, et al. 2021. Genome design of hybrid potato. *Cell.* 184(15):3873–3883.e3812. doi:10.1016/j.cell.2021.06.006.
- Zhang F, Qu L, Gu Y, Xu Z-H, Xue H-W. 2022. Resequencing and genome-wide association studies of autotetraploid potato. *Mol Hortic.* 2(1):6. doi:10.1186/s43897-022-00027-y.
- Zhang J, Yin J, Luo J, Tang D, Zhu X, Wang J, Liu Z, Wang P, Zhong Y, Liu C, et al. 2022. Construction of homozygous diploid potato through maternal haploid induction. *aBIOTECH.* 3(3):163–168. doi:10.1007/s42994-022-00080-7.
- Zorrilla C, Navarro F, Vega-Semorile S, Palta J. 2021. QTL for pitted scab, hollow heart, and tuber calcium identified in a tetraploid population of potato derived from an Atlanticx2 superior cross. *Crop Sci.* 61(3):1630–1651. doi:10.1002/csc2.20388.