



Final Report

**DEVELOPMENT OF A SNP PANEL FOR THE
IDENTIFICATION AND TRUENESS-TO-TYPE
DETERMINATION OF POTATO CULTIVARS IN SOUTH
AFRICA**

Compiled by

Dr I. Gazendam



30 September 2021

Disclaimer

Information contained in this document results from research funded wholly or in part by the Potato Industry Development Trust upon request of Potatoes South Africa acting in good faith. Opinions, attitudes and points of view expressed herein do not necessarily reflect the official position or policies of the Potatoes South Africa. Potatoes South Africa makes no claims, promises, or guarantees about the accuracy, completeness, or adequacy of the contents of this document and expressly disclaims liability for errors and omissions regarding the content thereof. No warranty of any kind, implied, expressed, or statutory, including but not limited to the warranties of non-infringement of third-party rights, title, merchantability, fitness for a particular purpose or freedom from computer virus is given with respect to the contents of this document in hard copy, electronic format or electronic links thereto. References made to any specific product, process, and service by trade name, trademark, manufacturer or another commercial commodity or entity are for informational purposes only and do not constitute or imply approval, endorsement or favouring by Potatoes South Africa.

Vrywaring

Die inligting vervat in hierdie dokument is die resultaat van navorsing wat ten volle of gedeeltelik befonds is deur die Potato Industry Development Trust op versoek van Aartappels Suid-Afrika, handelend in goeder trou. Menings, sienings en standpunte hierin vervat reflekteer nie noodwendig die amptelike standpunte of beleid van Aartappels Suid-Afrika nie. Aartappels Suid-Afrika maak geen aansprake, beloftes of waarborge oor die korrektheid, volledigheid of toereikendheid van die inhoud van hierdie dokument nie en ontken uitdruklik aanspreeklikheid vir foute en weglatings oor die inhoud daarvan. Geen waarborg van enige aard, hetsy geïmpliseerd, uitdruklik, of statutêr, insluitende maar nie beperk tot die waarborge van nie-skending van derdeparty-regte, titels, verkoopbaarheid, geskiktheid vir 'n bepaalde doel of vry van rekenaarvirus word gegee te opsigte van die inhoud van hierdie dokument in gedrukte formaat, elektroniese formaat of elektroniese koppeling daaraan. Verwysings na enige spesifieke produk, proses en diens by wyse van handelsnaam, handelsmerk, vervaardiger of ander kommersiële kommoditeit of entiteit is slegs vir inligtingsdoeleindes en vorm of impliseer nie goedkeuring, onderskrywing of begunstiging daarvan deur Aartappels Suid-Afrika nie.

TABLE OF CONTENTS

Disclaimer	2
Vrywaring	2
TABLE OF CONTENTS	3
List of Tables	4
List of Figures	4
ACKNOWLEDGEMENTS	6
EXECUTIVE SUMMARY	7
Chapter 1. INTRODUCTION	8
BACKGROUND	8
PROBLEM STATEMENT	12
STUDY AIM AND OBJECTIVES	12
PROJECT TEAM MEMBERS	12
Chapter 2 REPORT PER OBJECTIVE	13
Assemble a comprehensive set of commercial cultivars important to the South African potato industry for developing the SNP panel.	13
Review of literature, SNP data mining and identification for potato genotyping.	14
Send the list of SNP sequences to LGC Biosearch Technologies for SeqSNP probe development.	16
Ship potato leaf samples to LGC Biosearch Technologies after selection of successful SNPs.	18
Genotype the 190 selected potato germplasms at 500 SNP positions with SeqSNP assay.	18
Analyse the SeqSNP data to develop an optimum SNP panel to discriminate cultivars.	19
Validate SeqSNP genotyping assay results of selected samples using KASP SNP assays of the selected SNP panel.	28
Order KASP SNP assays	28
gDNA isolation	28
Running KASP SNP assays	29
Score SNP genotypes with appropriate software	29
Calculate genetic relatedness between cultivars and draw phylogenetic trees to indicate the relationships between the cultivars.	32
Compile a database of SNP genotype profiles for all the entries of the germplasm database.	34
Implement the KASP SNP assays of the selected SNP panel for future genotyping of new germplasm or when purity test requests are submitted to ARC-VIMP.	35
Assign putative cultivar identity to samples submitted for fingerprinting by comparing their SNP genotypes to the germplasm SNP genotype database.	37
Chapter 3. RECOMMENDATION AND IMPACT	40
Recommendation	40
Impacts of the project	40

OUTPUTS..... 44

Appendix A. (pages 1-5). Potato germplasm list selected for SeqSNP genotyping (190 cultivars)..... 45

Appendix B. (pages 1-2). Potato germplasm list selected for KASP SNP assay verification (78 cultivars)..... 50

List of Tables

Table 1. Number of SNP markers preliminarily selected for SeqSNP of potato samples, and the average density per chromosome..... 15

Table 2. Number of SNP markers per potato chromosome for SeqSNP of potato samples with 500 selected SNP markers 17

Table 3. Selection of the appropriate markers to use to distinguish between “Amethyst” and “Agatha 11” 36

Table 4. Example input file of KASP SNP assay dosage score results of a query sample 37

Table 5. Top 10 lowest Kosman genetic difference hits to the query sample “Unknown_4”..... 38

Table 6. SNP genotypes of the top 10 lowest Kosman genetic difference hits to the query sample “Unknown_4”38

Table 7. Calculation of the random match probability of a query sample after genotyping with a small number of SNP markers 39

List of Figures

Figure 1. Dosage of alleles in tetraploid potato 10

Figure 2. Illustration of the KASP SNP assay method (LGC Biosearch Technologies)..... 11

Figure 3. All the major potato varieties from the 2018/2019 season were selected for the potato SNP panel fingerprinting method development 14

Figure 4. Position of the final selection of 500 SNPs for SeqSNP, relative to the known potato genomic map..... 17

Figure 5. Sequencing read depth produced by SeqSNP of 500 variants (SNPs) over 190 potato samples and 2 empty wells..... 20

Figure 6. Genotype accumulation curve, used for determining the minimum number of loci necessary to discriminate between individuals in a population..... 21

Figure 7. Histogram (frequency distribution) of pairs (Y) vs Kosman’s distances (X) of 190 cultivars genotyped with SeqSNP at 500 SNP positions 21

Figure 8. (page 1 and 2). Complete cluster dendrogram of pairwise genetic distances calculated with the Kosman index of 190 cultivars genotyped with SeqSNP at 500 SNP positions..... 24

Figure 9. Histogram (frequency distribution) of pairs (Y) vs Kosman’s genetic distances (X) of potato cultivars genotyped with SeqSNP at selected panel of 25 SNPs 25

Figure 10. (page 1 and 2). Complete cluster dendrogram of pairwise genetic distances calculated with the Kosman index of 173 cultivars genotyped with SeqSNP at 25 selected SNP positions 27

Figure 11. Examples of scatter plots constructed with KASP data points in the colour of the expected SeqSNP allele dosage..... 30

Figure 12. Examples of fitPoly output graphs, used to cluster and call the genotype of KASP results..... 31

Figure 13. Complete cluster dendrogram of pairwise genetic distances 33

Figure 14. Segment of a complete cluster dendrogram of pairwise Kosman genetic distances between members of the SNP genotype database and the query sample “Unknown_4” 38

ACKNOWLEDGEMENTS

The following persons/institutions are acknowledged for their contribution to the successful execution and completion of this project:

1. Potatoes South Africa, for financing the research costs.
2. ARC-VIMP, for hosting the research, providing technical and financial support during the implementation of the research project.
3. ARC-VIMP gene bank personnel for maintaining and supplying *in vitro* potato cultivars.
4. South African potato growers, for providing plant material to be included in the method development.
5. LGC Biosearch Technologies for providing analytical services, data support, and replacing reagents that were not handled correctly during shipping.
6. Albert Gazendam, my husband, for sacrificing many hours providing guidance on how to use R coding for data analysis.
7. My supervisors, Dr Sunette Laurie and Dr Michael Bairu, and colleagues, Ms Pinkie Mojapelo and Dr Amele Assefa for editing the report.

EXECUTIVE SUMMARY

The suitability of single nucleotide polymorphism (SNP) panels for fingerprinting of tetraploid potato were investigated as a new high throughput method instead of simple sequence repeats (SSRs). Developing a SNP protocol to replace SSR and Polyacrylamide Gel Electrophoresis (PAGE) methods was necessary. This project entailed designing, optimising, and implementing a custom SNP panel useful for fingerprinting potato cultivars.

For the SeqSNP genotyping at LGC Biosearch Technologies, 500 SNP markers and 190 potato cultivars were selected. An optimal panel of 25 SNP markers was identified that could discriminate between South African potato cultivars on genetic allele dosage. The SeqSNP genotyping results of 25 selected SNPs were validated on selected potato genotypes using KASP SNP assays. The genetic relatedness between cultivars was calculated and phylogenetic trees were constructed to indicate the relationships between the cultivars. A database of SNP genotype profiles was compiled for all the entries of the germplasm database. Results showed that the KASP SNP assays designed for selected potato SNPs were effective for 21 out of 25 SNPs tested and matched SeqSNP genotypes in 97.7% of the samples used for verification. The 21 SNP panel accurately identified potato cultivars important in South Africa.

The advantages of the SNP panel method of potato DNA fingerprinting are as follows:

1. Increased throughput and phasing out of outdated technologies,
2. Ability to choose a smaller number of SNP markers to distinguish two cultivars with confidence,
3. Objective method for assigning putative cultivar identities to samples submitted for fingerprinting by comparing their SNP genotypes to the germplasm SNP genotype database.

The KASP SNP assays of the successful SNP panel are therefore available for potato DNA fingerprinting as new germplasm, or purity test requests are submitted to ARC-VIMP. The intellectual property right regarding the identity of the SNPs in the KASP SNP panel remains the property of the ARC.

Chapter 1

INTRODUCTION

Background

DNA fingerprinting is a molecular technique applied to identify genetic differences between cultivars or lines and used for genetic purity testing.

The benefit of potato DNA fingerprinting is that it can be done at very early stages of development, such as mini-tuber or *in vitro* leaf material, and it is less resource-intensive than morphological methods. Accidental mixes can therefore be identified early, before *in vitro* multiplication, to prevent costly mistakes later on. Cultivar genetic identity is important in the protection of plant breeders' rights.

The South African potato industry is the main client for which the Agricultural Research Council (ARC) provides genetic fingerprinting services. It is the largest vegetable commodity in South Africa, with a gross production value of R8.08 billion in 2019 (Food and Agriculture Organization of the United Nations, 2019). The South African Seed Potato Certification scheme certified about 6.8 million (25 kg) bags of seed potatoes during the 2017/2018 production season (Potato Certification Service, 2019). Seed potato growers must maintain their cultivars' genetic purity to provide true-to-type cultivars to the potato production industry. ARC-VIMP provides this service and is, therefore, essential to the industry.

The *in vitro* gene bank of the ARC-VIMP is also dependent on the fingerprinting service and cannot release material to the industry unless trueness-to-type is confirmed. In this sense, DNA fingerprinting of potato cultivars benefits everyone involved in the industry. The fingerprinting of newly introduced lines during the potato certification process is definitely essential but is not yet mandatory (Jamali et al., 2019; Singh et al., 2019).

The method currently being used at the ARC-VIMP (Vegetable, Industrial, and Medicinal Plants) is based on Polymerase Chain Reaction (PCR) of Simple Sequence Repeat (SSR) sequences. It produces a mixture of DNA fragments from each DNA sample. As a result of electrophoresis in a gel matrix, fragments are separated by size, resulting in a characteristic pattern (fingerprint) of bands from each sample. Differences are observed as the presence or absence of a particular fragment. If the fingerprint patterns of two plants differ, the plants are either not identical cultivars or their clones are not true-to-type. SSR fingerprinting is a popular method used for potato cultivar identification (McGregor et al., 2000a and b; Coombs et al., 2004; Moisan-Thiery et al., 2005) and evaluation of genetic diversity (Liao and Guo, 2014).

SSRs have been used to fingerprint potato cultivars in South Africa for many years, and five SSR markers have been used to effectively distinguish all South African cultivars. SSR markers have high mutation rates (due to DNA polymerase slippage during DNA replication) and provide high allele numbers per marker. The conventional SSR system has been extremely useful in identifying potato cultivars in South Africa and will continue to play an integral role in fingerprinting a variety of crops in the future (McGregor et al., 2000a). SSR analysis, however, has some limitations related to throughput, cost, and scoring of multiple alleles. A more robust, flexible system is needed with downstream options to apply test results.

The old PAGE method for fragment analysis has become outdated due to its limited sample-handling capacity. To better serve potato farmers and industry, a project was initiated to migrate to a modern method of liquid capillary electrophoresis and electronic scoring of fingerprinted profiles. This method is non-subjective, more sensitive, reproducible, and will increase efficiency due to a much higher throughput per day. An application for funding was developed and presented to Potatoes South Africa's research committee on 5 March 2019. Potatoes South Africa gave the project high priority since it addresses current issues, and on 18 March 2019, it approved ARC funding of R153 990 for two years.

Since the proposed project of modernising SSR fingerprinting using fragment analysis (February 2019) was conceived, Dr Gazendam, the principal investigator, has attended two courses on the suitability of SNP panels over SSRs in fingerprinting tetraploid potatoes. The two courses attended were the Microsatellite workshop held by the University of Pretoria from 5 April to 10 May 2019 and the second was held by the African Plant Breeding Academy (AfPBA) from 13 to 25 May 2019, in Nairobi, Kenya.

Dr Gazendam, therefore, investigated the possibility of SNP genotyping as the most modern method of potato fingerprinting. A document was prepared to summarise the rationale, work plan and activities that fit into the approved budget of the project selected for funding by PSA. Dr Fienie Niederwieser was consulted on 12 June 2019 with a revised proposal that incorporated a new approach to using SNP genotyping for fingerprinting potatoes, with budget and outputs remaining the same. Instead of developing fragment analysis for only 30 gene bank cultivars, as proposed in the previous approach, a SNP panel will be developed on 190 cultivars and allow us to discriminate all the cultivars maintained in the ARC-VIMP *in vitro* gene bank. Dr Niederwieser accepted the change in the methodology. A new name was given to the project, "Development of a SNP panel for the clonal identification and trueness-to-type determination of potato cultivars in South Africa". PSA and ARC signed the contract on 13 September 2019 and 22 November 2019, respectively.

SNP genotyping is the most modern method of determining genetic polymorphism, and due to advances in technology, it has become easier and cheaper to assay SNPs than ever before. Several authors suggest panels of SNP markers to replace SSR marker analysis for cultivar identification due to technical and economic reasons (Uitdewilligen et al., 2012; Vos et al., 2015; Chen, 2016; Sasaki et al., 2017).

SSR analysis is expensive per data point due to the low throughput and cost of reagents. A number of challenges were encountered in scoring SSR fragments in potato genotypes due to multiple alleles (tetraploid with four copies of the same chromosome), non-unique mapping of some of the SSR markers in the genome, and stutter bands. Stutter bands occur naturally when DNA polymerase amplifies SSR fragments and are difficult to prevent. The most informative SSR marker currently used (STK) is a complex mono- and di-nucleotide SSR, and STG is a tri-nucleotide SSR, so the lower separation of neighbouring alleles and the relatively high level of stutter bands are inevitable. According to Reid et al. (2011), one allele of STM3023 (di-nucleotide SSRs) is located at the stutter position for the other allele, resulting in a complication of the allele call.

The choice of DNA polymerase enzyme, genomic DNA quality and quantity, PCR multiplexing or mixing ratio of pseudo-multiplexed DNA fragments labelled with different fluorescent dyes require intensive protocol standardisation, which may still result in non-optimal fragment detection and scoring. The previously proposed strategy necessitated outsourcing SSR fragment analysis by liquid capillary electrophoresis to service providers outside of the ARC. Consideration would have to be given to logistics and turnaround times. The output files (called pherograms) produced by liquid capillary electrophoresis must then be analysed with software to generate an electronic database of fragment profiles. Due to the indirect method of determining fragment sizes, the SSR allele database would not contain fragments of identical sizes, making cultivar comparisons difficult. Hence, an improved method of electronically collecting and objectively scoring genotyping results is required to optimally serve the potato industry. Converting to the most modern genotyping method, such as SNP genotyping, was proposed since it will remain relevant for a long time.

SNP is the variation in a single nucleotide that occurs at a specific position in the genome of any organism. To be considered a SNP, the variation must be present in more than 1% of the population. If the variation is present at a lower frequency, it is considered a rare mutation (abnormal change). SNPs are highly abundant in plants and spread out evenly over the genome. Potato genomes are highly polymorphic, with one SNP occurring every 20 base pairs (bp). SNPs have been identified in significant quantities for potatoes using various high-throughput sequencing technologies (Hamilton et al., 2011; Uitdewilligen et al., 2012 and 2013). These SNPs have a fixed location on the reference genome of potatoes (The Potato Genome Sequencing Consortium, 2011) and are publicly available. They can be used in several types of flexible genotyping platforms, such as the KASP genotyping platform

(<https://www.biosearchtech.com>) (Lindhout et al., 2011) or the development of SNP arrays (Felcher et al., 2012). The Infinium SolCAP 12 K array has been successfully utilised to fingerprint and characterise diverse potato collections of the Northwest Potato Variety Development program in the United States (Bali et al., 2017), EEA INTA Balcarce in Argentina (Deperi et al., 2018) and the potato collection at the International Potato Centre (CIP) in Lima, Peru (Ellis et al., 2018). The following are some advantages of SNP markers over SSR or PAGE/CE for fingerprinting potatoes:

SNP assays are high throughput and cheaper than SSRs.

They have highly robust amplification from relatively crude DNA extracts.

SNP markers are able to estimate allele dosage (Figure 1), thus, nucleotide genotype and copy number can be determined from a polyploid genome (Chen, 2016; Sasaki et al., 2017). Copy number determination is impossible or produces highly variable results when SSR fragments are analysed.

SNP assays are reproducible. In-house endpoint KASP SNP assay fluorescent detection in 96-well plates is possible in the ARC-VIMP real-time PCR machine after running assays in normal PCR machines, which increases the throughput compared to PAGE.

Outsourced genotyping services, with a turnaround time of 1-2 months, are available when a large set of germplasm, such as large breeding lines and cultivars collections, needs to be genotyped with SNPs.

Software is used to cluster and assign SNP genotypes quickly and objectively.

SNP calling is robust and accurate since specific primers and probe combinations bind to their complementary sites in the potato genome.

SNP genotypes simplify the germplasm genotype database, allowing objective comparison to determine the suggested identity of unknown cultivars.

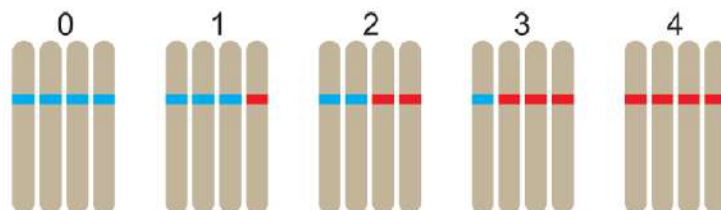


Figure 1. Dosage of alleles in tetraploid potato

Notes: Five distinct dosages are possible in a tetraploid at a bi-allelic marker position, ranging from 0 copies of the alternative allele to 4 copies. Here, the alternative allele is coloured red, with the reference allele coloured blue (Bourke et al., 2018).

Single-tube assays such as the Kompetitive Allele Specific PCR (KASP; LGC Biosearch Technologies, Hoddesdon, United Kingdom) eliminate all post-PCR sample handling, thereby reducing the cost and time of SNP genotyping while lowering error rates (Chen et al., 2010). Researchers found KASP to be a cost-effective and scalable SNP genotyping solution for small to moderate numbers of markers such as DNA fingerprinting for quality control analysis (Neelam et al., 2013; Semagn et al., 2014; Steele et al., 2018). **Figure 2** explains the process utilised during KASP SNP assays (LGC Biosearch Technologies). Performing the assay is, therefore, relatively easy. It entails adding together the correct amount of plant genomic DNA, the KASP SNP assay mix containing the three primers and the KASP-TF master mix, incubating this at a specific temperature profile and reading the end-point fluorescence of FAM and HEX dyes that correspond to alleles 1 and 2.

Identifying a small custom SNP panel that will be used to distinguish important potato cultivars in South Africa requires screening a larger number of SNPs on as many cultivars as the project budget can afford. The SeqSNP™

technique (LGC Biosearch Technologies) was chosen since it is more cost-effective for genotyping medium to large numbers of samples, making it cheaper for a greater number of SNPs than KASP SNP assays. SeqSNP is a targeted genotyping by sequencing (GBS) approach, which uses flexible, in-solution probe libraries to enrich targets before high-throughput sequencing. Additionally, SeqSNP provides flanking sequence information for each SNP.

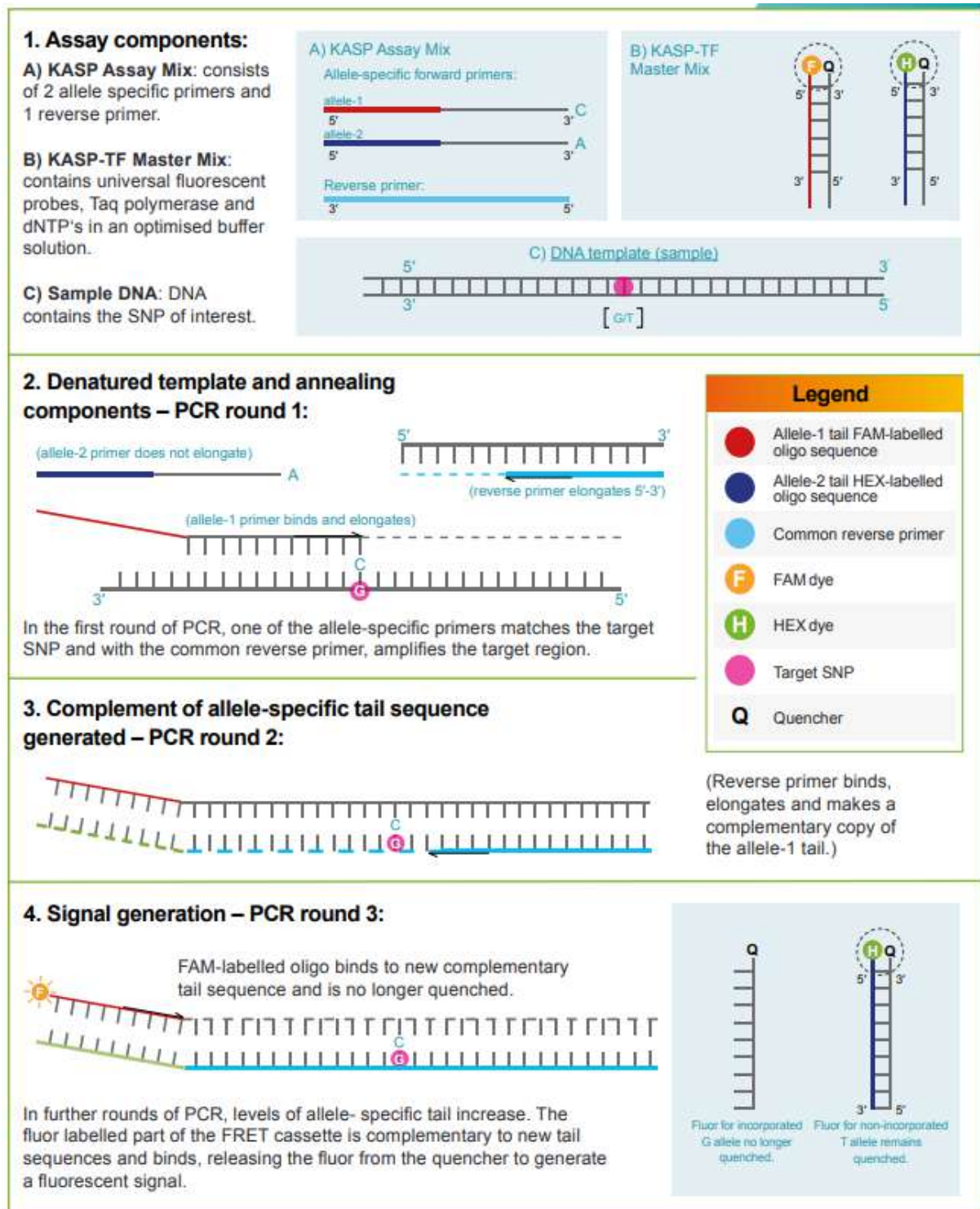


Figure 2. Illustration of the KASP SNP assay method (LGC Biosearch Technologies)

Problem statement

For potato farmers and the potato industry, the phasing out of the old SSR and PAGE methods for DNA fingerprinting necessitated protocol development to migrate to the most modern method of SNP genotyping.

Study aim and objectives

The study aimed to develop and optimise a protocol for migrating potato DNA fingerprinting from the old SSR and PAGE methods of fragment analysis to the most recent SNP genotyping methods. A custom SNP panel needs to be developed to discriminate between all the potato cultivars important in South Africa.

The specific objectives of the study were to:

1. Assemble a comprehensive set of commercial cultivars important to the South African potato industry for developing the SNP panel.
2. Review of literature, SNP data mining and identification for potato genotyping.
3. Send the list of SNPs sequences to LGC Biosearch Technologies for SeqSNP probe development.
4. Ship potato leaf samples to LGC Biosearch Technologies after selection of successful SNPs.
5. Genotype the 190 selected potato germplasms at 500 SNP positions with SeqSNP assay.
6. Analyse the SeqSNP data to develop an optimum SNP panel to discriminate cultivars.
7. Validate SeqSNP genotyping assay results of selected samples using KASP SNP assays of the selected SNP panel.
8. Calculate genetic relatedness between cultivars and draw phylogenetic trees to indicate the relationships between the cultivars.
9. Compile a database of SNP genotype profiles for all the entries of the germplasm database.
10. Implement the KASP SNP assays of the selected SNP panel for future genotyping of new germplasm or when purity test requests are submitted to ARC-VIMP.
11. Assign putative cultivar identity to samples submitted for fingerprinting by comparing their SNP genotypes to the germplasm SNP genotype database.

Project team members

Dr I Gazendam – Principal investigator
Ms P Mojapelo – Senior research technician
Mr F Steyn – Acting gene bank manager
Dr M Bairu – Plant breeding research team manager

Chapter 2

REPORT PER OBJECTIVE

Assemble a comprehensive set of commercial cultivars important to the South African potato industry for developing the SNP panel

Materials and methods

A germplasm set was obtained from the ARC-VIMP *in vitro* gene bank cultivar or contract collections. The germplasm list was annotated as follows: 1 = important commercial cultivars according to PSA annual reports, 2 = on the SA potato variety list (DAFF), 3 = if the cultivar has already been fingerprinted at the ARC-VIMP using SSRs, and 4 = cultivars overlapping with those used by Vos et al. (2015, 2016) were added. Due to the inaccessibility of all national cultivar collections and commercially important entries, suggestions and availability were considered, and some cultivars were revived from *in vitro* long storage. Additionally, commercial clients were invited to contribute and make requests. Germplasm received as tubers were planted in a greenhouse on 17 Aug 2020 after being treated with Rindite (ethylene chlorhydrin – ethylene dichloride – carbon tetrachloride 7 : 3 : 1) to break the dormancy.

Results

A set of 190 potato cultivars were selected. The list of the selected genotypes is presented in **Appendix A**, and the list includes the reason for selection of each genotype. Twenty potato cultivars were received from different companies:

- Eleven cultivars (7Four7, Belmonda, Connect, IIZA49A1, IIZASSA5, King Russet, Lanorma, Noya, Prada, Royal, Taisiya) were obtained from GWK Trading;
- Three (FL2006, FL2108, FL2476) from Pepsico (Mr Frank Osler);
- Five (Adato, Avalanche, Fianna, Markies and Sound) from First Potato Dynamics (FPD) (Mr Theuns van Rensburg) and
- Two (Panamera, Taurus) from Rascal Seed Research Laboratories (Mr Dawie Ras).

Additionally, McCain Foods (Ms I Vorster) requested inclusion of 96-0568-002 (Arno), Amigo, Crop60, Clearwater Russet, Dakota Trailblazer, Magnum, Monica Russet, Royal, and Teton Russet. Also included on the list are all 10 potato varieties that were deemed important during the 2018/2019 growing season (**Figure 3**, Potato Certification Service, 2019): Mondial, Sifra, Lanorma, FL2108, Panamera, Valor, Markies, Innovator, Up-to-Date and Taurus.

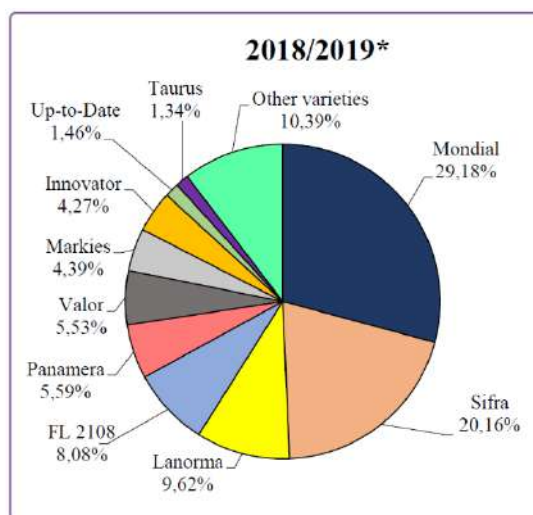


Figure 3. All the major potato varieties from the 2018/2019 season were selected for the potato SNP panel fingerprinting method development

(Source: Potato Certification Service, 2019).

Discussion

A representative sample from a germplasm population is essential for developing or validating DNA assays. This step involved selecting a set of South African potato cultivars that represent all varieties relevant to developing a method in South Africa. The 190 potato cultivars selected in this study could not contain all the important potato germplasms and is the minimum sample number for SeqSNP (LGC Biosearch Technologies). Higher numbers were not affordable with the funding available. The identities of the potato reference collection (~40 cultivars) maintained at the ARC-VIMP were withheld by DALRRD. The selected germplasm list may contain some of these important varieties, but some may have been excluded.

Review of literature, SNP data mining and identification for potato genotyping.

Materials and methods

This step involved selecting several SNPs with high polymorphism information content (PIC) (with a minor allele frequency (MAF) close to 50%) from the literature to develop a small panel of SNPs that could be used to differentiate South African potato cultivars.

A set of 20 000 SNPs, studied by Vos et al. (2015), were filtered to select over 1000 SNPs for SeqSNP assay design. Vos et al. (2015) used 20 000 SNPs to screen 569 cultivars and breeding lines. Of these 20 000, only the successful 14 530 SNPs were considered (the same approach used by Chen, 2016). The SNPs were filtered to be informative (high PIC value) and spread out evenly over all the 12 potato chromosomes. All non-coding SNPs were removed to minimise assay failure rates since coding regions had a lower assay failure rate than non-coding regions (Vos et al., 2015, 2016). No chloroplast, unmapped SNPs or SNPs designed for resistance genes were selected. The top 1500 SNPs with minor allele frequency values higher than 33.6% were further filtered to remove SNPs denser than 0.1 Mb. Low-density chromosomal areas were identified by looking at plotted coordinates of SNPs. More SNPs were chosen in these areas and for chromosome 12, which had a low marker density.

Results

A literature review was conducted to identify publications describing potato SNP marker sequences and presenting data on SNP genotypes and allele frequencies of large sets of potato germplasm. The study by Vos et al. (2015) provided supplementary tables of SNP genotype data at ~20 000 loci for 569 genotypes (representing commercial potato cultivars and advanced breeding lines from the Netherlands). These SNPs were mostly derived from Hamilton et al. (2011) and Uitdewilligen (2012 and 2013).

The preliminary selection from the 14 530 successful SNPs, after filtering for the highest PIC values ($33.6\% < \text{MAF} < 50\%$), resulted in 1210 SNPs (Table 1). This represented an average marker interval from 0.42 Mb (Chr02) to 0.8 Mb (Chr12) per chromosome.

Table 1. Number of SNP markers preliminarily selected for SeqSNP of potato samples, and the average density per chromosome

Chromosome	Number of markers	Largest coordinate	Ave marker interval (Mb)
1	134	88 572 440	0.66
2	114	48 436 012	0.42
3	105	62 269 671	0.59
4	107	72 145 421	0.67
5	103	51 998 374	0.50
6	79	58 462 796	0.74
7	115	56 629 628	0.49
8	102	56 753 185	0.56
9	96	61 466 245	0.64
10	79	59 403 404	0.75
11	89	45 092 249	0.51
12	76	61 086 761	0.80
Unknown	11		
	1210		

Discussion

As per Vos et al. (2015), the average minor allele frequency (MAF) of SNP found in recently released potato varieties is about 10 times smaller than the average MAF of SNP found in varieties released before 1945. Thus, it is reasonable to conclude that new SNPs found in the recent potato varieties reveal a low allele frequency. The MAF value may also be used to determine the age of the allele. Allele frequencies of pre-1945 SNPs are relatively stable, and over a century of selective breeding did not affect them. By filtering SNPs according to their high PIC values, we can select SNPs in the old founding cultivars and will be able to differentiate between other sets of potato germplasm, such as South African potato cultivars and other cultivars selected in this project.

Major and minor allele frequencies are influenced by the specific population. For variety identification, it is more appropriate to refer to population allele frequencies relative to the potato reference genome (REF and ALT alleles) since it is more stable.

PIC is dependent on the allele frequency. A balanced allele frequency contributes to the highest discriminatory power, and the population allele frequency of SNPs affects the probability of each allele dosage. According to the definition of Anderson et al. (1993), $\text{PIC}_i = 1 - \sum p_{ij}^2$; where p_{ij} is the frequency of the allele j for each marker i . For bi-allelic SNPs data, specifically, the formula can be rewritten as $\text{PIC} = 1 - \text{REF}^2 - (1 - \text{REF})^2$, where REF is the frequency

of the reference allele. Therefore, a PIC value of 0.5 corresponded to the theoretical maximum for bi-allelic markers. When REF closes to 50%, the PIC reaches the highest value (0.5).

Send the list of SNP sequences to LGC Biosearch Technologies for SeqSNP probe development

Materials and methods

The third step in developing a panel of SNPs useful for fingerprinting potatoes was to genotype the germplasm set with the selected SNPs using SeqSNP at LGC Biosearch Technologies. A list of 1210 sequences containing SNPs was compiled into a file format with specific headings (BED file) and sent to LGC Biosearch Technologies in Apr 2020 to design SeqSNP probe assays. On 4 Aug 2020, the list of SNPs was reduced to 500 and submitted for reagent synthesis.

Results

SeqSNP is a targeted GBS approach with the added advantage of providing flanking sequence information of each SNP. LGC Biosearch Technologies analysed the BED file list of 1210 SNPs' coordinates to determine if they yield the expected SNP alleles according to the potato reference genome (ST4.03, http://solanaceae.plantbiology.msu.edu/pgsc_download.shtml) (Sharma et al., 2013). Of the 1210, 11 markers with unknown chromosome positions were excluded and 1199 SNP assays were designed successfully. Of the successful SNPs, 94% (1130) were covered with two oligo probes and had no off-target hits to the potato genome. The final 1000 were selected from this list.

Due to the increase in the R/\$ exchange rate brought about by COVID-19, the number of SNPs that SeqSNP could genotype was reduced to 500. The number of germplasm samples could not be reduced further since 190 was the smallest set that could be sent for SeqSNP.

After LGC Biosearch Technologies designed the SeqSNP probe assays, the list of 1210 SNPs was filtered according to high specificity (no off-target hits allowed), primer annealing temperature outside the range of 45 - 60°C, and primer T_m differences of the probes (>10°C). Filtering for large MAF while considering spacing (according to the chromosome coordinate of the SNP) was done to ensure that the final selection of SNPs is not closely linked and spans as much of the genome as possible.

After further filtering, the final selection of SNPs was 500, corresponding to 31 to 61 SNPs per chromosome, according to their size (**Table 2**), representing an average marker interval of 1.40 - 1.46 Mb, respectively. This list was sent to LGC Biosearch Technologies on 4 Aug 2020.

The 500 SNP's position relative to the known potato map is plotted in **Figure 4** and indicates an even spacing of the selected SNPs over the 12 chromosomes. Low SNP density is characteristic around the centromeres of chromosomes since most non-coding SNPs were already removed.

Table 2. Number of SNP markers per potato chromosome for SeqSNP of potato samples with 500 selected SNP markers

	Nu of markers	Largest coordinate (Mb)	Ave marker interval (Mb)	Minimum minor allele frequency (MAF)	Maximum minor allele frequency (MAF)
Chr 1	61	88.6	1.45	34.9	50.0
Chr 2	33	48.3	1.46	32.7	49.9
Chr 3	43	61.9	1.44	26.3	49,9
Chr 4	50	71.8	1.44	31.0	49.8
Chr 5	36	51.9	1.44	28.6	49.9
Chr 6	41	58.5	1.43	37.1	49.2
Chr 7	40	55.9	1.40	31.2	49.8
Chr 8	39	54.7	1.40	27.3	50.0
Chr 9	43	61.5	1.43	28.4	49.3
Chr 10	41	59.4	1.45	29.6	49.9
Chr 11	31	45.1	1.45	31.5	49.7
Chr 12	42	59.3	1.41	32.0	49.8
Total:	500	716 Mb			

Notes: The average marker interval and minimum and maximum minor allele frequency (MAF) of selected SNPs (according to Vos et al., 2015) are also presented.

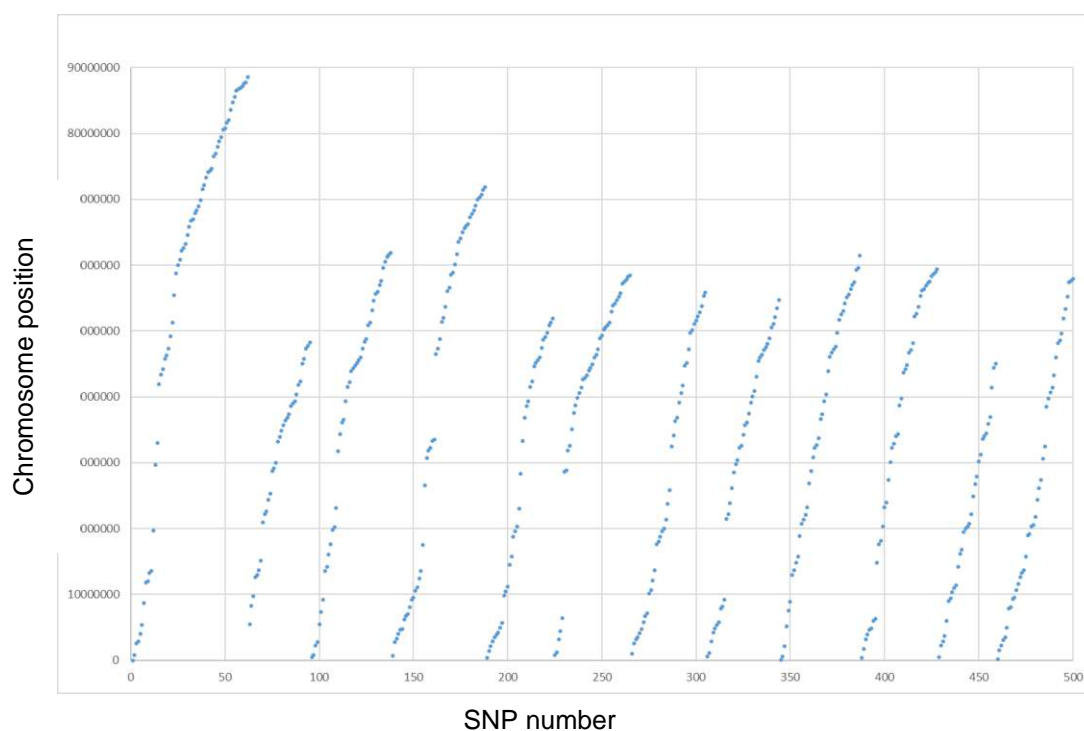


Figure 4. Position of the final selection of 500 SNPs for SeqSNP, relative to the known potato genomic map

Ship potato leaf samples to LGC Biosearch Technologies after selection of successful SNPs

Materials and methods

Many of the selected germplasm entries were sampled on 29 October 2020 as leaves from a field at Zeekoegat, Roodeplaat, Pretoria, planted for cultivar characterisation. Leaf disks were punched and placed into the LGC Biosearch Technologies 96-well sample collection plate; the desiccant was applied and sealed in a plastic bag. For the remaining accessions, leaves were collected from *in vitro* plants in batches between 2 to 25 November 2020. The samples were processed into a second plate and frozen before being freeze-dried and packed with the desiccant. The sampling plates with dried leaf samples, accompanied by a description for customs, order documentation and plate map file, were sent by courier to LGC Biosearch Technologies GmbH, Berlin, Germany, on 26 November 2020.

Results

SeqSNP was initially planned to be performed on genomic DNA sent to LGC Biosearch Technologies. Although genomic DNA is available in the freezer for many potato cultivars in the DNA database, the quality of the gDNA is important for Next Generation Sequencing (NGS) analysis. The sampling kits from LGC Biosearch Technologies, in 96-well format, were therefore ordered to perform the gDNA isolation for only \$1 per sample, which will increase the quality of gDNA and improve the success of NGS during SeqSNP.

Genotype the 190 selected potato germplasms at 500 SNP positions with SeqSNP assay

Materials and methods

Genotyping of the selected set of genotypes at LGC Biosearch Technologies was done using SeqSNP. 75 bp single reads were generated on an Illumina NextSeq 500/550 v2 sequencer. Demultiplexing of libraries was performed using barcodes, reads were clipped for adapter barcodes and quality trimmed, and aligned against the potato reference genome (ST5149G_2) with Bowtie2 v2.2.3. Variant discovery and genotyping of samples were performed with FreeBayes v1.2.0, a Bayesian variant caller that provides a most likely genotype (Garrison and Marth, 2012).

Genomic DNA extraction and SeqSNP genotyping were completed by LGC Biosearch Technologies on 21 December 2020, and the memory stick with data arrived at the ARC-VIMP on 4 January 2021. The data included a spreadsheet and variant call format (VCF) file of the genotypes and allele counts, sequence reads (FASTQ) and alignment files (BAM) produced for each sample at each variant by the NGS process.

Results

SeqSNP is a targeted GBS approach with the added benefit of providing flanking sequence information of each SNP and is cheaper for larger numbers of SNPs than KASP SNP assays. As mentioned earlier, genotyping was done at a reduced number of 500 SNPs due to a R\$ exchange rate hike since the budget calculation. The turnaround time for SeqSNP was one week after receiving the samples for gDNA extraction. Data generation was completed on 21 December 2021 after the package arrived in Germany on 14 December 2021.

Analyse the SeqSNP data to develop an optimum SNP panel to discriminate cultivars

Materials and methods

SeqSNP data (VCF file) was successfully imported into the R software package and filtered for read depth and quality (Grunwald et al., 2016). Allele frequencies were extracted with `gt.to.popsum` of the `vcfR` package in R (Knaus and Grunwald, 2017). The PIC value for each SNP was calculated using the formula $PIC = 1 - REF^2 - (1-REF)^2$, where REF is the frequency of the reference allele. The genotype accumulation curves up to a maximum number of 100 loci were drawn in R with `poppr`.

The Kosman genetic distance method was implemented to calculate the pairwise difference in genotypes between individuals. This method considers the allele dosage scores and averages the value over all loci (Kosman and Leonard, 2005); `gd.Kosman` in the `PopGenReport` package in R (Gruber and Adamack, 2014)).

The flanking variant environment was determined from variant calling performed by LGC Biosearch Technologies on the alignment of the raw SeqSNP sequencing reads. R and Microsoft Excel were used to determine the number of flanking variants and distance to each target SNP.

Various parameters were considered and an iterative software pipeline was developed to enable the selection of an optimal panel for discriminating among South African potato cultivars on genetic allele dosage. SNPs were marked in the flanking sequences, and a maximum number of one SNP both upstream and downstream of the target SNP, but no SNP closer than 20 bases, was allowed.

Results

Analysing SeqSNP data was necessary to select a panel of 25 SNPs that could discriminate between South African potato cultivars in specific combinations as required.

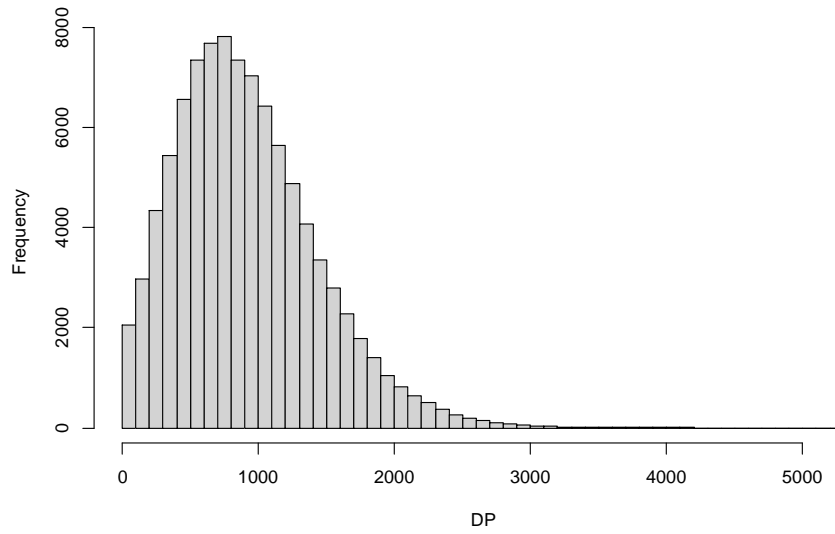
Raw SeqSNP read number per sample ranged from 11 K to 1.2 million reads. 79-82% of reads per sample were retained after adapter clipping and quality trimming. The average effective target SNP coverage was 923-fold, much higher than what LGC Biosearch Technologies usually aim for (400-fold in tetraploids) (**Figure 5A and B**). One cultivar ("Connect") yielded the most missing data, caused by lower than average read depth. Read depth was filtered to >100 reads before proceeding with the SeqSNP data analysis.

Of the 500 SNPs genotyped, 23 were not bi-allelic, i.e., two or more alternative alleles were observed. Among the 23, 6 had one or two occurrences of the third allele. There was only one monomorphic SNP among the 190 cultivars. In the three cultivars sent in duplicate (BP1, Mondial, and Up to Date), there were 28-30 SNP genotype differences per cultivar. These duplicate samples were from different sources and ages kept in the ARC-VIMP *in vitro* genebank (see Appendix A).

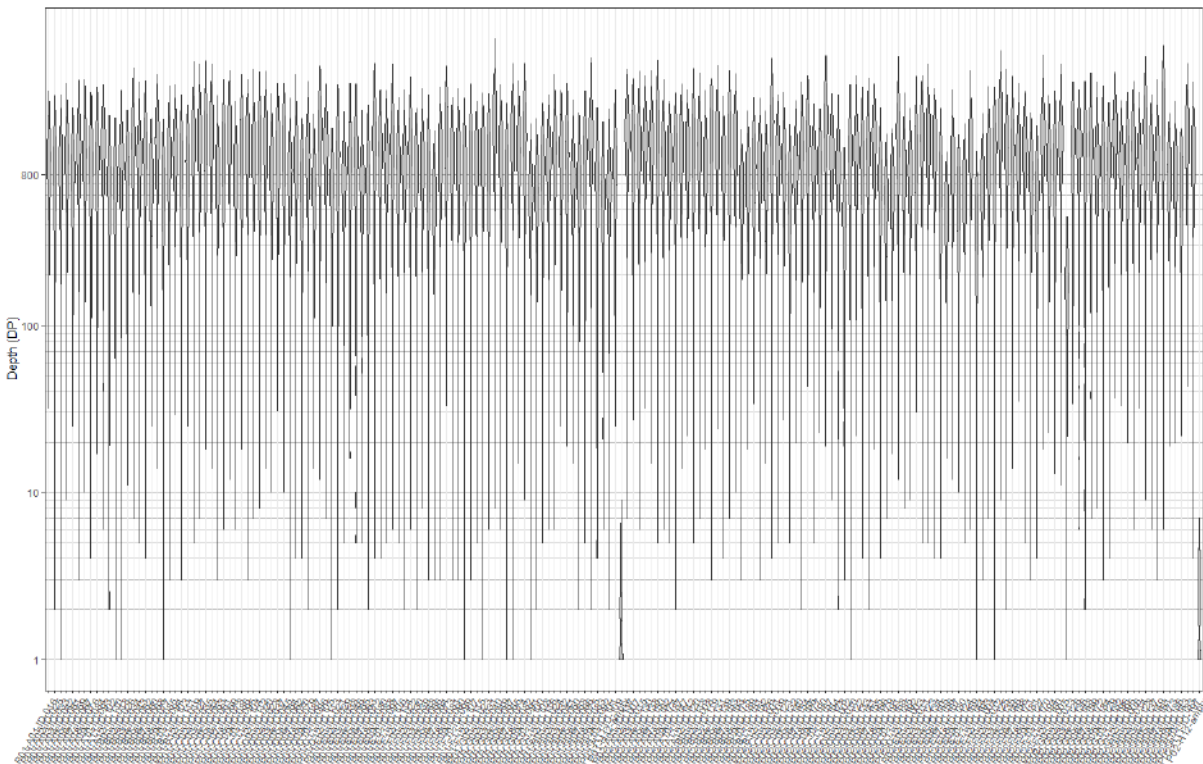
In the genotyped population of 190 cultivars, 500 SNPs were analysed for diversity. REF for the 500 SNPs ranged from 0.25 to 1. All the SNPs (except the lowest four) had PIC values between 0.22 and 0.5. Sixty-six percent (328 out of the 500) had REF values between 0.4 and 0.6, corresponding to $0.48 < PIC < 0.5$.

Genotype accumulation curves are useful for determining the minimum number of loci necessary to discriminate between individuals in a population. The function in R randomly samples loci without replacement and counts the number of observed multi-locus genotypes. According to the genotype accumulation curve (**Figure 6**), 185.73 out of 190 samples can be distinguished if only 25 loci are randomly chosen. With a number so close to 190, it is likely that a 25 SNP panel will be sufficient to discriminate genotypes and that a panel size of 25 is appropriate.

Distribution of Read depth (DP) of vcf file



A



B

Figure 5. Sequencing read depth produced by SeqSNP of 500 variants (SNPs) over 190 potato samples and 2 empty wells

Notes: A. Histogram and B. Violin plot of sequencing read depth.

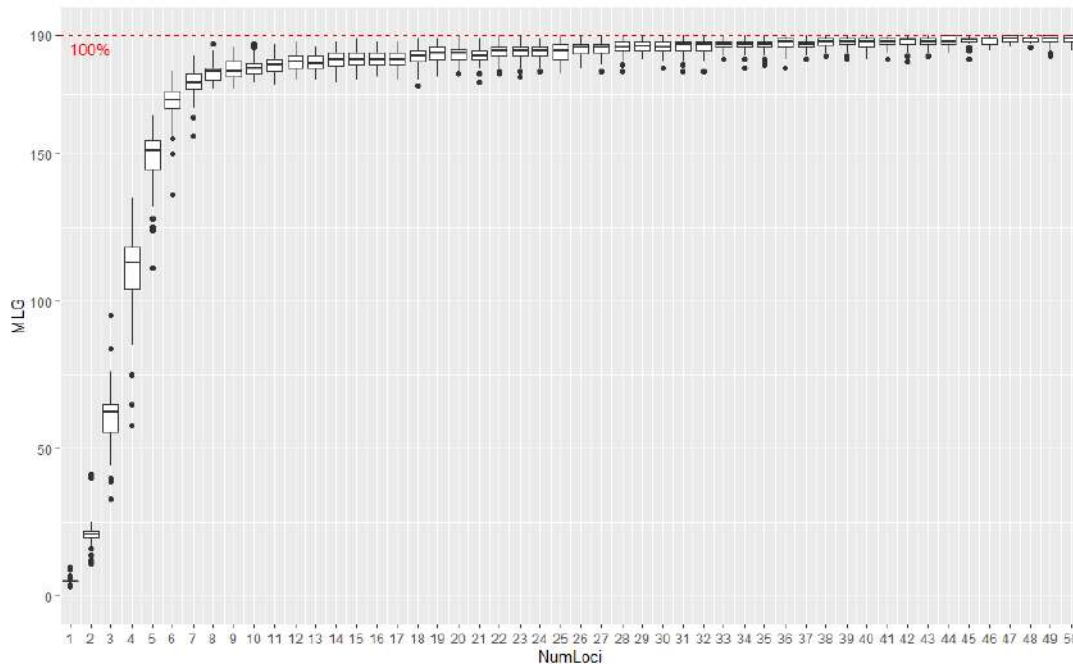


Figure 6. Genotype accumulation curve, used for determining the minimum number of loci necessary to discriminate between individuals in a population

The pairwise genetic distances between cultivars were calculated (using Kosman's index) to indicate the genetic relationships. The largest similarities are expected for similar or identical cultivars, as expected between the three duplicated pairs. Based on the Kosman genetic distances, 190 cultivars had 20 additional pairs of similar cultivars ($0.008 < D < 0.020$), possibly because of identical parental genotypes, being mixed, or having been mislabelled during *in vitro* maintenance or field planting (Figure 7). The genetic distances between the other unique pairs ranged from 0.154 to 0.374, with a mean of 0.265 ± 0.025 (Figure 7).

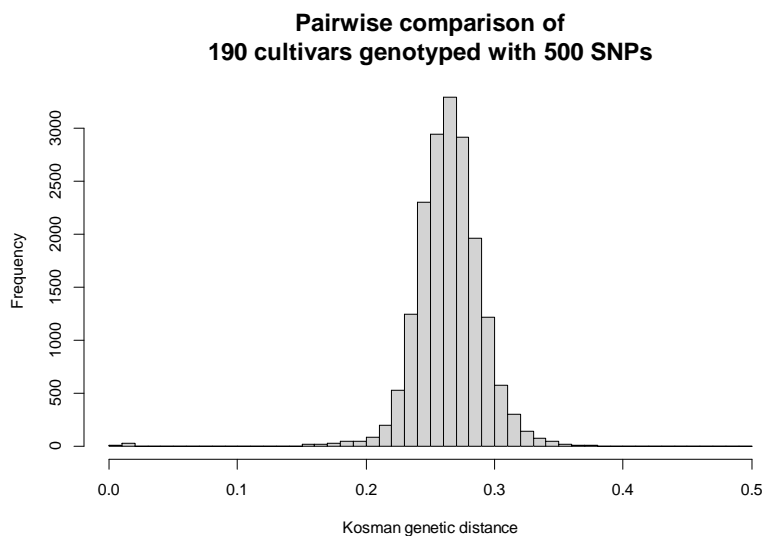


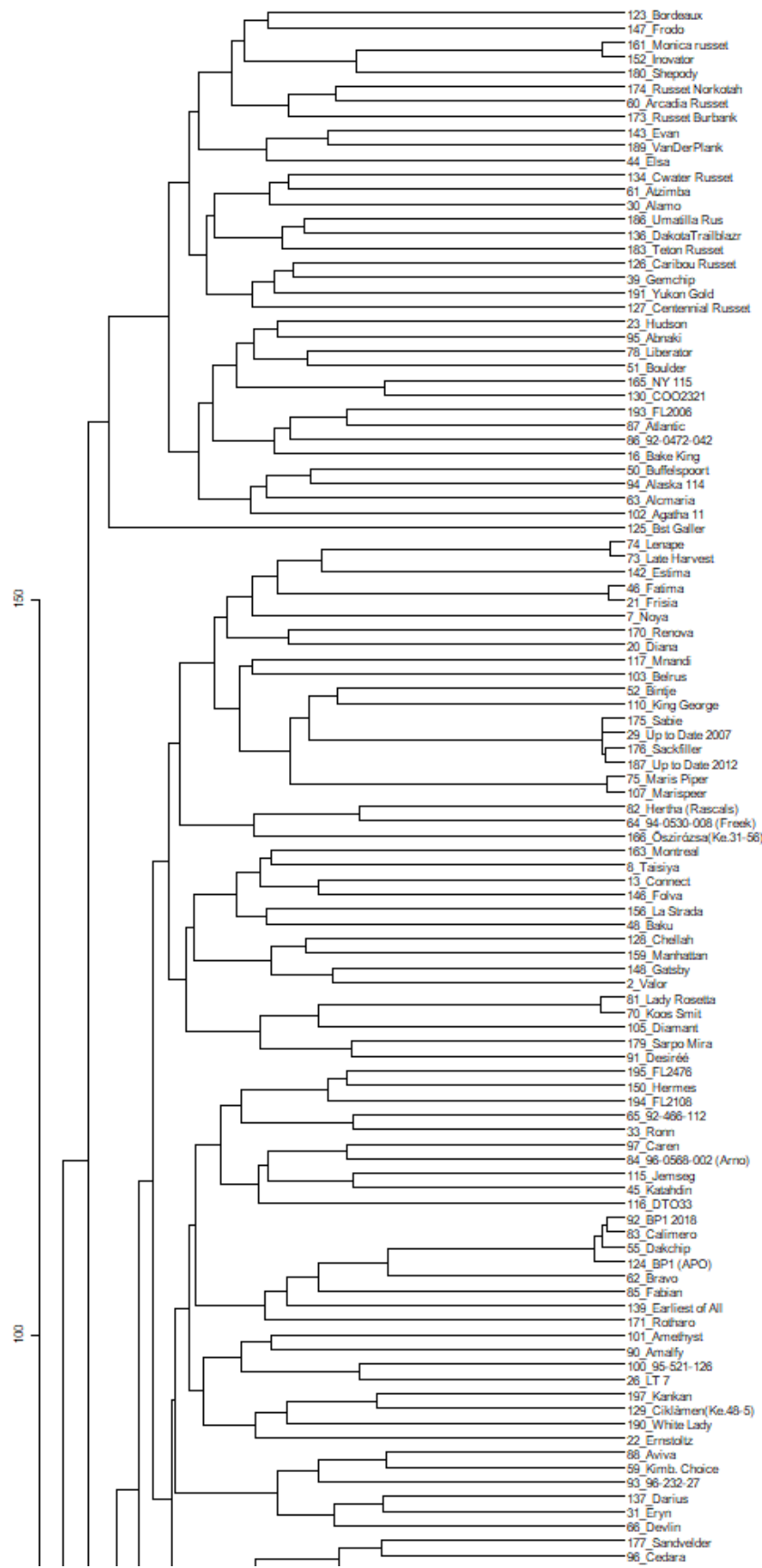
Figure 7. Histogram (frequency distribution) of pairs (Y) vs Kosman's distances (X) of 190 cultivars genotyped with SeqSNP at 500 SNP positions

Notes: Average distance = 0.265 ± 0.025

A dendrogram was constructed from the pairwise Kosman genetic distances to visually indicate the genetic relationships between the 190 cultivars genotyped using the 500 SNPs with SeqSNP (**Figure 8**).

Cluster analysis divided the 190 genotypes into three major groups with 36, 152, and 2 genotypes, respectively. In the first cluster, 36 genotypes were further divided into three sub-clusters. The second cluster contained 80% of the tested genotypes and was further divided into two big sub-clusters. The last cluster only contained Apache and Atacama.

Additionally, to the three duplicated cultivar pairs, Up to Date, BP1 and Mondial, other similar/identical cultivar pairs with very low genetic distances were observed. These included the following: Monica Russet & Innovator; Lenape & Late Harvest; Fatima & Frisia; Up to Date, Sabie and Sackfiller; Maris Piper & Marispeer; Lady Rosetta & Koos Smit; BP1, Calimero and Dakchip; Esco & Escourt; Hydra & Mokgotlong; Nicola & Nooksack; Advira & 890/20; Liseta & Norchip. In some cases, these pairs included material sourced from field-grown potatoes and from *in vitro* cultures.



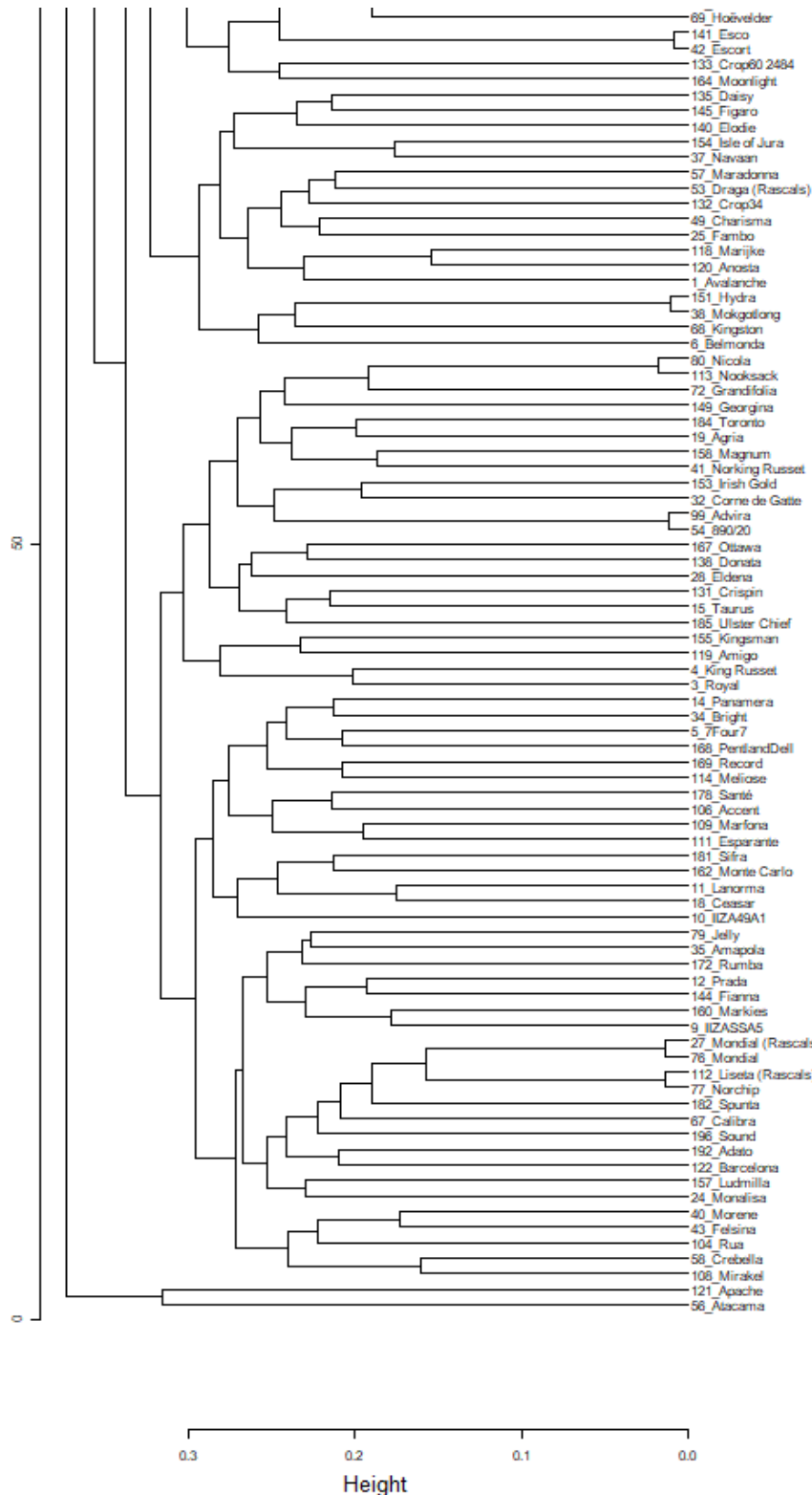


Figure 8. (page 1 and 2). Complete cluster dendrogram of pairwise genetic distances calculated with the Kosman index of 190 cultivars genotyped with SeqSNP at 500 SNP positions

The flanking sequence information obtained with SeqSNP was taken into account in selecting SNPs to be included in the KASP SNP assay design. The SNP diversity of the selected 25 SNP panel had REF ranging between 0.41 and 0.62 and all had PIC > 0.468, with 23 having PIC > 0.48. Distances between SNPs in the same linkage group were at least 0.58 Mb.

The pairwise Kosman genetic distances calculated from the SeqSNP genotypes of 190 cultivars genotyped with the 25 SNP panel indicate an upward shift in the genetic distance between the pairs, ranging between 0.13 and 0.44, and an average value of 0.271 ± 0.044 (Figure 9A). As a result, the panel is better at distinguishing cultivars based on genetic distance, despite using fewer SNP genotypes.

One member of each similar pair and duplicated samples were removed to yield 173 distinct cultivars. For the 173 unique samples, the genetic distances between the pairs ranged between 0.13 and 0.446, with an average of 0.274 ± 0.044 (Figure 9B).

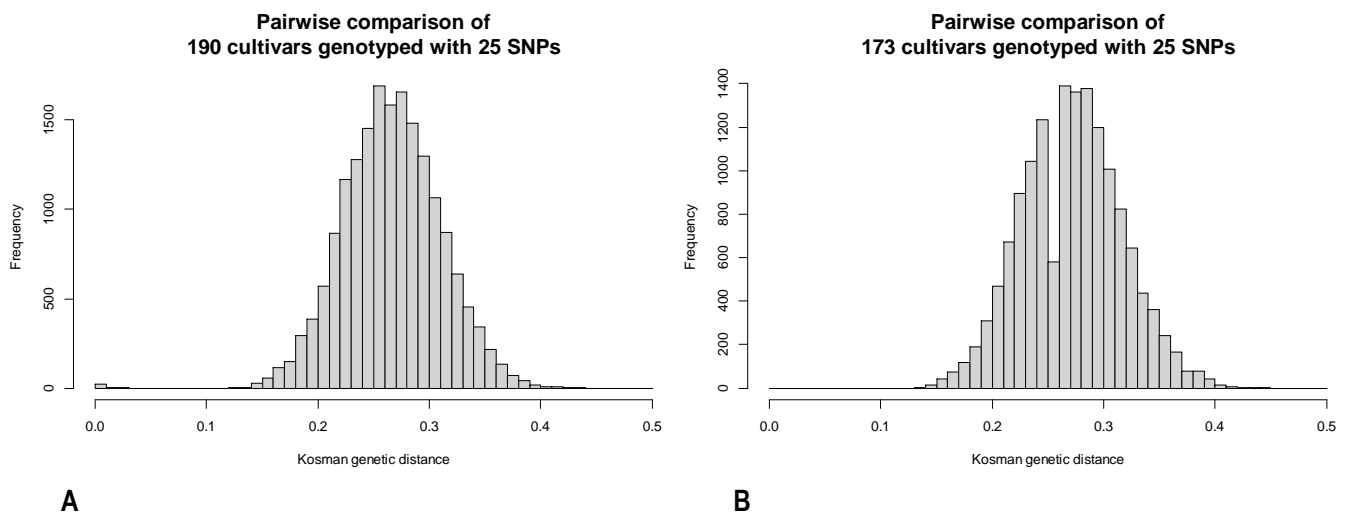
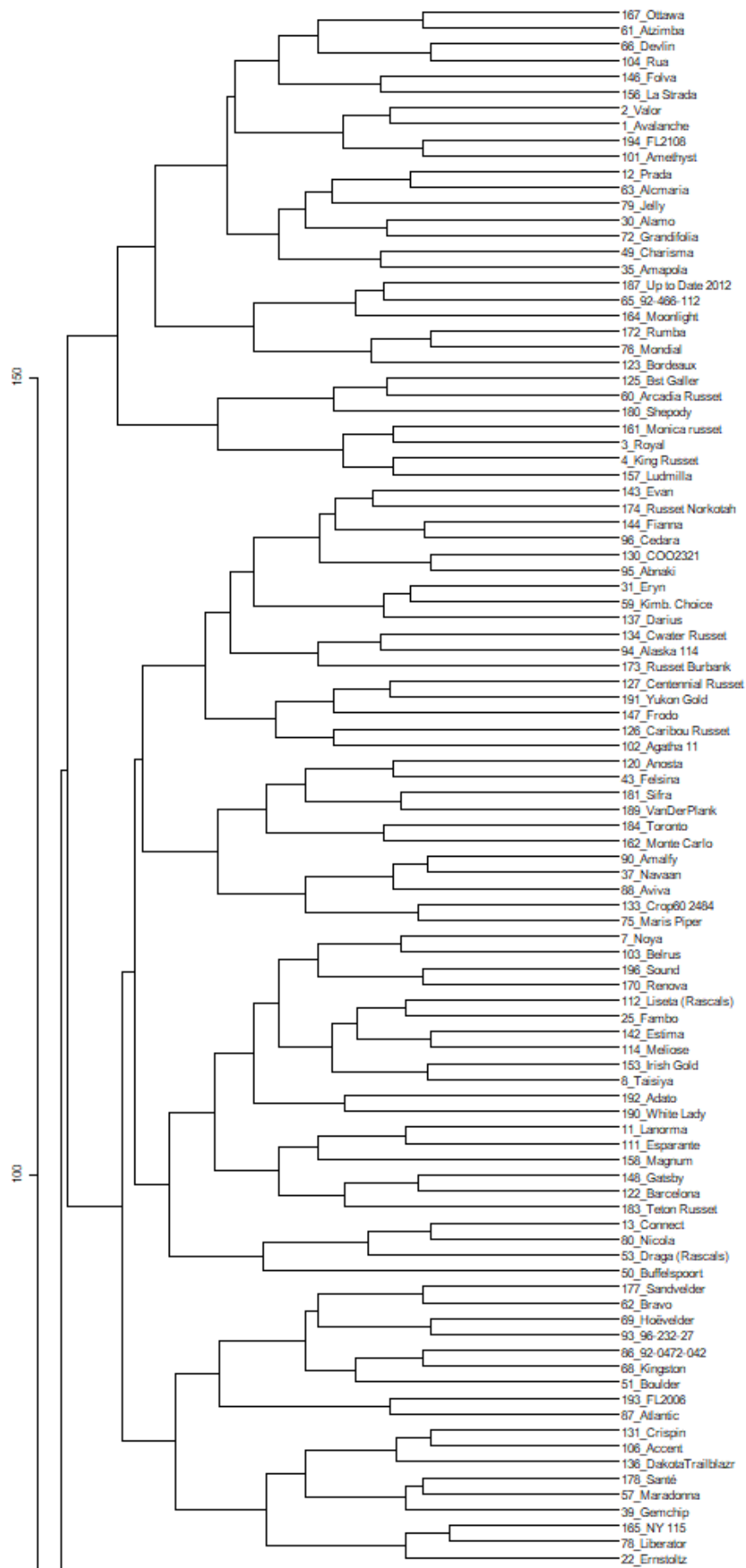


Figure 9. Histogram (frequency distribution) of pairs (Y) vs Kosman’s genetic distances (X) of potato cultivars genotyped with SeqSNP at selected panel of 25 SNPs

Notes: A. For all 190 cultivars, average distance = 0.271 ± 0.044 . B. For 173 unique cultivars, average distance = 0.274 ± 0.044 .

The ability of the selected panel to distinguish between unique cultivars (173 out of the 190) is demonstrated with a dendrogram of the pairwise genetic distances between cultivars using the selected 25 SNP panel (Figure 10). The 25 selected SNPs grouped the genotypes into two major clusters.

When the final 25 SNP panel was selected, the 23 pairs of similar/identical cultivars (identified previously) differed by only two or fewer allele dosages. In the other pairwise comparisons, all cultivars differed by at least 10 allele dosages ($D \geq 0.13$).



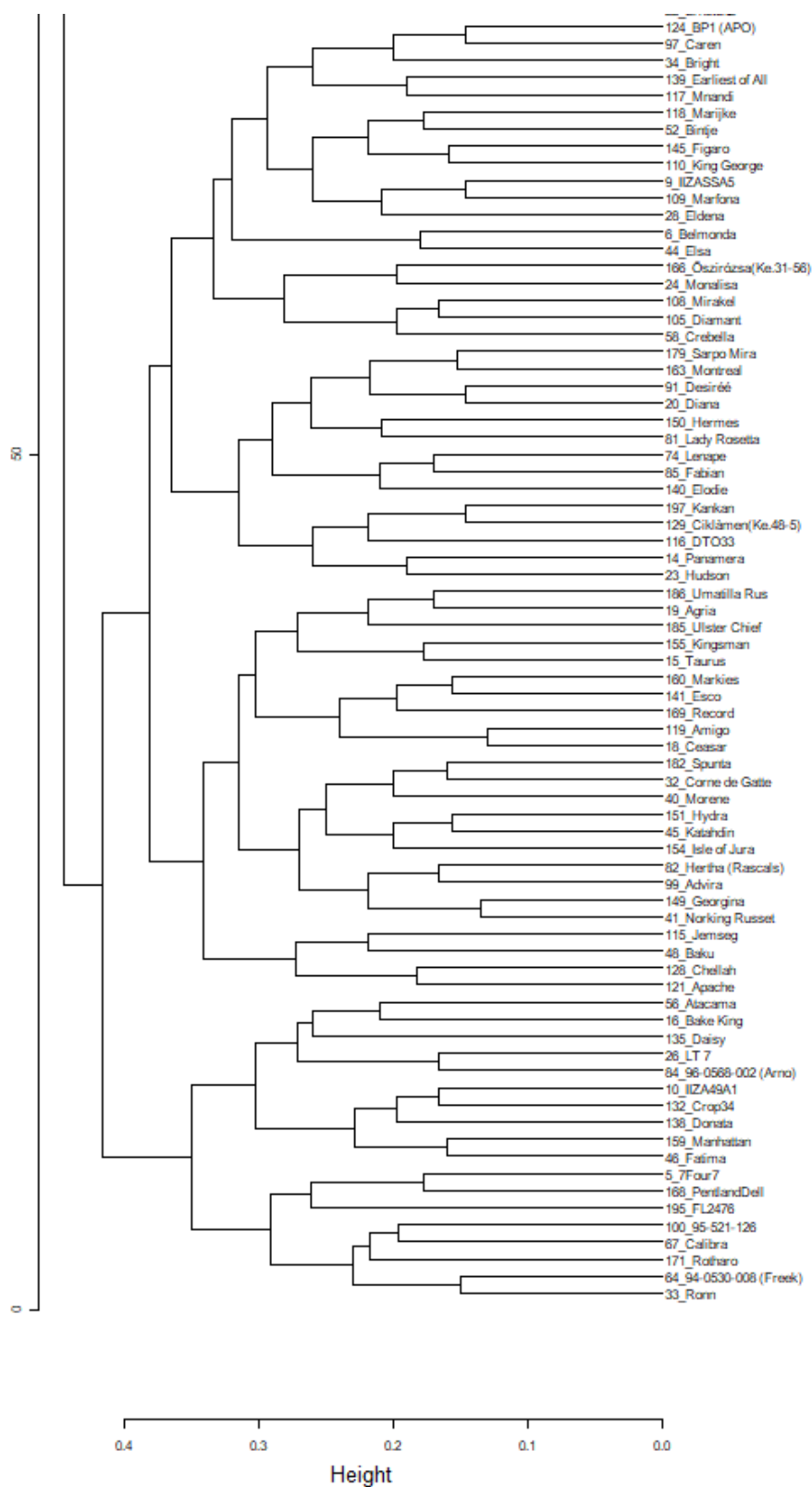


Figure 10. (page 1 and 2). Complete cluster dendrogram of pairwise genetic distances calculated with the Kosman index of 173 cultivars genotyped with SeqSNP at 25 selected SNP positions

Discussion

The application of SSR/SNP markers in crop improvement will depend on the quality of the information they provide regarding genetic diversity and population structure parameters. SNP fingerprinting provides both nucleotide genotype and copy number data for each allele, which is an advantage over SSR fingerprinting. SSR markers tend to have more alleles per locus than bi-allelic SNPs. Therefore, more SNPs (7-11 for maize per SSR) are needed to replace a single SSR. Due to their unique features, such as abundance in the genome and the ability to generate polymorphism at the single-base level, SNP markers are more cost-effective, technically feasible, and high throughput to measure. Chen (2016) proposed a panel size of 40-50 SNPs with a minor allele frequency between 40 and 60% for potatoes.

The population allele frequency affects the discriminatory power of SNP loci, where a balanced allele frequency (50%) yields a higher discriminatory power (maximum PIC = 0.5). However, selecting SNP loci with unbalanced allele frequencies (low PIC) can hardly identify differences between varieties (average Kosman similarity coefficients tend toward 1.00 as REF drops, according to Figure 2 in Chen, 2016). Another factor that affects SNP's discriminatory power is the number of markers in the panel. Chen (2016) found that a larger panel size (>50) did not significantly improve pairwise comparisons' average similarity value and variance. However, a too-small panel may result in many similar pairs (Figure 3 in Chen, 2016). Despite this, a panel of 25 SNPs was proposed for this study due to cost concerns.

For polyploid data, the Kosman similarity coefficient is superior to the Jaccard coefficient (for binary data, such as the presence/absence of an SSR allele) since it compares every allele dosage from different genotypes. It is, therefore, more sensitive to detect dissimilarity within a small SNP panel. As recommended by Chen (2016), a Kosman genetic similarity (1 - dissimilarity) of less than 0.85 fails to find a similar variety. Based on the chosen SNP panel, all unique cultivars were genetically distant from one another by at least 0.13.

Validate SeqSNP genotyping assay results of selected samples using KASP SNP assays of the selected SNP panel.

Materials and methods

Order KASP SNP assays

The completion of SeqSNP analysis enabled the procurement of an optimal panel of KASP SNP assays. With nearby SNPs marked in the flanking sequences, and the target SNP indicated with square brackets [/], the sequences were submitted to LGC Biosearch Technologies on 19 March 2021 for their KASP assay design software. They confirmed that they were able to design assays for all 25 SNPs.

gDNA isolation

Using the CTAB extraction method following standard laboratory protocols, genomic DNA was collected from field-grown, *in vitro*, and producer-supplied potato lines for verification with KASP SNP assays. The leaf samples (100 mg) were ground in the Genogrinder (SpexSamplePrep) at 1500 rpm for 4 minutes before adding the CTAB isolation buffer [2% CTAB, 1.5M NaCl, 20mM EDTA, 0.1M Tris.Cl, 0.2% β -mercaptoethanol]. A 30-minute incubation period at 60°C was followed by extraction with an equal volume of chloroform and isoamyl alcohol (24:1). The samples were centrifuged for 10 minutes at 10 000g and the supernatant was transferred to a clean tube. Genomic DNA was precipitated with 0.6 volumes of isopropanol for 30 minutes at -20°C. After centrifuging at 10 000g for 10 minutes, the pellet was washed with 70% ethanol. The DNA pellet was air dried and resuspended in

1x TE buffer overnight, and DNA concentrations were determined with the Nanodrop ND-1000. Distilled water was used to prepare dilutions of 20 ng/ul.

Running KASP SNP assays

KASP SNP assays were run according to the manufacturer's manuals and recommendations (LGC Biosearch Technologies 2013a, 2013b, 2014). 10 ul of KASP SNP assay reaction volumes consisting of 1x KASP genotyping master mix, 1x KASP probe mix, and 100 ng template gDNA per KASP SNP assay reaction were added to 96-well PCR plates (Biorad Hard-shell PCR plates 96-well, thin wall, clear well, HSP9601). Those cultivars that are important to the industry and represent all the allele dosage groups for each SNP marker were selected from the SeqSNP genotype dosage data (Appendix B). Plates were sealed with optically clear seals (Biorad Microseal 'B' adhesive sealing film, MSB1001). Reactions were run on a Biorad CFX96 Connect real-time PCR machine. The cycling conditions were 15 minutes at 94°C for hot-start Taq activation, a 2-step 65-57°C touchdown protocol over 10 cycles, and 26 cycles after touchdown at 94°C and 57°C. The end-point fluorescence data for FAM and HEX were read at 30°C.

Score SNP genotypes with appropriate software

R scripts were developed to analyse KASP SNP assay data. Scatter plots were constructed with the data points in the expected SeqSNP allele dosage colour. The software package in R called "fitPoly" (<https://cran.r-project.org/package=fitPoly>) was used for clustering and SNP dosage calling. It can fit the most accurate model for clustering polyploid genotyping data.

Results

Decisions on the most informative and best-performing SNPs needed to be made before an optimum panel of KASP SNP assays, able to discriminate between cultivars, could be ordered. LGC Biosearch Technologies cannot guarantee the success of the KASP SNP assay design or whether a successfully designed assay will produce a functional KASP SNP assay. The chosen SNP panel still needed to be validated by running real-time PCR experiments on each KASP SNP assay. The KASP SNP assays were ordered on 31 March 2021 and received on 6 July 2021.

The 25 KASP SNP assays were validated on 78 selected potato germplasm (**Appendix B**), with three duplicated and additional germplasm selected per marker to represent all the expected genotypic classes. **Figure 11 A** represents an ideal result in which KASP SNP assay values cluster into five distinct clusters. In this example of marker K, two genotypes did not cluster together as expected, as indicated by the coloured data points that do not match the rest of the cluster they grouped in.

Only two of the 25 KASP SNP assays failed to cluster into the five gene dosage classes (**Figure 11 B**). In the case of marker I, allele 2 (HEX) competed with allele 1 (FAM) probably due to allele 2 primers' preferential binding and amplification relative to allele 1. Conversely, Marker B exhibited preferential amplification of allele 1 (results not shown).

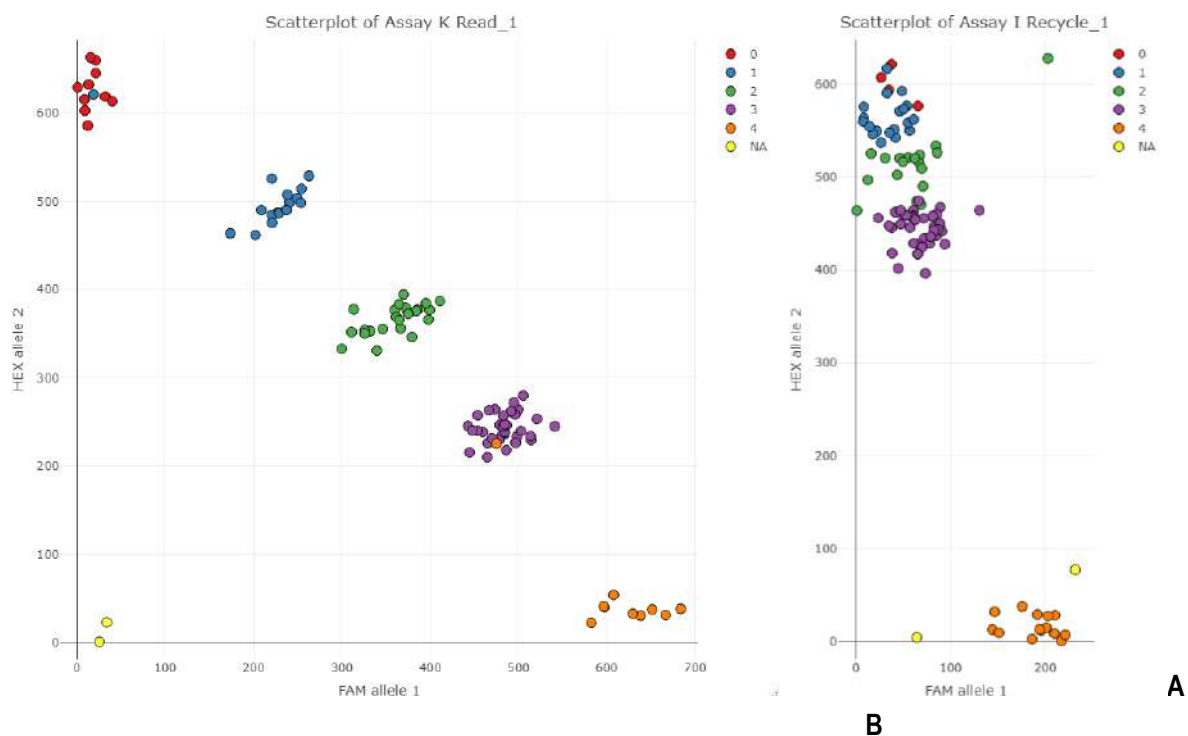


Figure 11. Examples of scatter plots constructed with KASP data points in the colour of the expected SeqSNP allele dosage

Notes: A. Example of a successful KASP SNP assay, Marker K, able to cluster potato genotypes into the five gene dosage classes; B. Scatter plot of one of the two unsuccessful KASP SNP assays, Marker I.

Figure 12 presents examples of plots that fitPoly generates as part of its output when clustering ratios between FAM and HEX fluorescence (“signal ratio”) is used to score the genotypes into the five dosage classes. The clusters of marker K in **Figure 12 A** are clearly discernible, resulting in confident scoring of the KASP dosage classes. **Figure 12 B** illustrates an example of a marker (marker G) where no models could be found to cluster the KASP data; hence, no genotyping of the dosage classes could be performed. When other “readfiles” (i.e., after the standard cycling or after additional cycles of re-cycling) of KASP data were analysed for the same marker, successful clustering could be obtained (results not shown) and dosage class genotyping could be performed.

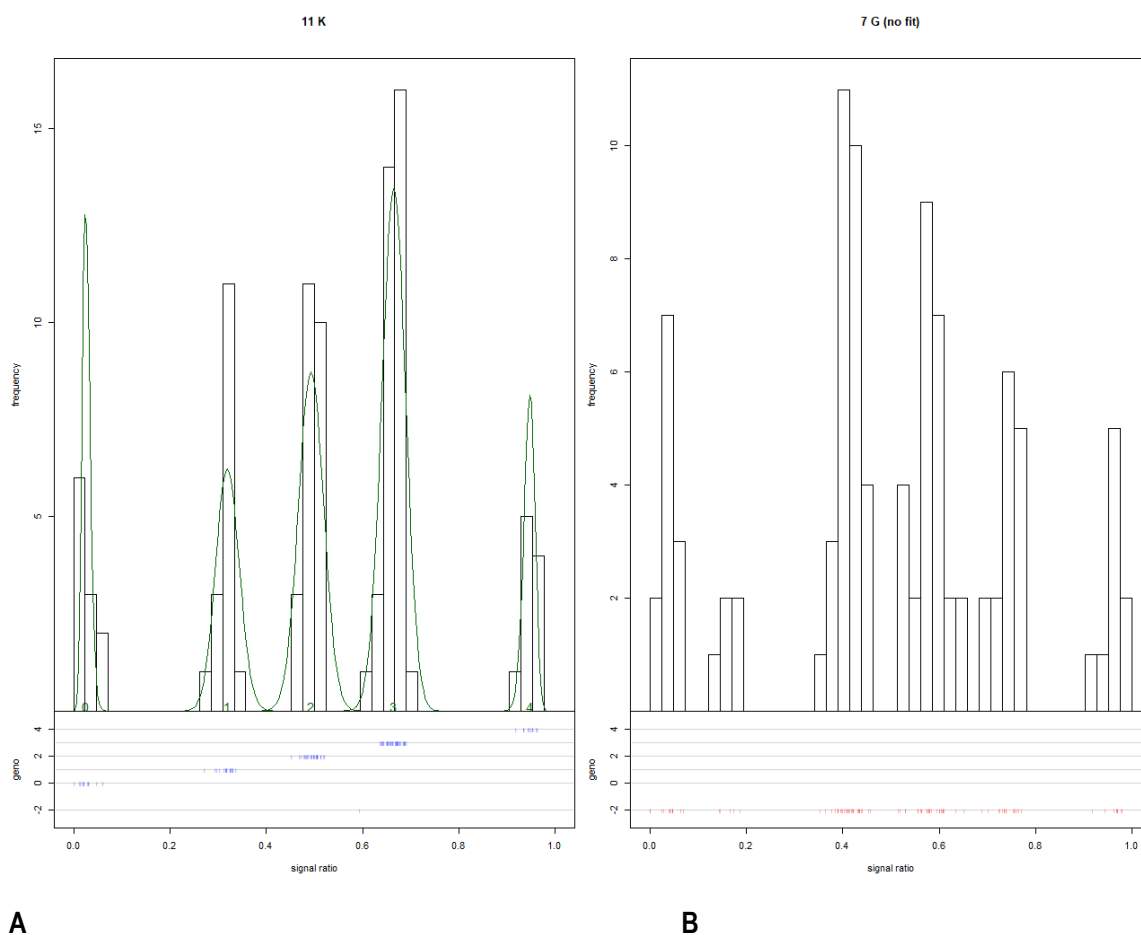


Figure 12. Examples of fitPoly output graphs, used to cluster and call the genotype of KASP results

Notes: A. KASP data could be clustered successfully, as with marker K. B. fitPoly could not find a model to cluster potato genotypes into five gene dosage classes, as with marker G.

Discussion

The two failed KASP markers (B and I) had no flanking variants that would have interfered with the assay. There was competition between one allele primer and the other, caused by preferential binding and amplification of one allele over the other. Therefore, these assays are not useful to genotype and classify potato cultivars into their dosage classes.

Since potato is a tetraploid, the original fitTetra tool only allows genotyping of autotetraploids and clustering of SNP genotypes into five gene dosages (Voorrips et al., 2011; Zych et al., 2019). In the meantime, an extension to higher levels of auto-polyploidy was implemented into a more advanced version of the package called fitPoly. This package was used to cluster and dose-call the KASP genotyping results.

Calculate genetic relatedness between cultivars and draw phylogenetic trees to indicate the relationships between the cultivars

Materials and methods

Kosman's index was used to calculate pairwise genetic distances between cultivars' KASP genotypes. Dendrograms were constructed from the allele dosage scores of the SNP markers using R software, the pairwise Kosman genetic distances and the complete clustering method. The SNP markers were also combined in the appropriate panels for some comparative analysis.

Results

Clustering of genotypes using Kosman's index was done to demonstrate the ability of the 25 SNP panel to distinguish between cultivars. The largest similarities were expected between duplicated pairs.

Since KASP markers B and I were unsuccessful, they were excluded from further analysis. The 78 samples that were genotyped with all the remaining 23 markers were selected. Cluster analysis using pairwise genetic distances (**Figure 13**) revealed that all cultivars could be distinguished from each other with at least one dosage difference, except in the following cases:

- One sample of Mondial (sample 76) differed from the others (27, 27_d, 76_d) with one dosage at Marker E. Both fitPoly and the scatterplot results scored this sample as dosage 2, even though it was expected to be 1 like the others.
- Maris Piper and Marispeer differed by only 1 dosage when compared with Marker M. These were members of the similar/identical pairs identified by SeqSNP genotyping using 500 SNPs. On the scatterplot of marker M, Maris Piper (sample 75) is one of the seven cultivars grouped in dosage class 3, while it was expected to be in dosage class 2. However, it was separated from the other dosage 3 class cultivars after increased cycle numbers (results not shown).
- No distinction could be made with the 23 SNP panel between Innovator and Monica Russet. They had also previously been identified as members of the similar/identical pairs.

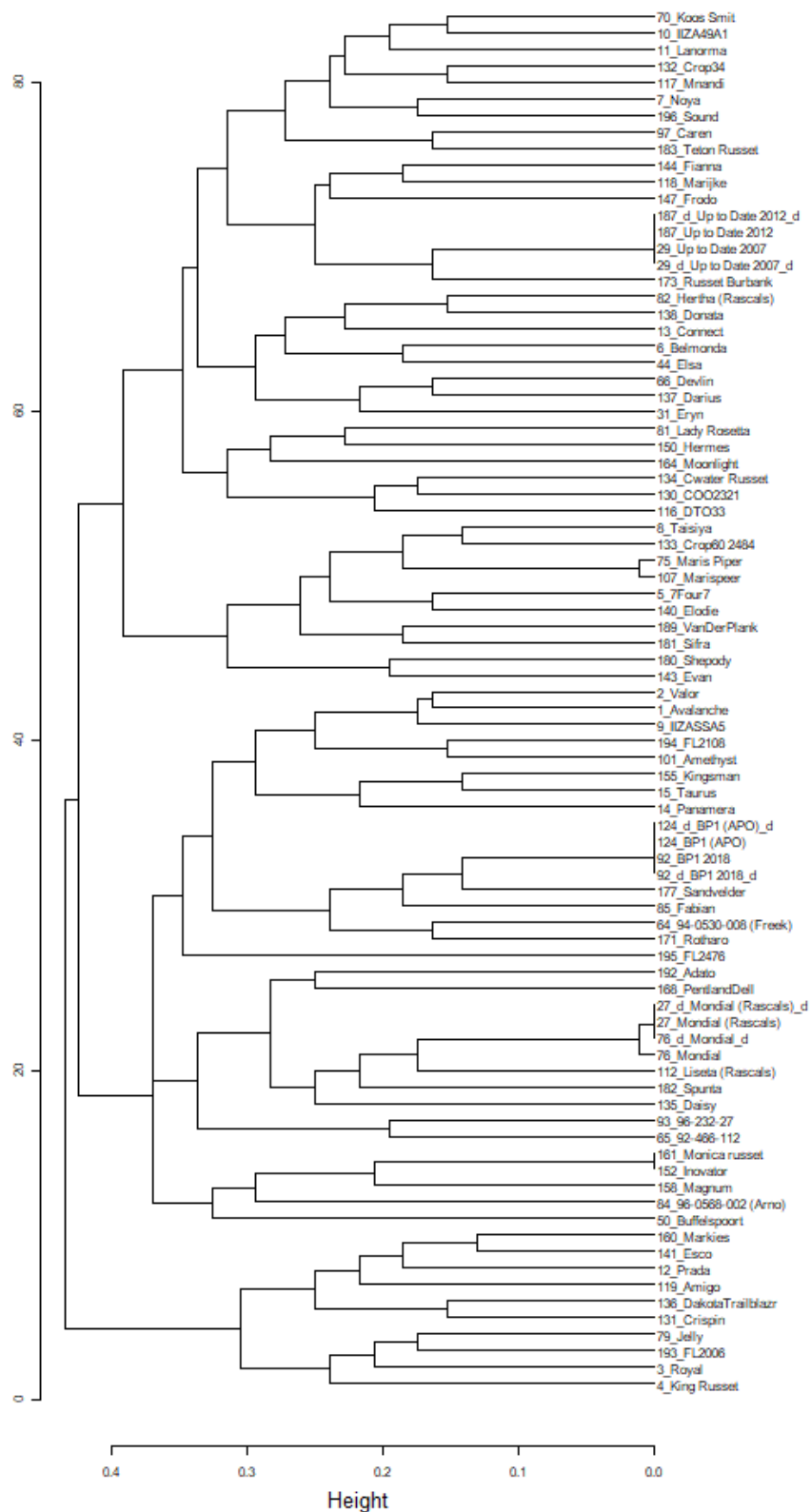


Figure 13. Complete cluster dendrogram of pairwise genetic distances of 78 selected commercially important cultivars genotyped with the 23 successful KASP SNP assays

Notes: Genetic distances were calculated with the Kosman index. Six cultivars were duplicated as indicated with “d” in their label.

KASP genotypes and SeqSNP genotypes of all the 190 samples were combined. A dendrogram made from pairwise genetic distances indicated that most cultivars' KASP genotype corresponded with their SeqSNP genotype, with a few exceptions (results not shown). Samples that matched a different cultivar's SeqSNP genotype will be repeated to confirm if DNA samples could have been mixed. For example, Koos Smit (sample 70) did not match its expected SeqSNP genotype with 15 out of the 23 KASP SNP markers. From the combined dendrogram, Koos Smit was most closely matched with Northing Russet (sample 41). Similarly, Atacama (sample 56) was most closely matched with Mnandi (sample 117); however, it was only tested with 7 KASP SNP markers.

Discussion

Chen (2016) suggested a Kosman's coefficient of 0.85 as a threshold for discriminating between similar and different varieties. Genetic similarities (similarity = 1 – distance) between non-duplicated potato samples using the 23 successful KASP SNP assays were in all cases above 0.87 (results not shown), indicating that all the 78 selected cultivars could be distinguished from each other.

Compile a database of SNP genotype profiles for all the entries of the germplasm database

Materials and methods

Set up of a SNP genotype database for all the entries in the germplasm database was done. Potato is a tetraploid, so the reference allele dosage is represented by a number between 0 and 4, as shown in **Figure 1**. Both expected SeqSNP and obtained KASP dosages (only cultivars assayed with 6 or more KASP SNP assays) were included in the database.

Results

The SNP genotype database will be useful to assign putative cultivar identity to samples submitted for fingerprinting by comparing their SNP genotypes to the germplasm SNP genotype database.

Not all the 190 samples were assayed with KASP SNP assays. Therefore, the SNP genotype database contains all the expected SNP genotypes and the KASP genotypes of samples assayed with 6 or more KASP SNP assays (**Appendix C**).

The obtained KASP results and expected SeqSNP genotypes were compared for the samples chosen for KASP verification. About 96% of the genotypes obtained with the 23 successful KASP SNP assays corresponded to the genotypes expected with SeqSNP. Only 4% (88 out of 2139) reactions resulted in different genotypes, as scored by fitPoly. Of the 23 successful KASP SNP assays, three had six or more mismatches between the expected SeqSNP genotype and obtained KASP genotype, each assayed with 93 samples. Those were assay G with six, M with 17 and H with 26 mismatches. All four KASP samples of the duplicated Up to Date samples (29 and 187) differed from the SeqSNP results at marker G. For BP1 duplicates (92 and 124), both 124 KASP samples differed from the rest at marker H.

After removing the two markers with high mismatch rates (H and M) from the panel, there were only 2.3% mismatches (45 out of 1953 samples) between KASP and SeqSNP genotypes.

Discussion

Marker O was expected to be the only marker to distinguish a single dosage difference between Monica Russet and Innovator. However, when both genotypes were assessed using the KASP SNP assays, they showed the same dosage. Consequently, the SNP panel fails to distinguish between these two cultivars.

Marker H resulted in 26 dosage differences from the expected over the 93 samples assayed, suggesting it may be detecting a different SNP than the target assayed by SeqSNP. Marker M had the second-highest number of mismatches (17) between the expected SeqSNP genotype and obtained KASP genotypes. Therefore, these two markers are suggested to be dropped and the remaining 21 SNP panel is proposed to be implemented for potato SNP fingerprinting. A dendrogram of pairwise genetic distances using the selected 21 SNP panel could still distinguish between all cultivars, except the previously detected similar or duplicated pairs (results not shown).

Implement the KASP SNP assays of the selected SNP panel for future genotyping of new germplasm or when purity test requests are submitted to ARC-VIMP

Materials and methods

A tool was developed in R software that calculates the differences in allele dosages between two cultivars of interest at each of the SNP markers in the panel. Both SeqSNP and KASP dosages were included, if available. The dosage differences per marker were sorted from high to low to allow ARC-VIMP to select the smallest appropriate subset of markers to use for fingerprinting if a purity test request is received.

Results

A tool was developed to assist in selecting the most appropriate subset of markers to use if ARC-VIMP receives a purity test request for fingerprinting. The input is the names of two cultivars that need to be distinguished with DNA fingerprinting. The output is a list of markers that will give the highest confidence in the results if a difference in genotype between two suspected cultivars is detected. With this technology, ARC-VIMP can generate scientifically significant SNP fingerprint profiles to distinguish a cultivar from a suspected case, requiring fewer markers than the total panel. For the client, this is beneficial.

An example is presented in **Table 3**, where the markers required to distinguish between “Amethyst” (sample 101) and “Agatha 11” (sample 102) were determined. “Amethyst” was fully analysed by all 21 KASP SNP assays; therefore, two comparisons could be made between the cultivars of interest. The absolute dosage difference was calculated for each SNP marker in the panel and sorted from high to low. The dosage differences were the same at each comparison (**Table 3**) since the KASP results of “Amethyst” corresponded to the expected SeqSNP genotypes at all the markers. In this example, the most appropriate markers for SNP fingerprinting would be at least markers A, S and E. Genotyping with these 3 markers only is expected to result in an 8-dosage difference, which gives high confidence that the two suspected cultivars are different. Genotyping with markers C, N, Q, or T is not recommended since it will not result in any dosage differences between the two cultivars.

Table 3. Selection of the appropriate markers to use to distinguish between “Amethyst” and “Agatha 11”

Marker	101_SeqSNP	101_KASP
	102_SeqSNP	102_SeqSNP
Dosage difference between pair		
A	3	3
S	3	3
E	2	2
H	2	2
O	2	2
V	2	2
X	2	2
D	1	1
F	1	1
G	1	1
J	1	1
K	1	1
L	1	1
M	1	1
P	1	1
R	1	1
U	1	1
W	1	1
Y	1	1
C	0	0
N	0	0
Q	0	0
T	0	0

Notes: Dosage differences were calculated per marker in the 21 SNP panel and sorted from high to low.

Discussion

The SNP genotype database includes both the SeqSNP and KASP genotypes since there are incidences where verification by KASP SNP assays did not match the expected SeqSNP genotypes. By using this tool, researchers will be able to select the most appropriate subset of markers if ARC-VIMP receives a purity test request for fingerprinting.

Assign putative cultivar identity to samples submitted for fingerprinting by comparing their SNP genotypes to the germplasm SNP genotype database

Materials and methods

A second tool developed in R software determines a query sample's most likely cultivar identity after genotyping it with several KASP markers. The possible identity of a given cultivar can be searched from the SNP genotype database and the members of the database with the lowest pairwise genetic difference can be identified. The probability that a match to a particular multiple-locus genotype would occur by chance will be calculated using the “product rule”, taking the specific marker and its REF into account.

Results

With the second tool, the ARC-VIMP can objectively assign putative cultivar identities to samples submitted for fingerprinting by comparing their SNP genotypes with the germplasm SNP genotype database.

The input is the KASP dosage score of the query sample that needs to be identified. The input file is combined with KASP and SeqSNP genotype databases. The pairwise Kosman genetic distances are then calculated and sorted, and the lowest pairwise genetic differences that contain the query sample are listed for possible identification.

An example of a cultivar called “Unknown_4” is presented in **Table 4**. **Table 5** contains the result of the top 10 hits with the lowest genetic difference to the query sample. Occasionally, hits to KASP genotyped samples will occur if one or two marker genotypes matched perfectly to the query, in this case only marker O, although other marker KASP genotypes are absent since KASP was not verified for all 21 markers for some samples (**Table 6**). Therefore, it is presumed that the putative identity of “Unknown_4” is one of the two SeqSNP genotypes, in this case, Eldena or Desirée. According to the complete cluster dendrogram of genetic differences, “Unknown_4” showed no genetic difference from Eldena and Desirée (**Figure 14**).

Table 4. Example input file of KASP SNP assay dosage score results of a query sample

SampleName	Marker	Dosage
Unknown_4	X	2
Unknown_4	Y	2
Unknown_4	F	2
Unknown_4	O	3

Table 5. Top 10 lowest Kosman genetic difference hits to the query sample “Unknown_4”

Match in SNP genotype database	Kosman genetic distance
SeqSNP_28_Eldena	0.000
SeqSNP_91_Desir���	0.000
KASP_166_��szir��zsa(Ke.31-56)	0.000
KASP_175_Sabie	0.000
KASP_176_Sackfiller	0.000
SeqSNP_106_Accent	0.0625
SeqSNP_39_Gemchip	0.0625
SeqSNP_90_Amalfy	0.0625
SeqSNP_40_Morene	0.0625
SeqSNP_100_95-521-126	0.0625

Table 6. SNP genotypes of the top 10 lowest Kosman genetic difference hits to the query sample “Unknown_4”

gDNA #	Cultivar name	Genotype	SNP marker			
			X	F	Y	O
	Unknown_4	KASP	2	2	2	3
91	Desir���	SeqSNP	2	2	2	3
28	Eldena	SeqSNP	2	2	2	3
166	��szir��zsa(Ke.31-56)	KASP	NA	NA	NA	3
175	Sabie	KASP	NA	NA	NA	3
176	Sackfiller	KASP	NA	NA	NA	3
106	Accent	SeqSNP	2	2	2	2
39	Gemchip	SeqSNP	2	3	2	3
90	Amalfy	SeqSNP	1	2	2	3
40	Morene	SeqSNP	2	2	2	2
100	95-521-126	SeqSNP	2	2	3	3

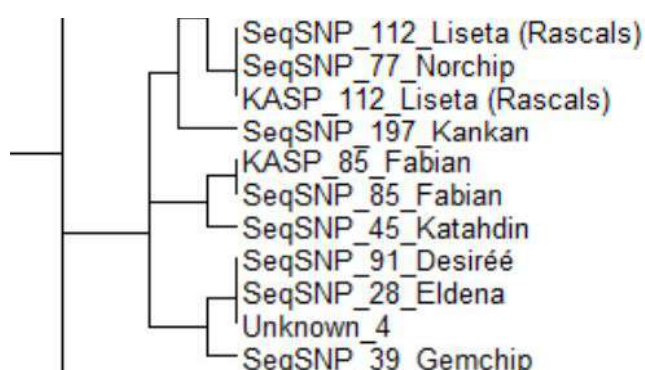


Figure 14. Segment of a complete cluster dendrogram of pairwise Kosman genetic distances between members of the SNP genotype database and the query sample “Unknown_4”

Notes: Genotypes at four SNP positions (X, F, Y and O) were used.

The “product rule” is used to calculate the probability that a match to a particular multiple-locus genotype would occur by chance. Per-locus genotype frequencies are multiplied together to determine the match probability. For a

diploid organism, the addition of a factor of 2 for each heterozygous locus is also included. Tetraploids are calculated differently but follow the same rule. For tetraploid individuals, the expected frequencies of genotype classes in progeny after random mating can be mono-allelic for one allele (AAAA) $R_{iiii} = p_i^4$, bi-allelic simplex (AAAB) $R_{iiij} = 4p_i^3p_j$, bi-allelic duplex (AABB) $R_{iiij} = 6p_i^2p_j^2$, bi-allelic triplex (ABBB) $R_{ijjj} = p_i4p_j^3$ and mono-allelic for the second allele (BBBB) $R_{jjjj} = p_j^4$. If the REF (p_i) is 0.5, the probability of a variety having 0, 1, 2, 3 or 4 allele dosages can be computed as 0.0625, 0.25, 0.375, 0.25 and 0.0625, respectively.

The probability that a random sample of potatoes in a population will have a particular DNA profile is therefore dependent on the number of markers used and the allele frequency in the potato population. If we use more markers, or the rarer the allele frequencies are, the lower the match probability.

Table 7 illustrates the application of the product rule where the frequencies of the per-locus genotypes (or independent SNP markers in this case) are multiplied together to get the match probability. Random match probabilities are interpreted as the one in X chance of an unrelated cultivar having the same DNA profile as the unknown sample. Given the REF of each marker, there is a one in 87 chance that an unrelated cultivar will have the same DNA profile as “Unknown_4” purely by chance.

Table 7. Calculation of the random match probability of a query sample after genotyping with a small number of SNP markers

Sample	Genotype: KASP dosage per marker				Match probability	Interpretation
	X	F	Y	O		
Unknown_4	2	2	2	3		
Genotype frequency per locus:					Match probability	Interpretation
if REF = $p_i = 0.5$	0.375	0.375	0.375	0.25		
use specific allele frequency for marker	0.3495	0.3671	0.3318	0.2704	0.01151	one in 87

Notes: This calculation illustrates the difference between using REF as 0.5 or as the genotype allele frequency as determined for 190 potato cultivars.

Discussion

For bi-allelic markers, such as the SNPs employed here, a REF value of 0.5 corresponds to the theoretical maximum, resulting in the highest polymorphism information content (PIC) of 0.5. SNP markers were specifically selected for this project to have a balanced allele frequency. All SNPs selected for the panel had REFs between 0.41 and 0.62 among 190 samples, and all PICs were above 0.46.

The probability of 0 (zero) dosage differences between two samples, if REF is 0.5, is 0.27 (Chen, 2016). Therefore, the probability of 0 dosage differences between two samples over 10 markers in potato is 0.27^{10} , corresponding to one pair from a panel of 986 samples being indistinguishable by chance alone.

Chapter 3

RECOMMENDATION AND IMPACT

Recommendation

The implementation of a panel of 21 SNP markers for potato fingerprinting at the ARC is proposed, after eliminating the ineffective KASP SNP markers B, H, I, and M. This panel is able to distinguish between all potato cultivars, except the previously detected similar or duplicated pairs.

Impacts of the project

The development of a SNP genotype database for a large number of potato cultivars, which are crucial for the South African potato industry, was one of the main outputs of this project. SNP genotypes simplify the germplasm genotype database and enable us to objectively compare the genetic identity of the unknown cultivar to the databased genotypes to determine the suggested identity of the cultivar.

The KASP SNP assays developed for the selected SNP panel are suitable for genotyping samples locally as new germplasm, or purity test requests, are submitted to the ARC-VIMP. DNA fingerprinting based on SNP technology will streamline the process. Compared to conventional SSR and PAGE, the new technology will offer improved efficiency, reliability, sensitivity, higher throughput and lower cost per sample.

ARC-VIMP currently charges R250 per SSR marker for potato fingerprinting, and three SSR primers are usually required to confirm trueness-to-type. Therefore, the price is R930, including gDNA isolation per sample. The ARC's revised costing model has to account for personnel and overhead costs, which were not adequately reflected in the outdated SSR prices. Control samples and water controls will be needed for each KASP experiment, and there are some fixed costs and labour components involved. When only three SNPs (for scientific robustness) are required to discriminate between two known cultivars, the price will be less than currently from 4 samples and up, and almost half of the current price for 12 samples. It is necessary to genotype new cultivars with all 21 SNP assays to add it to the SNP genotype database, instead of the 5 SSR markers traditionally performed. The price will match SSR at 12 samples, so it is recommended to expand the database every 2-3 years with 12 or more samples rather than submitting samples one at a time.

ARC-VIMP will continue to serve the South African potato industry, including private farmers, public and private laboratories, by providing genetic fingerprinting services for clonal identification and trueness-to-type determination of potato cultivars.

REFERENCES

- Anderson, J. A., Churchill, G. A., Autrique, J. E., Tanksley, S. D., & Sorrells, M. E. (1993). Optimizing parental selection for genetic linkage maps. *Genome* 36(1): 181-186.
- Bali, S., Sathuvalli, V., Brown, C. R., Novy, R. G., Ewing, L., Debons, J., Douches, D. S., Coombs, J. J., Navarre, D. A., Whitworth, J. L., Charlton, B. A., Yilma, S., Shock, C. C., Stark, J. C., Pavek, M. J., & Knowles, N. R. (2017). Genetic fingerprinting of potato varieties from the Northwest Potato Variety Development Program. *American Journal of Potato Research* 94: 54-63. <https://doi.org/10.1007/s12230-016-9547-z>
- Bourke, P. M., Voorrips, R. E., Visser, R. G., & Maliepaard, C. (2018). Tools for genetic studies in experimental populations of polyploids. *Frontiers in plant science* 9: 513. <https://doi.org/10.3389/fpls.2018.00513>
- Chen, L. (2016). Potato Variety Identification with a Panel of SNP Markers. MSc Thesis, Plant Breeding, Wageningen University.
- Chen, W., Mingus, J., Mammadov, J., Backlund, J. E., Greene, T., Thompson, S., & Kumpatla, S. (2010). KASPar: a simple and cost-effective system for SNP genotyping. In: *Proceedings of plant and animal genomes XVIII conference*. San Diego, p 94
- Coombs, J. J., Frank, L. M., & Douches, D. S. (2004). An applied fingerprinting system for cultivated potato using simple sequence repeats. *American Journal of Potato Research* 81: 243-250.
- Deperi, S. I., Tagliotti, M. E., Bedogni, M. C., Manrique-Carpintero, N. C., Coombs, J., Zhang, R., Douches, D. S., & Huarte, M. (2018). Discriminant analysis of principal components and pedigree assessment of genetic diversity and population structure in a tetraploid potato panel using SNPs. *PLoS ONE* 13(3): e0194398.
- Ellis, D., Chavez, O., Coombs, J., Soto, J., Gomez, R., Douches, D., Panta, A., Silvestre, R., & Anglin, N. L. (2018). Genetic identity in genebanks: application of the SolCAP 12K SNP array in fingerprinting and diversity analysis in the global in trust potato collection. *Genome* 61: 523–537.
- Felcher, K. J., Coombs, J. J., Massa, A. N., Hansey, C. N., Hamilton, J. P., Veilleux, R. E., Buell, C. R., & Douches, D. S. (2012). Integration of two diploid potato linkage maps with the potato genome sequence. *PLoS One* 7: e36347.
- Food and Agriculture Organization of the United Nations. (2019). FAOSTAT statistical database. FAO. URL: <http://www.fao.org/faostat/en/#data/QV>. Accessed 24 August 2021.
- Garrison, E., & Marth, G. (2012). "Haplotype-based variant detection from short-read sequencing". In: arXiv preprint arXiv:1207.3907. URL: <http://github.com/ekg/freebayes>.
- Gruber, B., & Adamack, A. (2014). Introduction to PopGenReport using PopGenReport Ver.2.0. p 1-29.
- Grunwald, N., Kamvar, Z. N., & Everhart, S. E. (2016). grunwaldlab/Population_Genetics_in_R: First release [Data set]. URL: https://grunwaldlab.github.io/Population_Genetics_in_R/.

- Hamilton, J. P., Hansey, C. N., Whitty, B. R., Stoffel, K., Massa, A. N., Van Deynze, A., De Jong, W. S., Douches, D. S., & Buell, C. R. (2011). Single nucleotide polymorphism discovery in elite North American potato germplasm. *BMC genomics* 12: 1.
- Jamali, S. H., Cockram, J., & Hickey, L. T. (2019). Insights into deployment of DNA markers in plant variety protection and registration. *Theoretical and Applied Genetics* 132(7): 1911-1929. doi: 10.1007/s00122-019-03348-7
- Knaus, B. J., & Grunwald, N. J. (2017). VcfR: An R package to manipulate and visualize VCF format data. *Molecular Ecology Resources* 17: 44-53.
- Kosman, E., & Leonard, K. J. (2005). Similarity coefficients for molecular markers in studies of genetic relationships between individuals for haploid, diploid, and polyploidy species. *Molecular Ecology* 14: 415-424.
- LGC Biosearch Technologies. (2013a). Guide to running KASP genotyping on the BIO-RAD CFX-series instruments. URL: <https://biosearch-cdn.azureedge.net/assetsv6/running-KASP-on-BIO-RAD-CFX.pdf>. Accessed 10 March 2021.
- LGC Biosearch Technologies. (2013b). KASP genotyping chemistry User guide and manual. URL: <https://biosearch-cdn.azureedge.net/assetsv6/KASP-genotyping-chemistry-User-guide.pdf>. Accessed 13 May 2021.
- LGC Biosearch Technologies. (2014). KASP thermal cycling conditions. URL: <https://biosearch-cdn.azureedge.net/assetsv6/KASP-thermal-cycling-conditions-all-protocols.pdf>. Accessed 13 May 2021.
- Liao, H., & Guo, H. (2014). Using SSR to evaluate the genetic diversity of potato cultivars from Yunnan Province (SW China). *Acta Biologica Cracoviensia Series Botanica* 56(1): 16-27.
- Lindhout, P., Meijer, D., Schotte, T., Hutten, R. C. B., Visser, R. G. F., & van Eck, H. J. (2011). Towards F1 hybrid seed potato breeding. *Potato Research* 54: 301-312.
- McGregor, C. E., Greyling, M. M., & Warnich, L. (2000a). The use of simple sequence repeats (SSRs) to identify commercially important potato (*Solanum tuberosum* L.) cultivars in South Africa. *South African Journal of Plant and Soil* 17(4): 177-179.
- McGregor, C. E., Lambert, C. A., Greyling, M. M., Louw, J. H., & Warnich, L. (2000b). A comparative assessment of DNA fingerprinting techniques (RAPD, ISSR, AFLP and SSR) in tetraploid potato (*Solanum tuberosum* L.) germplasm. *Euphytica* 113: 135-144.
- Moisan-Thiery, M., Marhadour, S., Kerlan, M. C., Dessenne, N., Perramant, M., Gokelaere, T., & Le Hingrat, Y. (2005). Potato cultivar identification using simple sequence repeats markers (SSR). *Potato Research* 48: 191-200.
- Neelam, K., Brown-Guedira, G., & Huang, L. (2013). Development and validation of a breeder-friendly KASPar marker for wheat leaf rust resistance locus *Lr21*. *Molecular Breeding* 31: 233.
- Potato Certification Service. (2019). *2018-2019 annual report*. URL: <https://potatocertification.co.za/annual-report/>.

- Reid, A., Hof, L., Felix, G., Ruecker, B., Tams, S., Milczynska, E., Esselink, D., Uenk, G., Vosman, B., & Weitz, A. (2011). Construction of an integrated microsatellite and key morphological characteristic database of potato varieties on the EU Common Catalogue. *Euphytica* 182: 239–249.
- Sasaki, H., Sanetomo, R., & Hosaka, K. (2017). A system for identification of potato varieties using SNP dosage. *American Journal of Potato Research* 94(5): 513.
- Semagn, K., Babu, R., Hearne, S., & Olsen, M. (2014). Single nucleotide polymorphism genotyping using kompetitive allele specific PCR (KASP): overview of the technology and its application in crop improvement. *Molecular Breeding* 33(1): 1–14.
- Sharma, S. K., Bolser, D., de Boer, J., Sønderkær, M., Amoros, W., Carboni, M. F., D'Ambrosio, J. M., de la Cruz, G., Di Genova, A., Douches, D. S., Eguiluz, M., Guo, X., Guzman, F., Hackett, C. A., Hamilton, J. P., Li, G., Li, Y., Lozano, R., Maass, A., ... Bryan, G. J. (2013). Construction of reference chromosome-scale pseudomolecules for potato: Integrating the potato genome with genetic and physical maps. *G3: Genes, Genomes, Genetics* 3: 2031-2047.
- Singh, R., Iquebal, M. A., Mishra, C. N., Jaiswal, S., Kumar, D., Raghav, N., & Tiwari, V. (2019). Development of model web-server for crop variety identification using throughput SNP genotyping data. *Scientific reports* 9(1): 5122.
- Steele, K. A., Quinton-Tulloch, M. J., Amgai, R. B., Dhakal, R., Khatiwada, S. P., Vyas, D., Heine, M., & Witcombe, J. R. (2018). Accelerating public sector rice breeding with high-density KASP markers derived from whole genome sequencing of indica rice. *Molecular breeding: new strategies in plant improvement* 38(4): 38. <https://doi.org/10.1007/s11032-018-0777-2>
- The Potato Genome Sequencing Consortium. (2011). Genome sequence and analysis of the tuber crop potato. *Nature* 475: 189–195. <https://doi.org/10.1038/nature10158>.
- Uitdewilligen, J. (2012). Discovery and genotyping of existing and induced DNA sequence variation in potato. Wageningen University.
- Uitdewilligen, J. G., Wolters, A. M. A., D'hoop, B. B., Borm, T. J., Visser, R. G., & Van Eck, H. J. (2013). A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. *PloS one* 8(5): e62355.
- Voorrips, R. E., Gort, G., & Vosman, B. (2011). Genotype calling in tetraploid species from bi-allelic marker data using mixture models. *BMC Bioinformatics* 12: 172. <https://doi.org/10.1186/1471-2105-12-172>
- Vos, P. (2016). Development and application of a 20K SNP array in potato. PhD Thesis, Plant Breeding, Wageningen University.
- Vos, P. G., Uitdewilligen, J. G., Voorrips, R. E., Visser, R. G., & van Eck, H. J. (2015). Development and analysis of a 20K SNP array for potato (*Solanum tuberosum*): an insight into the breeding history. *Theoretical and applied genetics* 128(12): 2387-2401.
- Zych, K., Gort, G., Maliepaard, C. A., Jansen, R. C., & Voorrips, R. E. (2019). FitTetra 2.0 – improved genotype calling for tetraploids with multiple population and parental data support. *BMC Bioinformatics* 20: 148.

OUTPUTS

1. KASP SNP assays verified for fingerprinting of potato cultivars in the certification process of seed potatoes.
2. Report-back at the annual meetings of the Nuclear Material Producers (NUMPRO), regarding the progress of the SNP fingerprinting project.
3. Presentation titled “Migrating potato fingerprinting from microsatellite markers to a panel of single nucleotide polymorphisms” on 9 December 2020 at the LGC Biosearch Technologies webinar with the theme “How agrigenomics ‘SuperSheroes’ (women scientists) are seeding global change”.
http://info.biosearchtech.com/supersheroes-championing-modern-agrigenomic-technology?utm_source=pardot&utm_medium=email&utm_campaign=gglobal20apr_kasp&utm_content=cta

Anticipated outputs

1. Provide fingerprinting services to clients in the potato industry using the new SNP panel.
2. Informative/ popular article for publication in industry-specific press, namely CHIPS, explaining the similarities and differences between the old PAGE method for SSR fragment analysis and modern SNP-based method for potato cultivar fingerprinting and reporting on the outcome and use of this new potato fingerprinting method.
3. Scientific article on the development and validation of the method using a custom SNP panel for the fingerprinting of potato.

Appendix A. (pages 1-5). Potato germplasm list selected for SeqSNP genotyping (190 cultivars)

GWK, FPD and FL at the “Reason for choice” refers to contributions made by GWK Trading, First Potato Dynamics and Pepsico (FL lines) respectively.

gDNA#	Cultivar name	Source	Reason for choice
5	7Four7	Greenhouse – dried	GWK
54	890/20	Field	Did SSR
86	92-0472-042	Field	Did SSR
65	92-466-112	Field	Did SSR
64	94-0530-008 (Freek)	Field	F. Steyn request
100	95-521-126	Field	Did SSR
84	96-0568-002 (Arno)	Field	McCain request
93	96-232-27	Field	Did SSR
95	Abnaki	Field	Choose more
106	Accent	Field	Choose more
192	Adato	<i>In vitro</i> planted in greenhouse	FPD
99	Advira	Field	Choose more
102	Agatha 11	Field	Choose more
19	Agria	Field	Vos gt
30	Alamo	Field	Choose more
94	Alaska 114	Field	Choose more
63	Alcmaria	Field	Vos gt
90	Amalfy	Field	Choose more
35	Amapola	Field	Choose more
101	Amethyst	Field	F. Steyn request
119	Amigo	<i>In vitro</i>	McCain request
120	Anosta	<i>In vitro</i> planted in greenhouse	Vos gt
121	Apache	<i>In vitro</i>	Choose more
60	Arcadia Russet	Field	Choose more
56	Atacama	Field	Choose more
87	Atlantic	Field	Vos gt
61	Atzimba	Field	Choose more
1	Avalanche	Greenhouse	FPD
88	Aviva	Field	SA Variety list
16	Bake King	Field	Choose more
48	Baku	Field	Choose more
122	Barcelona	<i>In vitro</i>	Choose more
6	Belmonda	Greenhouse	GWK
103	Belrus	Field	Choose more
52	Bintje	Field	Vos gt
123	Bordeaux	<i>In vitro</i>	Choose more
51	Boulder	Field	Choose more

124	BP1 (APO)	<i>In vitro</i>	Commercially important
92	BP1 2018	Field	Commercially important
62	Bravo	Field	F. Steyn request
34	Bright	Field	Choose more
125	Bst Galler	<i>In vitro</i>	Choose more
50	Buffelspoort	Field	Commercially important
67	Calibra	Field	SA Variety list
83	Calimero	Field	Choose more
97	Caren	Field	Commercially important
126	Caribou Russet	<i>In vitro</i>	Did SSR
18	Ceasar	Field	Choose more
96	Cedara	Field	Choose more
127	Centennial Russet	<i>In vitro</i>	Choose more
49	Charisma	Field	Choose more
128	Chellah	<i>In vitro</i>	SA Variety list
129	Ciklâmen(Ke.48-5)	<i>In vitro</i>	Choose more
13	Connect	Greenhouse	GWK
130	COO2321	<i>In vitro</i>	Did SSR
32	Corne de Gatte	Field	F. Steyn request
58	Crebella	Field	Did SSR
131	Crispin	<i>In vitro</i>	Did SSR
132	Crop34	<i>In vitro</i>	SA Variety list
133	Crop60 2484	<i>In vitro</i>	McCain request
134	Cwater Russet	<i>In vitro</i>	McCain request
135	Daisy	<i>In vitro</i>	Commercially smaller
55	Dakchip	Field	Choose more
136	Dakota Trailblazr	<i>In vitro</i>	McCain request
137	Darius	<i>In vitro</i>	Commercially important
91	Desir��e	Field	Choose more
66	Devlin	Field	SA Variety list
105	Diamant	Field	Vos gt
20	Diana	Field	Choose more
138	Donata	<i>In vitro</i>	SA Variety list
53 A	Draga (Rascals)	Field	Choose more
116	DTO33	Field	F. Steyn request
139	Earliest of All	<i>In vitro</i>	Choose more
28 (B)	Eldena	Field	Vos gt
140	Elodie	<i>In vitro</i>	Did SSR
44	Elsa	Field	F. Steyn request
22	Ernstoltz	Field	Choose more
31	Eryn	Field	Commercially smaller
141	Esco	<i>In vitro</i>	SA Variety list
42	Escort	Field	Vos gt
111	Esparante	Field	Choose more

142	Estima	<i>In vitro</i> planted in greenhouse	Vos gt
143	Evan	<i>In vitro</i>	SA Variety list
85	Fabian	Field	SA Variety list
25	Fambo	Field	Choose more
46	Fatima	Field	Choose more
43	Felsina	Field	Vos gt
144	Fianna	<i>In vitro</i> planted in greenhouse	Commercially important
145	Figaro	<i>In vitro</i>	SA Variety list
193	FL2006	<i>In vitro</i> planted in greenhouse	FL
194	FL2108	<i>In vitro</i> planted in greenhouse	FL
195	FL2476	<i>In vitro</i>	FL
146	Folva	<i>In vitro</i>	Vos gt
21	Frisia	Field	Vos gt
147	Frodo	<i>In vitro</i>	SA Variety list
148	Gatsby	<i>In vitro</i>	Choose more
39	Gemchip	Field	Choose more
149	Georgina	<i>In vitro</i>	Did SSR
72	Grandifolia	Field	Choose more
150	Hermes	<i>In vitro</i>	Commercially smaller
82	Hertha (Rascals)	Field	Commercially important
69	Hoëvelder	Field	SA Variety list
23	Hudson	Field	Choose more
151	Hydra	<i>In vitro</i>	Choose more
10	IIZA49A1	Greenhouse	GWK
9	IIZASSA5	Greenhouse	GWK
152	Innovator	<i>In vitro</i> planted in greenhouse	Commercially important
153	Irish Gold	<i>In vitro</i>	Did SSR
154	Isle of Jura	<i>In vitro</i>	Choose more
79	Jelly	Field	SA Variety list
115	Jemseg	Field	Choose more
197	Kankan	<i>In vitro</i> planted in greenhouse	Choose more
45	Katahdin	Field	Vos gt
59	Kimb. Choice	Field	Choose more
110	King George	Field	Vos gt
4	King Russet	Greenhouse	GWK
155	Kingsman	<i>In vitro</i>	Did SSR
68	Kingston	Field	Choose more
70	Koos Smit	Field	F. Steyn request
156	La Strada	<i>In vitro</i>	Did SSR
81	Lady Rosetta	Field	Commercially important
11	Lanorma	Greenhouse	GWK
73	Late Harvest	Field	Choose more
74	Lenape	Field	Vos gt
78	Liberator	Field	Choose more

112 (B)	Liseta (Rascals)	Field	Commercially smaller
26	LT 7	Field	F. Steyn request
157	Ludmilla	<i>In vitro</i>	SA Variety list
158	Magnum	<i>In vitro</i>	McCain request
159	Manhattan	<i>In vitro</i>	Did SSR
57	Maradonna	Field	Choose more
109	Marfona	Field	Vos gt
118	Marijke	Field	SA Variety list
75	Maris Piper	Field	Vos gt
107	Marispeer	Field	Choose more
160	Markies	<i>In vitro</i>	FPD
114	Meliose	Field	Choose more
108	Mirakel	Field	Vos gt
117	Mnandi	Field	Commercially smaller
38	Mokgotlong	Field	Choose more
24	Monalisa	Field	Vos gt
76	Mondial	Field	Commercially important
27	Mondial (Rascals)	Field	Commercially important
161	Monica russet	<i>In vitro</i>	McCain request
162	Monte Carlo	<i>In vitro</i>	Vos gt
163	Montreal	<i>In vitro</i>	Choose more
164	Moonlight	<i>In vitro</i> planted in greenhouse	Did SSR
40	Morene	Field	Vos gt
37	Navaan	Field	Choose more
80	Nicola	Field	SA Variety list
113	Nooksack	Field	Choose more
77	Norchip	Field	Choose more
41	Norking Russet	Field	Choose more
7	Noya	Greenhouse	GWK
165	NY 115	<i>In vitro</i>	Choose more
166	Őszirózsa(Ke.31-56)	<i>In vitro</i>	Choose more
167	Ottawa	<i>In vitro</i>	SA Variety list
14	Panamera	Leaf sample – dried	Commercially important
168	PentlandDell	<i>In vitro</i>	Commercially important
12	Prada	Greenhouse	GWK
169	Record	<i>In vitro</i>	Vos gt
170	Renova	<i>In vitro</i>	Choose more
33	Ronn	Field	F. Steyn request
171	Rotharo	<i>In vitro</i>	SA Variety list
3	Royal	Greenhouse	GWK
104	Rua	Field	Choose more
172	Rumba	<i>In vitro</i>	Choose more
173	Russet Burbank	<i>In vitro</i>	SA Variety list
174	Russet Norkotah	<i>In vitro</i>	Vos gt

175	Sabie	<i>In vitro</i>	Choose more
176	Sackfiller	<i>In vitro</i> planted in greenhouse	Choose more
177	Sandvelder	<i>In vitro</i>	SA Variety list
178	Santé	<i>In vitro</i> planted in greenhouse	Choose more
179	Sarpo Mira	<i>In vitro</i>	Vos gt
180	Shepody	<i>In vitro</i>	Commercially smaller
181	Sifra	<i>In vitro</i>	Commercially important
196	Sound	<i>In vitro</i> planted in greenhouse	FPD
182	Spunta	<i>In vitro</i>	SA Variety list
8	Taisiya	Greenhouse – dried	GWK
15	Taurus	Leaf sample – dried	Commercially important
183	Teton Russet	<i>In vitro</i>	McCain request
184	Toronto	<i>In vitro</i>	Choose more
185	Ulster Chief	<i>In vitro</i> planted in greenhouse	Choose more
186	Umatilla Rus	<i>In vitro</i>	Choose more
29	Up to Date 2007	Field	Commercially important
187	Up to Date 2012	<i>In vitro</i>	Commercially important
2	Valor	Greenhouse	Commercially important
189	VanDerPlank	<i>In vitro</i>	Commercially important
190	White Lady	<i>In vitro</i>	Choose more
191	Yukon Gold	<i>In vitro</i>	Vos gt

Appendix B. (pages 1-2). Potato germplasm list selected for KASP SNP assay verification (78 cultivars)

GWK, FPD and FL at the “Reason for choice” refer to contributions made by GWK Trading, First Potato Dynamics and Pepsico (FL lines), respectively.

gDNA#	Cultivar Name	Source	Reason for choice
5	7Four7	Greenhouse – dried	GWK
65	92-466-112	Field	Did SSR
64	94-0530-008 (Freek)	Field	F. Steyn request
84	96-0568-002 (Arno)	Field	McCain request
93	96-232-27	Field	Did SSR
192	Adato	<i>In vitro</i> planted in greenhouse	FPD
101	Amethyst	Field	F. Steyn request
119	Amigo	<i>In vitro</i>	McCain request
1	Avalanche	Greenhouse	FPD
6	Belmonda	Greenhouse	GWK
124	BP1 (APO)	<i>In vitro</i>	Commercially important
92	BP1 2018	Field	Commercially important
50	Buffelspoort	Field	Commercially important
97	Caren	Field	Commercially important
13	Connect	Greenhouse	GWK
130	COO2321	<i>In vitro</i>	Did SSR
131	Crispin	<i>In vitro</i>	Did SSR
132	Crop34	<i>In vitro</i>	SA Variety list
133	Crop60 2484	<i>In vitro</i>	McCain request
134	Cwater Russet	<i>In vitro</i>	McCain request
135	Daisy	<i>In vitro</i>	Commercially smaller
136	DakotaTrailblazr	<i>In vitro</i>	McCain request
137	Darius	<i>In vitro</i>	Commercially important
66	Devlin	Field	SA Variety list
138	Donata	<i>In vitro</i>	SA Variety list
116	DTO33	Field	F. Steyn request
140	Elodie	<i>In vitro</i>	Did SSR
44	Elsa	Field	F. Steyn request
31	Eryn	Field	Commercially smaller
141	Esco	<i>In vitro</i>	SA Variety list
143	Evan	<i>In vitro</i>	SA Variety list
85	Fabian	Field	SA Variety list
144	Fianna	<i>In vitro</i> planted in greenhouse	Commercially important
193	FL2006	<i>In vitro</i> planted in greenhouse	FL
194	FL2108	<i>In vitro</i> planted in greenhouse	FL
195	FL2476	<i>In vitro</i>	FL
147	Frodo	<i>In vitro</i>	SA Variety list
150	Hermes	<i>In vitro</i>	Commercially smaller
82	Hertha (Rascals)	Field	Commercially important

10	IIZA49A1	Greenhouse	GWK
9	IIZASSA5	Greenhouse	GWK
152	Innovator	<i>In vitro</i> planted in greenhouse	Commercially important
79	Jelly	Field	SA Variety list
4	King Russet	Greenhouse	GWK
155	Kingsman	<i>In vitro</i>	Did SSR
70	Koos Smit	Field	F. Steyn request
81	Lady Rosetta	Field	Commercially important
11	Lanorma	Greenhouse	GWK
112	Liseta (Rascals)	Field	Commercially smaller
158	Magnum	<i>In vitro</i>	McCain request
118	Marijke	Field	SA Variety list
75	Maris Piper	Field	Vos gt
107	Marispeer	Field	Choose more
160	Markies	<i>In vitro</i>	FPD
117	Mnandi	Field	Commercially smaller
76	Mondial	Field	Commercially important
27	Mondial (Rascals)	Field	Commercially important
161	Monica russet	<i>In vitro</i>	McCain request
164	Moonlight	<i>In vitro</i> planted in greenhouse	Did SSR
7	Noya	Greenhouse	GWK
14	Panamera	Leaf sample - dried	Commercially important
168	PentlandDell	<i>In vitro</i>	Commercially important
12	Prada	Greenhouse	GWK
171	Rotharo	<i>In vitro</i>	SA Variety list
3	Royal	Greenhouse	GWK
173	Russet Burbank	<i>In vitro</i>	SA Variety list
177	Sandvelder	<i>In vitro</i>	SA Variety list
180	Shepody	<i>In vitro</i>	Commercially smaller
181	Sifra	<i>In vitro</i>	Commercially important
196	Sound	<i>In vitro</i> planted in greenhouse	FPD
182	Spunta	<i>In vitro</i>	SA Variety list
8	Taisiya	Greenhouse - dried	GWK
15	Taurus	Leaf sample - dried	Commercially important
183	Teton Russet	<i>In vitro</i>	McCain request
29	Up to Date 2007	Field	Commercially important
187	Up to Date 2012	<i>In vitro</i>	Commercially important
2	Valor	Greenhouse	Commercially important
189	VanDerPlank	<i>In vitro</i>	Commercially important

Appendix C (pages 1-4). Database of SNP genotypes of selected potato cultivars as obtained from SeqSNP and KASP SNP assays at 23 SNP positions

The reference allele dosage is indicated as a number between 0 and 4.

gDNA#	Cultivar name	Genotype	KASP SNP assay																							
			A	C	D	E	F	G	H	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	
5	7Four7	SeqSNP	4	0	3	3	3	0	3	1	3	3	2	2	3	1	3	2	3	2	0	3	2	2	3	
5	7Four7	KASP	4	0	3	3	3	0	3	1	3	3	2	2	3	1	3	2	3	2	0	3	2	2	3	
54	890/20	SeqSNP	3	1	3	1	2	1	3	3	4	2	3	4	2	1	1	2	2	2	2	2	3	3	1	
86	92-0472-042	SeqSNP	1	1	1	1	2	2	3	1	1	2	2	3	2	1	1	2	1	2	0	1	2	2	3	
65	92-466-112	SeqSNP	3	2	1	0	4	1	0	1	1	0	3	2	3	1	0	2	3	2	2	1	3	3	3	
65	92-466-112	KASP	3	2	1	0	4	1	0	1	1	0	3	2	3	1	0	2	3	2	2	1	3	3	3	
64	94-0530-008 (Freek)	SeqSNP	2	0	2	1	2	2	2	1	2	2	3	2	4	2	1	1	1	1	1	1	2	3	3	
64	94-0530-008 (Freek)	KASP	2	0	2	1	2	2	2	1	2	2	3	2	4	2	1	1	1	1	1	1	2	3	3	
100	95-521-126	SeqSNP	2	0	1	0	2	3	2	1	1	2	1	2	3	0	1	3	3	2	0	3	2	3	3	
84	96-0568-002 (Arno)	SeqSNP	3	4	2	2	1	2	3	2	1	4	0	3	1	1	3	2	1	4	3	1	2	2	4	
84	96-0568-002 (Arno)	KASP	3	4	2	2	1	2	3	2	1	4	0	3	1	1	3	2	1	4	3	1	2	2	4	
93	96-232-27	SeqSNP	3	3	0	0	1	1	1	2	1	1	1	1	2	2	1	2	2	1	2	1	1	3	3	
93	96-232-27	KASP	3	3	0	0	1	1	1	2	1	1	1	1	2	2	1	2	2	1	2	1	1	3	3	
95	Abnaki	SeqSNP	1	1	0	1	2	0	0	2	2	0	2	1	2	2	3	0	3	2	2	2	2	2	4	
106	Accent	SeqSNP	2	0	2	2	2	3	3	4	1	2	2	2	2	2	2	4	2	2	2	3	1	2	2	
192	Adato	SeqSNP	3	3	2	3	4	4	3	0	1	3	1	1	1	1	1	1	4	1	1	3	1	4	3	
192	Adato	KASP	3	3	2	3	4	4	3	0	1	3	1	1	1	1	1	1	4	1	1	3	1	4	3	
99	Advira	SeqSNP	3	1	3	1	2	1	3	3	4	2	3	4	2	1	1	2	2	2	2	2	3	3	1	
102	Agatha 11	SeqSNP	1	2	0	4	2	1	3	2	3	1	0	2	4	1	2	0	4	1	1	1	4	1	1	
19	Agria	SeqSNP	2	1	2	2	3	4	3	3	2	2	3	1	1	2	3	2	2	3	1	2	3	2	2	
30	Alamo	SeqSNP	3	2	1	4	3	0	1	3	3	2	2	2	1	1	1	3	1	3	1	3	2	1	3	
94	Alaska 114	SeqSNP	1	2	0	2	4	1	3	4	2	0	3	1	2	2	2	0	3	1	3	1	1	2	2	
63	Alcmaria	SeqSNP	3	2	1	2	3	2	3	2	2	2	4	3	1	1	1	1	1	2	3	2	3	2	2	
90	Amalfy	SeqSNP	1	3	1	1	2	1	3	2	2	2	1	1	3	3	1	2	2	2	1	1	3	1	2	
35	Amapola	SeqSNP	3	4	0	2	3	3	2	3	3	2	1	2	1	2	0	0	3	3	1	2	2	2	1	
101	Amethyst	SeqSNP	4	2	1	2	3	2	1	1	2	2	1	2	2	0	2	1	1	1	0	3	3	3	2	
101	Amethyst	KASP	4	2	1	2	3	2	1	1	2	2	1	2	2	0	2	1	1	1	0	3	3	3	2	
119	Amigo	SeqSNP	3	1	0	3	3	4	0	2	2	3	3	1	3	2	2	3	1	2	2	3	3	3	2	
119	Amigo	KASP	3	1	0	3	3	4	0	2	2	3	3	1	3	2	2	3	1	2	2	3	3	3	2	
120	Anosta	SeqSNP	1	3	3	1	1	1	3	2	1	3	1	2	2	1	4	2	1	1	2	3	3	2	1	
121	Apache	SeqSNP	1	2	3	4	2	2	3	2	2	2	3	3	3	1	1	3	4	3	0	1	4	2	3	
60	Arcadia Russet	SeqSNP	2	3	1	3	2	0	2	1	0	3	3	2	3	3	4	2	2	2	2	2	1	1	3	
56	Atacama	SeqSNP	2	2	3	2	2	3	4	2	1	2	4	4	2	1	3	0	4	1	1	2	3	3	3	
56	Atacama	KASP	NA	NA	NA	NA	NA	1	2	NA	3	2	NA	2	NA	NA	NA	NA	3	2	NA	NA	NA	NA	NA	
87	Atlantic	SeqSNP	2	0	0	0	0	4	3	2	2	2	2	2	2	3	1	1	1	2	1	1	3	1	2	
87	Atlantic	KASP	NA	0	0	0	0	4	NA	NA	NA	NA	NA	NA	NA	3	NA	NA	1	NA	NA	NA	1	NA		
61	Atzimba	SeqSNP	0	3	2	1	3	2	1	3	3	3	1	1	3	1	2	2	2	2	2	2	3	1	4	
1	Avalanche	SeqSNP	2	3	3	2	3	2	3	2	3	0	2	3	1	0	2	0	1	1	1	2	2	2	2	
1	Avalanche	KASP	2	3	3	2	3	2	3	2	3	0	2	3	1	0	2	0	1	1	1	2	2	2	2	
88	Aviva	SeqSNP	1	3	3	1	2	0	3	2	1	1	0	3	3	2	2	2	2	2	3	0	2	4	2	
16	Bake King	SeqSNP	2	2	0	0	1	2	3	3	2	3	3	3	4	3	3	3	1	2	1	1	2	3	3	
48	Baku	SeqSNP	2	1	3	3	3	1	4	3	3	2	2	1	4	2	3	3	2	1	1	0	2	4	2	
122	Barcelona	SeqSNP	2	1	3	2	1	2	4	2	2	3	2	2	1	1	0	2	1	1	1	3	4	2	3	2
6	Belmonda	SeqSNP	1	3	3	4	2	3	1	3	4	3	3	2	1	2	2	1	2	1	1	2	2	2	3	
6	Belmonda	KASP	1	3	3	4	2	3	1	3	4	3	3	2	1	2	1	1	2	1	1	2	2	2	3	
103	Belrus	SeqSNP	3	3	3	2	3	1	3	1	0	3	3	1	2	2	1	1	3	3	3	1	0	2	4	
52	Binje	SeqSNP	4	3	3	3	2	1	3	1	3	2	3	0	3	2	2	2	3	2	2	2	2	4	3	
123	Bordeaux	SeqSNP	2	1	2	1	3	0	2	3	1	1	3	0	3	4	1	1	2	2	2	1	1	3	3	
51	Boulder	SeqSNP	1	0	0	2	1	2	2	1	2	1	1	1	1	1	3	2	1	0	1	1	3	2	1	
124	BP1 (APO)	SeqSNP	2	3	1	2	1	2	3	1	3	2	1	1	4	2	3	1	2	1	0	0	2	2	4	
124	BP1 (APO)	KASP	2	3	1	2	1	2	2	1	3	2	1	1	4	2	3	1	2	1	0	0	2	2	4	
124_d	BP1 (APO)_d	SeqSNP	2	3	1	2	1	2	3	1	3	2	1	1	4	2	3	1	2	1	0	0	2	2	4	
124_d	BP1 (APO)_d	KASP	2	3	1	2	1	2	2	1	3	2	1	1	4	2	3	1	2	1	0	0	2	2	4	
92	BP1 2018	SeqSNP	2	3	1	2	1	2	2	1	3	2	1	1	4	2	3	1	2	1	0	0	2	2	4	
92	BP1 2018	KASP	2	3	1	2	1	2	2	1	3	2	1	1	4	2	3	1	2	1	0	0	2	2	4	
92_d	BP1 2018_d	SeqSNP	2	3	1	2	1	2	2	1	3	2	1	1	4	2	3	1	2	1	0	0	2	2	4	
92_d	BP1 2018_d	KASP	2	3	1	2	1	2	2	1	3	2	1	1	4	2	3	1	2	1	0	0	2	2	4	
62	Bravo	SeqSNP	3	2	1	2	1	1	3	0	1	2	2	1	3	2	3	2	2	2	0	1	1	1	3	
34	Bright	SeqSNP	2	2	3	1	1	2	3	1	3	2	2	3	3	1	2	0	4	2	1	3	2	2	3	
125	Bst Galler	SeqSNP	2	1	2	2	1	0	3	1	0	2	2	1	3	3	3	4	1	1	3	2	1	2	2	
50	Buffelspoort	SeqSNP	1	2	0	3	3	3	2	1	1	2	1	3	0	0	3	1	3	2	3	1	4	0	2	
50	Buffelspoort	KASP	1	2	0	3	3	3	2	1	1	2	1	3	0	0	2	1	3	2	3	1	4	0	2	
67	Calibra	SeqSNP	3	2	2	2	1	2	2	0	0	2	1	2	4	0	2	1	3	1	2	1	2	3	3	
83	Calimero	SeqSNP	2	3	1	2	1	2	3	1	3	2	1	1	4	2	3	1	2	1	0	0	2	2	4	
97	Caren	SeqSNP	2	2	1	2	1	2	3	2	3	3	1	1	2	0	3	1	3	2	2	2	2	1	4	
97	Caren	KASP	2	2	1	2	1	2	3	NA	3	3	1	1	2	0	3	1	3	2	2	2	2	1	4	
126	Caribou Russet	SeqSNP	2	2	3	2	1	1	3	4	2	1	1	2	2	2	1	1	3	0	0	1	3	1	0	
126	Caribou Russet	KASP	2	NA	NA	NA	1	NA	NA	4	2	1	1	NA	NA	NA	1	NA	NA	0	0	1	3	1	0	
18	Ceasar	SeqSNP	2	1	2	3	3	4	2	2	3	2	3	1	3	3	3	1	1	2	2	2	2	4	1	
96	Cedara	SeqSNP	3	2	0	2	1	1	0	2	2	2	2	1	2	1	0	2	2	3	2	2	2	2	2	
127	Centennial Russet	SeqSNP	2	1	2	3	4	1	2	4	2	3	1	1	1	1	2	0	4	0	2	3	3	2	4	
127	Centennial Russet	KASP	NA	1	NA	NA	4	NA	2	4	NA	NA	NA	NA	NA	NA	NA	4	0	NA	NA	NA	NA	NA	NA	
4																										

KASP SNP assay

gDNA#	Cultivar name	Genotype	A	C	D	E	F	G	H	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	
128	Chellah	SeqSNP	2	4	2	2	3	1	3	2	3	2	3	3	0	2	2	3	2	3	1	3	3	3	3	
129	Ciklāmen (Ke.48-5)	SeqSNP	1	2	1	2	2	1	3	2	3	3	0	3	3	2	2	2	4	2	2	2	1	3	1	
13	Connect	SeqSNP	0	1	3	2	1	1	2	0	3	2	2	3	0	1	1	2	3	2	2	1	3	3	1	
13	Connect	KASP	0	1	2	2	1	1	0	3	2	3	3	1	1	1	2	3	2	2	1	3	3	1		
130	COO2321	SeqSNP	2	1	1	1	3	0	1	1	3	2	2	1	2	3	4	1	3	2	2	2	2	2	2	
130	COO2321	KASP	2	1	1	1	3	0	1	1	3	2	2	1	2	3	4	1	3	2	2	2	2	2	2	
32	Corne de Gatte	SeqSNP	2	2	3	2	3	1	3	1	0	3	2	3	2	2	2	2	2	3	2	2	4	1	3	
58	Crebella	SeqSNP	3	3	1	3	4	1	1	1	4	2	3	3	1	1	2	3	2	1	3	3	0	3	2	
131	Crispin	SeqSNP	3	0	2	2	2	1	1	3	4	1	3	0	2	2	2	1	1	2	3	3	2	2	1	
131	Crispin	KASP	3	0	2	2	2	1	1	3	4	1	3	0	2	2	2	1	1	2	3	3	2	2	1	
132	Crop34	SeqSNP	1	2	2	3	0	2	2	2	2	2	1	2	2	2	3	1	2	2	3	1	2	3	2	
132	Crop34	KASP	1	2	2	3	0	2	2	2	2	2	1	2	2	2	3	1	2	2	3	1	2	3	2	
133	Crop60 2484	SeqSNP	2	2	3	3	3	0	3	1	2	2	1	2	2	1	2	1	3	2	0	1	2	1	1	
133	Crop60 2484	KASP	2	2	3	3	3	0	3	1	2	2	1	2	2	1	2	1	3	2	0	1	2	1	1	
134	Cwater Russet	SeqSNP	1	1	1	2	3	1	3	3	3	1	1	2	2	3	3	1	4	1	1	2	1	4	2	
134	Cwater Russet	KASP	1	1	1	2	3	1	2	3	3	1	1	2	2	3	3	1	4	1	1	2	1	4	2	
135	Daisy	SeqSNP	2	3	1	1	2	3	3	1	0	3	0	3	2	2	0	3	3	1	2	2	2	3	2	
135	Daisy	KASP	2	NA	1	1	2	3	3	NA	0	3	0	3	2	2	0	3	3	1	2	2	2	3	2	
55	Dakchip	SeqSNP	2	3	1	2	1	2	3	1	3	2	1	0	4	3	3	1	2	1	0	0	2	2	4	
136	DakotaTrailblazr	SeqSNP	2	0	1	2	2	2	3	2	3	1	2	0	1	2	2	2	1	2	3	0	2	1	2	
136	DakotaTrailblazr	KASP	2	0	1	NA	2	2	2	2	3	1	2	0	1	2	2	2	1	2	3	0	2	1	2	
137	Darius	SeqSNP	3	2	1	3	0	0	1	3	3	1	0	2	2	3	2	2	4	2	2	2	3	2	4	
137	Darius	KASP	2	2	1	3	0	0	1	3	3	1	0	2	2	3	2	2	4	2	2	2	3	2	4	
91	Desirée	SeqSNP	0	3	2	2	2	3	1	3	1	2	1	3	3	2	2	3	2	2	1	3	1	2	2	
66	Devlin	SeqSNP	2	3	2	3	1	1	1	3	2	2	0	2	2	1	0	2	3	2	2	2	0	2	3	
66	Devlin	KASP	2	3	2	3	1	1	1	3	2	2	0	2	2	1	0	2	3	2	2	2	0	2	3	
105	Diamant	SeqSNP	2	3	1	4	3	1	3	1	3	1	3	3	3	2	3	2	4	2	3	3	0	2	3	
20	Diana	SeqSNP	0	2	1	1	2	3	1	1	2	2	2	4	2	2	3	2	3	2	1	3	1	3	3	
138	Donata	SeqSNP	1	2	2	2	2	1	2	3	2	4	3	2	2	2	1	1	1	2	0	1	4	2	2	
138	Donata	KASP	1	2	2	2	2	1	1	3	2	4	4	2	2	2	1	1	1	2	0	1	4	2	2	
53	Draga (Rascals)	SeqSNP	2	0	3	4	3	0	0	2	2	3	1	2	0	1	3	1	3	2	2	2	4	3	1	
53	Draga (Rascals)	KASP	NA	0	NA	4	NA	NA	0	2	NA	NA	1	NA	0	NA	NA	NA	NA	NA	NA	4	NA	1		
116	DTO33	SeqSNP	2	1	1	2	3	0	2	3	3	4	4	4	3	3	3	3	3	3	1	1	1	3	2	
116	DTO33	KASP	2	1	1	2	3	0	1	3	3	4	4	4	3	3	3	3	3	3	1	1	1	3	2	
139	Earliest of All	SeqSNP	1	3	1	3	3	1	1	2	4	2	0	2	3	2	0	2	1	2	1	1	1	2	3	
28	Eldena	SeqSNP	0	3	1	3	2	1	2	1	2	1	3	0	3	2	2	2	3	2	1	2	4	2	2	
140	Elodie	SeqSNP	4	1	2	2	2	2	4	2	3	3	2	2	1	1	2	3	3	1	0	3	2	2	2	
140	Elodie	KASP	4	1	2	2	2	2	4	2	3	3	3	2	1	1	2	3	3	1	0	3	2	2	2	
44	Elsa	SeqSNP	0	4	2	1	2	1	0	3	2	3	4	2	1	3	1	1	2	2	1	3	1	2	3	
44	Elsa	KASP	0	4	2	1	2	1	0	3	2	3	4	2	1	3	1	1	3	2	1	3	1	2	3	
22	Ernstoltz	SeqSNP	1	2	2	2	2	0	1	4	1	2	3	1	4	3	2	3	1	4	1	3	2	2	2	
31	Eryn	SeqSNP	2	3	1	4	0	1	3	3	1	1	2	1	3	3	1	2	4	3	2	2	2	3	2	
31	Eryn	KASP	2	NA	1	4	0	1	2	3	1	1	3	1	3	3	1	2	4	3	2	2	2	3	2	
141	Esco	SeqSNP	2	3	1	1	2	3	2	3	2	2	3	0	2	1	1	1	2	2	2	3	2	1	2	
141	Esco	KASP	2	3	1	1	2	3	1	3	2	2	3	0	2	1	1	1	2	2	2	3	2	1	2	
42	Escort	SeqSNP	2	3	1	1	2	3	2	3	2	2	3	0	2	1	1	1	2	2	2	3	2	1	2	
111	Esparante	SeqSNP	2	1	3	3	2	1	2	1	3	2	2	2	1	3	2	2	2	3	2	4	2	4	3	
142	Estima	SeqSNP	4	2	3	2	3	2	2	1	2	3	2	2	2	2	2	2	2	1	1	1	2	3	4	
143	Evan	SeqSNP	2	3	2	2	1	1	2	1	0	1	2	1	1	2	3	2	4	3	2	3	2	0	2	
143	Evan	KASP	2	3	2	2	1	1	1	1	0	1	2	1	1	2	3	2	4	3	2	3	2	0	2	
85	Fabian	SeqSNP	3	2	0	2	3	1	1	2	3	2	2	2	4	2	2	3	3	1	0	0	1	3	2	
85	Fabian	KASP	3	2	0	2	3	1	1	2	3	2	2	2	4	2	2	2	3	1	0	0	1	3	2	
25	Fambo	SeqSNP	2	2	3	3	2	1	1	2	0	1	1	2	3	2	1	2	1	2	1	2	2	2	4	4
46	Falima	SeqSNP	2	2	1	2	3	3	1	2	2	3	1	1	3	3	2	0	1	2	1	2	3	4	4	
43	Felsina	SeqSNP	0	2	4	3	1	1	2	2	4	2	1	3	2	1	1	2	1	1	3	4	3	2	1	
43	Felsina	KASP	0	NA	4	NA	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	NA	3	4	NA	NA	1	
144	Fianna	SeqSNP	1	2	2	3	1	1	1	2	2	0	2	1	2	2	2	1	2	2	2	3	3	4	2	
144	Fianna	KASP	1	2	2	3	1	0	1	2	2	0	2	1	2	2	2	1	2	2	2	3	3	4	2	
145	Figaro	SeqSNP	1	3	2	1	1	1	3	1	4	2	3	2	1	2	1	2	1	1	2	2	3	3	2	
193	FL2006	SeqSNP	2	2	1	1	0	4	1	3	3	2	2	1	1	2	1	0	2	1	2	3	3	1	1	
193	FL2006	KASP	2	2	1	1	0	4	1	3	3	2	2	1	1	2	1	0	2	1	2	3	3	1	1	
194	FL2108	SeqSNP	2	2	2	2	2	3	0	0	2	2	1	4	2	0	2	1	3	1	1	2	3	3	2	
194	FL2108	KASP	2	2	2	2	2	3	0	0	2	2	2	4	2	0	2	1	3	1	1	2	3	3	2	
195	FL2476	SeqSNP	3	0	2	2	3	4	1	2	1	2	2	3	4	1	2	3	4	3	0	2	4	4	3	
195	FL2476	KASP	3	0	2	2	3	4	1	2	1	2	3	3	4	1	2	3	4	3	0	2	4	4	3	
146	Folva	SeqSNP	1	2	3	1	1	2	2	0	3	2	1	1	0	0	0	3	2	1	2	2	3	2	3	
21	Frisia	SeqSNP	2	2	1	2	3	3	1	2	2	3	1	1	3	3	2	0	1	2	1	2	3	4	4	
147	Frodo	SeqSNP	1	1	2	3	2	0	3	2	3	3	2	0	2	3	2	3	4	1	2	0	3	2	2	
147	Frodo	KASP	1	1	2	3	2	0	2	2	3	3	2	0	2	3	2	3	4	1	2	0	3	2	2	
148	Gatsby	SeqSNP	2	1	1	1	2	2	3	2	2	2	2	2	1	0	1	2	1	2	2	3	2	3	4	
39	Gemchip	SeqSNP	1	0	1	1	3	1	3	2	4	2	2	2	3	2	3	2	3	1	3	2	3	2	2	
149	Georgina	SeqSNP	3	2	3	2	3	2	3	0	4	3	3	3	2	3	3	3	1	1	1	2	2	2	2	
72	Grandifolia	SeqSNP	1	2	2	3	3	3	1	1	3	1	3	1	2	2	1	0	3	1	2	2	4	1	3	

KASP SNP assay

gDNA#	Cultivar name	Genotype	A	C	D	E	F	G	H	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	
150	Hermes	SeqSNP	3	2	4	2	2	2	0	2	3	1	2	2	4	1	3	2	3	3	1	1	2	3	1	
150	Hermes	KASP	3	2	4	2	2	2	0	2	3	1	4	2	4	1	3	2	3	3	1	1	2	3	1	
82	Hertha (Rascals)	SeqSNP	1	1	2	2	2	2	3	2	3	4	3	3	3	2	1	2	2	2	1	2	4	2	4	
82	Hertha (Rascals)	KASP	1	1	2	2	2	2	2	2	3	4	3	3	3	2	1	2	2	2	1	2	4	2	4	
69	Hoëvelder	SeqSNP	1	2	1	1	1	2	0	2	1	1	1	2	2	3	1	2	4	2	1	1	1	2	3	
23	Hudson	SeqSNP	3	1	2	1	3	1	2	2	2	0	0	1	3	2	3	2	3	3	0	3	1	3	3	
151	Hydra	SeqSNP	2	2	1	0	4	2	3	3	3	2	1	2	2	0	1	2	2	0	2	2	2	1	1	
10	IIZA49A1	SeqSNP	2	2	3	1	4	1	2	2	2	1	1	2	1	2	3	1	2	2	1	2	2	1	3	
10	IIZA49A1	KASP	2	2	3	1	4	1	2	2	2	1	1	2	1	2	3	1	2	2	1	2	2	1	3	
9	IIZASSA5	SeqSNP	2	3	2	0	3	2	3	2	4	1	3	1	2	1	2	1	2	2	1	2	1	3	3	
9	IIZASSA5	KASP	2	3	2	0	3	2	3	2	4	1	3	1	2	1	2	1	2	2	1	2	1	3	3	
152	Innovator	SeqSNP	1	4	1	1	1	2	3	0	2	3	4	2	0	1	3	3	2	2	1	2	1	1	3	
152	Innovator	KASP	1	4	1	1	1	2	2	0	2	3	4	2	1	1	3	3	2	2	1	2	1	1	3	
153	Irish Gold	SeqSNP	3	2	4	3	4	3	3	2	1	2	3	2	3	3	3	1	1	2	2	2	2	2	3	
154	Isle of Jura	SeqSNP	2	2	3	1	2	2	3	2	3	2	3	2	1	0	1	2	0	0	2	0	3	0	2	
79	Jelly	SeqSNP	4	2	1	0	1	3	2	1	3	2	2	2	0	2	1	0	2	2	3	3	1	1	2	
79	Jelly	KASP	4	2	1	0	1	3	2	1	3	2	3	2	0	2	1	0	2	2	3	3	1	1	2	
115	Jemseg	SeqSNP	1	2	2	3	4	2	1	3	3	2	3	1	3	3	2	2	0	2	0	1	3	0	3	
197	Kankan	SeqSNP	2	2	0	3	3	0	2	3	2	3	0	3	3	2	3	2	3	0	1	1	1	3	1	
45	Katahdin	SeqSNP	2	2	1	1	3	1	3	3	3	2	2	2	4	1	2	0	2	1	1	0	2	2	2	
59	Kimb. Choice	SeqSNP	3	4	2	2	2	0	2	3	1	1	1	3	3	3	1	3	4	2	2	1	2	3	3	
110	King George	SeqSNP	2	4	2	2	1	1	3	1	4	2	3	1	4	2	3	3	2	2	0	1	3	3	2	
110	King George	KASP	NA	4	NA	NA	1	NA	NA	1	4	NA	4	1	4	NA	NA	3	NA	NA	NA	NA	NA	NA	NA	
4	King Russet	SeqSNP	4	4	1	1	0	1	1	0	1	2	4	0	0	2	2	3	2	2	2	3	3	2	1	
4	King Russet	KASP	4	4	1	1	0	1	1	0	1	2	4	0	0	2	2	3	2	2	2	3	3	2	1	
155	Kingsman	SeqSNP	1	0	3	3	3	4	2	4	3	2	2	1	2	1	2	0	1	2	2	2	2	3	2	
155	Kingsman	KASP	1	0	3	3	3	4	2	4	3	2	2	1	2	1	2	0	1	2	2	2	2	3	2	
68	Kingston	SeqSNP	1	2	1	2	1	2	1	1	3	1	2	2	2	0	1	2	2	1	0	1	3	1	1	
70	Koos Smit	SeqSNP	2	2	3	2	2	1	4	2	4	1	2	3	3	2	2	2	3	3	4	3	1	3	0	
70	Koos Smit	KASP	2	3	3	2	3	1	3	0	3	2	1	3	2	2	3	2	2	2	2	2	2	3	3	
156	La Strada	SeqSNP	3	0	3	2	4	1	0	1	3	1	1	1	1	0	3	2	0	2	2	2	2	2	3	
81	Lady Rosetta	SeqSNP	2	2	3	2	2	1	4	2	4	1	2	3	3	2	2	2	3	3	4	3	1	3	0	
81	Lady Rosetta	KASP	2	2	3	2	2	1	4	2	4	1	2	3	3	2	2	2	3	3	4	3	1	3	0	
11	Lanorma	SeqSNP	1	1	3	1	2	2	3	1	2	1	2	2	2	4	3	1	2	3	3	2	1	4	3	
11	Lanorma	KASP	1	1	3	1	2	2	3	1	2	1	2	2	2	4	3	1	2	3	3	2	1	4	3	
73	Late Harvest	SeqSNP	2	2	2	2	3	2	2	1	3	4	2	2	2	1	2	3	1	1	2	2	1	3	2	
74	Lenape	SeqSNP	2	2	2	2	3	2	2	1	3	4	2	2	2	1	2	3	1	1	2	2	1	3	2	
78	Liberator	SeqSNP	3	2	2	2	0	2	2	1	3	1	2	1	0	4	2	2	2	1	2	3	3	2	1	
112	Liseta (Rascals)	SeqSNP	3	2	2	2	4	2	4	2	0	2	1	1	3	2	2	1	2	1	1	3	2	3	2	
112	Liseta (Rascals)	KASP	3	2	1	2	4	2	4	2	0	2	1	1	3	2	2	1	2	1	1	3	2	3	2	
26	LT 7	SeqSNP	3	1	1	1	1	3	3	3	1	3	0	3	1	1	3	3	2	2	3	0	2	3	3	
157	Ludmilla	SeqSNP	2	2	2	1	1	1	1	0	2	2	4	1	0	1	2	2	1	1	0	3	2	2	2	
158	Magnum	SeqSNP	2	1	1	3	2	3	3	1	3	3	2	2	0	2	3	3	2	4	1	2	2	4	4	
158	Magnum	KASP	2	NA	1	3	2	3	2	1	3	3	4	2	0	2	3	3	2	4	1	2	2	4	4	
159	Manhattan	SeqSNP	2	1	2	2	3	3	2	1	1	3	1	2	1	1	4	2	1	2	1	0	3	4	4	
159	Manhattan	KASP	NA	1	NA	NA	NA	3	2	NA	1	NA	2	NA	1	1	4	NA	1	NA	1	0	3	4	4	
57	Maradonna	SeqSNP	3	0	1	3	3	1	3	1	3	2	4	2	2	1	4	3	3	2	3	3	3	2	2	
109	Marfona	SeqSNP	2	3	2	3	2	2	3	4	1	3	1	1	3	3	2	3	2	1	3	1	3	1	2	3
118	Marijke	SeqSNP	1	3	2	0	2	1	3	1	2	2	1	1	3	2	2	2	3	1	2	4	1	4	2	
118	Marijke	KASP	1	3	2	0	2	1	2	1	2	2	2	1	3	2	2	2	3	1	2	4	1	4	2	
75	Maris Piper	SeqSNP	4	3	3	3	2	0	3	1	3	3	2	3	4	2	2	2	3	3	1	1	2	1	2	
75	Maris Piper	KASP	4	3	3	3	2	0	2	1	3	3	3	3	4	2	2	2	NA	3	1	1	2	1	2	
107	Marispeer	SeqSNP	4	3	3	3	2	0	3	1	3	3	2	3	4	2	2	2	3	3	1	1	2	1	2	
107	Marispeer	KASP	4	3	3	3	2	0	2	1	3	3	2	3	4	2	2	2	3	3	1	1	2	1	2	
160	Markies	SeqSNP	1	3	2	2	2	3	1	3	3	2	2	0	2	2	1	3	1	2	2	3	3	4	1	
160	Markies	KASP	1	3	1	2	2	3	1	3	3	2	3	0	2	2	1	3	1	2	2	3	3	4	1	
114	Meliose	SeqSNP	2	2	3	3	2	2	2	2	1	3	2	2	4	2	4	2	2	1	2	3	3	3	4	
108	Mirakel	SeqSNP	2	1	0	2	2	1	2	2	4	2	2	3	3	2	3	2	2	2	3	4	0	2	3	
117	Mnandi	SeqSNP	1	3	2	3	2	1	3	0	3	2	2	2	1	2	0	1	2	2	3	0	2	3	3	
117	Mnandi	KASP	1	3	2	3	2	1	2	1	3	2	2	2	1	2	0	1	3	2	3	0	2	3	3	
38	Mokgotlong	SeqSNP	2	2	1	0	4	2	3	3	3	2	1	2	2	0	1	2	2	0	2	2	2	1	2	
24	Monalisa	SeqSNP	1	2	2	3	3	3	1	2	2	2	4	2	3	2	4	1	0	0	2	2	0	4	2	
24	Monalisa	KASP	NA	NA	NA	NA	NA	3	NA	NA	NA	NA	4	NA	NA	NA	4	NA	0	0	NA	NA	0	4	NA	
76	Mondial	SeqSNP	2	2	2	1	3	1	4	1	0	3	3	1	2	3	2	1	1	2	2	3	1	3	3	
76	Mondial	KASP	2	2	2	2	3	1	4	1	0	3	3	1	2	3	2	1	1	2	2	3	1	3	3	
27	Mondial (Rascals)	SeqSNP	2	2	2	1	3	1	4	1	0	3	3	1	2	3	2	1	1	2	2	3	1	3	3	
27	Mondial (Rascals)	KASP	2	2	2	1	3	1	4	1	0	3	3	1	2	3	2	1	1	2	2	3	1	3	3	
27_d	Mondial (Rascals)_d	SeqSNP	2	2	2	1	3	1	4	1	0	3	3	1	2	3	2	1	1	2	2	3	1	3	3	
27_d	Mondial (Rascals)_d	KASP	2	2	2	1	3	1	4	1	0	3	3	1	2	3	2	1	1	2	2	3	1	3	3	
76_d	Mondial_d	SeqSNP	2	2	2	1	3	1	4	1	0	3	3	1	2	3	2	1	1	2	2	3	1	3	3	
76_d	Mondial_d	KASP	2	2	2	1	3	1	4	1	0	3	3	1	2	3	2	1	1	2	2	3	1	3	3	
161	Monica russet	SeqSNP	1	4	1	1	1	2	3	0	2	3	4	2	1	1	3	3	2	2	1	2	1	1	3	
161	Monica russet	KASP	1	4	1	1	1	2	2	0	2	3	4	2	1	1	3	3	2	2	1	2	1	1	3	

