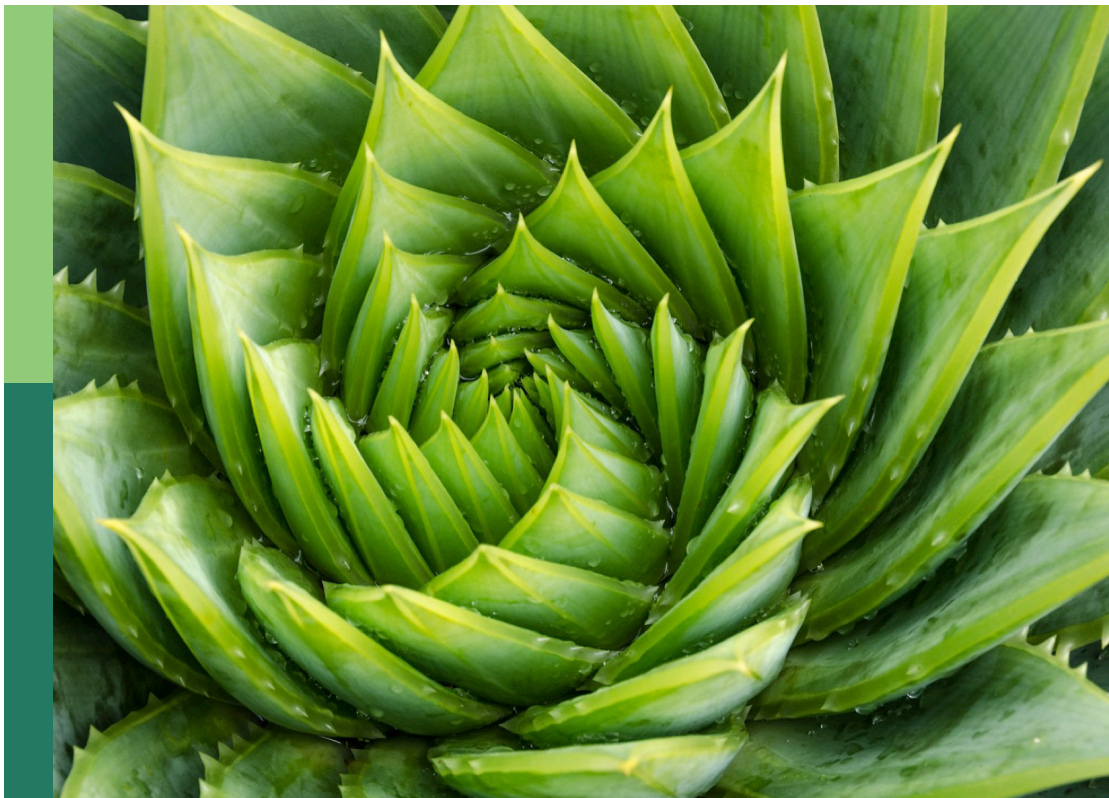
frontiers | Research Topics

Advances in conservation and use of potato genetic resources

Edited by
Iris Edith Peralta, Alfonso H. Del Rio, Gustavo Heiden, Dave Ellis and Peter Giovannini

Published in
Frontiers in Plant Science



FRONTIERS EBOOK COPYRIGHT STATEMENT

The copyright in the text of individual articles in this ebook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this ebook is the property of Frontiers.

Each article within this ebook, and the ebook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this ebook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or ebook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-8325-5711-2
DOI 10.3389/978-2-8325-5711-2

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Advances in conservation and use of potato genetic resources

Topic editors

Iris Edith Peralta — National University of Cuyo, Argentina
Alfonso H. Del Rio — University of Wisconsin-Madison, United States
Gustavo Heiden — Brazilian Agricultural Research Corporation (EMBRAPA), Brazil
Dave Ellis — International Potato Center, Peru
Peter Giovannini — Global Crop Diversity Trust, Germany

Citation

Peralta, I. E., Del Rio, A. H., Heiden, G., Ellis, D., Giovannini, P., eds. (2024). *Advances in conservation and use of potato genetic resources*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-5711-2

Table of contents

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 23 September 2022
DOI 10.3389/fpls.2022.954162

Check for updates

OPEN ACCESS

EDITED BY
Dave Ellis,
International Potato Center, Peru

REVIEWED BY
Noelle Anglin,
International Potato Center, Peru
Juliano Lino Ferreira,
Embrapa Pecuária Sul, Brazil

*CORRESPONDENCE
Liping Jin
jinliping@caas.cn

SPECIALTY SECTION
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

RECEIVED 27 May 2022
ACCEPTED 02 September 2022
PUBLISHED 23 September 2022

CITATION
Hu J, Mei M, Jin F, Xu J, Duan S, Bian C, Li G, Wang X and Jin L (2022) Phenotypic variability and genetic diversity analysis of cultivated potatoes in China.
Front. Plant Sci. 13:954162.
doi: 10.3389/fpls.2022.954162

COPYRIGHT
© 2022 Hu, Mei, Jin, Xu, Duan, Bian, Li, Wang and Jin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Phenotypic variability and genetic diversity analysis of cultivated potatoes in China

Jun Hu[1], Meng Mei[2], Fang Jin[3], Jianfei Xu[1], Shaoguang Duan[1], Chunsong Bian[1], Guangcun Li[1], Xiyao Wang[2] and Liping Jin[1]*

[1]Key Laboratory of Biology and Genetic Improvement of Tuber and Root Crops, Ministry of Agriculture and Rural Affairs of the People's Republic of China, Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing, China, [2]College of Agronomy, Sichuan Agricultural University, Chengdu, China, [3]National Agro-Tech Extension and Service Center, Ministry of Agriculture and Rural Affairs of the People's Republic of China, Beijing, China

Phenotypic evaluation and molecular biotechnology are both important in the identification and utilization of crop germplasm resources. In this study, the phenotypic variation and genetic diversity of 149 main potato cultivars in China were investigated with 12 phenotypic traits and 24 SSR markers. The coefficient of variation of 12 phenotypic traits ranged from 12.11% to 156.93%. The results of SSR markers exhibited a relatively high level of genetic variation (*Na* =5.458 ± 1.499, *Ne* =3.300 ± 1.087, *I* =1.397 ± 0.298, *Ho* =0.797 ± 0.178, *He* = 0.660 ± 0.117, and *PIC*=0.702 ± 0.087). Population structure and phylogenetic tree analysis divided the varieties into three subgroups. The results indicated that ninety percent of the molecular variance was attributed to within-group differences, and the remaining 10% was attributed to variation among groups. Consistent with previous report, alleles of the STI032 marker were significantly associated with tuber starch content and growth period traits in the population. The results of this study could facilitate the utilization of potato germplasm resources, molecular genetic breeding and improvement.

KEYWORDS

potato, SSR marker, genetic diversity, population structure, variety identification

Introduction

China produces 25.5% of the world's potato (FAOSTAT, 2020). With a total acreage of 4.2 million hectares and production of 78.2 million tons in 2020 (FAOSTAT, 2020), potatoes are widely distributed throughout the country in four agro-ecological zones, including the Northern single-cropping zone, the Southwestern mixed-cropping zone, the Central Plain double-cropping zone and the Southern winter-cropping zone (Jansky et al., 2009). There are different potato breeding objectives for key traits of varieties in different zones (Jansky et al., 2009). For example, early maturity and yellow tuber flesh

are favored in the Central Plain double-cropping zone and the Southern winter-cropping zone, whereas late maturity and high tuber starch contents are common in the Northern single-cropping zone. In addition, high resistance to late blight is necessary in the Southwestern mixed-cropping zone (Jansky et al., 2009). Since the 1960s, more than 800 new potato varieties have been released from breeding programs in China, most of which are suitable for table consumption (Xu and Jin, 2017). It is still difficult to meet the demands for the development of the potato processing industry in China.

The effective exploitation of plant genetic resources is built on the accurate identification of phenotypic and genotypic variation from target germplasms. Determination of genetic diversity is helpful for the effective use of germplasm, especially in plant breeding research. There have been some studies on potato germplasm and variety at both the morphological and molecular levels (Wang et al., 2017; Duan et al., 2019; Wang et al., 2019; Duan et al., 2021). A total of 288 potato genotypes from around the world were classified into three to eight groups with 20 SSR and 10 AFLP markers (Wang et al., 2017). A population of 292 genotypes from the International Potato Center (CIP), Europe and Yunnan Province in China was clustered into two main groups and subdivided into seven groups using 30 SSR markers (Wang et al., 2019). However, the systematic phenotypic variability and genetic diversity and population structure of potato varieties bred from various regions of China need more study.

Pedigree analysis is an important method for selecting parents and confirming the genetic relationships of their offspring (Li et al., 2018a). However, due to the lack of effective pedigree information, it is difficult to determine genetic relationships by the pedigree method in many varieties (Duan et al., 2019). Molecular genetic diversity studies can evaluate all levels of genetic structure from the relationship between complex components of species to the origin of specific genotypes. Different marker systems, such as SSRs, AFLPs and SNPs (single nucleotide polymorphisms), have been employed to study different crops in addition to potato. However, traditional SSR marker polymorphisms are analyzed by polyacrylamide gel electrophoresis (PAGE) and manual reading, which is time-consuming, labor-intensive and nonautomated. Moreover, data collection and analysis of multiple batch samples are associated with numerous difficulties. Capillary electrophoresis with fluorescent-labeled SSR primers has been widely used (Chandra et al., 2014; Liang et al., 2018; Tiwari et al., 2018; Verma et al., 2019; Ayenan et al., 2021) owing to its high efficiency and automation.

Plant maturity is an important agronomic trait in potato. Maturity is related to starch content and late blight resistance, which is an important factor in determining the regional adaptability of varieties (Domański, 2001; van Eck, 2007). Studies focused on quantitative trait loci (QTLs) for plant maturity both in diploid and tetraploid potato have been reported (Bradshaw et al., 2008; McCord et al., 2011; Kloosterman et al., 2013; D'Hoop et al., 2014; Massa et al., 2015; Li et al., 2018b). A major QTL related to potato plant maturity was located on chromosome V (Kloosterman et al., 2013; Li et al., 2018b). Two overlapping QTLs for potato maturity and tuber dry matter content were identified on chromosome V in a tetraploid population (Bradshaw et al., 2008). Fourteen loci for tuber starch content were identified on seven chromosomes by association analysis in 243 tetraploid potato germplasms (Li et al., 2008). Although an abundance of excellent genetic resources exist in diploid wild germplasm (Duan et al., 2021), it is difficult to deploy these resources in tetraploid cultivated varieties by introgression due to reproductive barriers and/or unfavorable linkage drag.

In this study, we examined the phenotypic variability and genetic diversity of 149 main potato cultivars to uncover their genetic diversity and relationships among populations for the utilization and identification of potato germplasm and breeding programs. The information in the study provides data support for variety identification and selection of cross-combinations among cultivars to improve potato breeding in China.

Materials and methods

Plant materials and phenotypic traits

A brief description of the 149 main potato varieties grown in China since the 1970s is summarized in Supplementary Table S1. Of the 149 cultivars used in the study, 144 were developed by breeding programs in different regions of China, and five (Atlantic, Favorita, Shepody, Mira and Desiree) were imported from abroad and commonly planted in China. The phenotypic data of the varieties were collected from the information provided by the national variety approval and registration in China through a big data platform (http://202.127.42.47:6010/index.aspx) and the National Potato Variety Resources Catalogue (Hongzhu Yang, 1989). Growth period (GP), plant height (PH), tuber dry matter content (DM), starch content (SC), reducing sugar content (RS), protein content (PC) and vitamin C content (VC) data were recorded as numerical values in the database. Tuber shape (TS), eye depth (ED), tuber skin color (TSC), tuber flesh color (TFC) and flower color (FC) were determined according to Domański L. (2001) with minor modifications. TS was scored as follows: 1: compressed, 2: round, 3: round-oval, 4: oval, 5: long-oval, and 6: long. ED was evaluated on a scale ranging from 1 to 9, where 1: greater deep bud eye, 3: deep bud eye, 5: moderate bud eye, 7: shallow bud eye and 9: greater shallow bud eye. TSC and TFC were

evaluated according to a 0- to 5-point scale, where 0: white, 1: pale yellow, 2: yellow, 3: deep yellow, 4: red, and 5: purple. FC was scored as follows: 0: white, 1: pale violet, 2: violet, 3: dark violet, and 4: blue corolla. Analysis of word cloud was finished by *wordcloud* package in R software to determine the frequency of a potato cultivar used as a parent. Network analysis was performed using the *igraph* package in R software.

DNA extraction and SSR analysis

Approximately 2 g of fresh young leaves was collected. Genomic DNA was extracted according to the modified cetyltrimethylammonium bromide (CTAB) procedure. All DNA samples were diluted to 25 ng μL^{-1} and stored at -20°C until use. Twenty-four SSR primer pairs covering all 12 chromosomes were labeled with FAM, HEX, ROX and TAMRA fluorescent dyes reseparately. Primers were synthesized by Qingke Biotech Company (Beijing, China). The sequences of these primer pairs were obtained from previous reports (Supplementary Table S2). PCR amplifications were performed in a 20 μL reaction mixture that consisted of 4.0 μL of DNA template, 0.5 μL of forward primer (10 pmol μL^{-1}), 0.5 μL of reverse primer (10 pmol μL^{-1}), 10 μL of 2 × Taq Master Mix (Vazyme, Nanjing, China), and 5 μL of ddH_2O. A touchdown PCR was used: 5 min at 95°C, 13 cycles of 30 s for denaturation at 95°C, 45 s for gradient annealing from 60°C to 50°C (each cycle reduced by 0.8°C), 30 s for extension at 72°C, followed by 24 cycles of 95°C for 30 s, 50°C for 45 s, and 72°C for 30 s, with a final step at 72°C for 5 min. The amplified SSR products were fragment analyzed on a '3730 Genetic Analyzer' (Applied Biosystems, Foster City, California, USA). The results of the peak patterns produced were analyzed by an SSR Analyzer (Wang et al., 2018), which is based on the commercialized software GeneMarker. A 500-bp LIZ500 standard was used to estimate the molecular size of the amplification fragments. For each locus, peaks were recorded in order from smallest to largest (Supplementary Table S3). The number of peaks and the number of profiles per marker were scored based on amplification of the cultivars. A data matrix of 149 cultivars was constructed based on the presence (1) or absence (0) of the amplified SSR fragments.

Genetic diversity, population structure and phylogeny analysis

The number of alleles (*Na*), the effective number of alleles (*Ne*), Shannon's information index (*I*), observed heterozygosity (*Ho*), expected heterozygosity (*He*), and *F*-Statistics (*Fis*) were determined using the GenAlEx 6.502 program (Peakall and Smouse, 2012), whereas the polymorphism information content (*PIC*) value was estimated using PowerMarkerv3.25 software (Liu and Muse, 2005). Population structure was inferred by Bayesian clustering implemented in STRUCTURE v.2.3.4 (Pritchard et al., 2000). K values from 1 to 20 were tested for twenty independent runs with a burn-in length of 100,000 and MCMC repetitions of 100,000. The most likely K value was analyzed by the method in Evanno et al. (2005) using STRUCTURE Harvester (Earl and Vonholdt, 2012). Furthermore, analysis of molecular variance (AMOVA) was conducted to estimate variation among the three groups using GeneAlEx software with 1000 permutations. A cluster analysis based on the neighbor-joining method was also conducted using PowerMarker software, and then an unrooted tree was constructed and analyzed using MEGA X software with default setting parameters (Kumar et al., 2018).

Marker–trait associations

Association analysis was carried out for the identification of significant marker–trait association (MTA). The phenotype trait data along with SSR markers were analyzed in the TASSEL 5.2.77 program (Bradbury et al., 2007) to identify significant marker–trait associations (MTAs). The analysis of MTAs was performed using a mixed linear model (MLM) with TASSEL software based on the Q-matrix and kinship matrix (K-matrix). The relative kinship matrix (K) was determined using SPAGeDi software v1.2 (Hardy and Vekemans, 2002). The significance of MTAs was tested in terms of *P* value ($P < 0.01$ for significant markers). False discovery rate (FDR) was analyzed using the Q-values in R 3.5.3 (R Core Team, 2019).

Results

Analysis of phenotypic traits and pedigree information

In this study, a total of 12 traits, such as growth period (GP), plant height (PH), flower color (FC), tuber shape (TS), tuber skin color (TSC), tuber flesh color (TFC), eye depth (ED), starch content (SC), dry matter content (DM), protein content (PC), reducing sugar content (RS) and vitamin C content (VC), in the population are listed in Table 1. The coefficient of variance for the traits ranged from 12.11% in dry matter content to 156.93% in flower color (Table 1). The analysis of the traits revealed a broad spectrum of variation in the population. Correlation analysis showed 10 extremely significant ($P < 0.01$) differences in correlation between traits (Figure 1, Supplementary Table S4). Growth period traits showed an extremely significant correlation with the tuber dry matter content and starch content ($P < 0.001$) (Figure 1). Principal component analysis of the phenotypic data of the 12 traits showed that six principal components potentially contributed to the variation of the traits, with a cumulative

TABLE 1 12 phenotypic variation parameters of 149 potato cultivars in China.

Traits	Min.	Max.	Average	SD	CV
Growth Period (d)	50.00	144.00	97.38	19.85	20.39
Plant Height (cm)	32.00	113.00	62.80	13.44	21.40
Starch Content (%)	10.30	22.86	15.49	2.49	16.06
Dry matter Content (%)	14.80	28.20	21.76	2.64	12.11
Reducing Sugar Content (%)	0.02	1.54	0.35	0.25	72.57
Vitamin C Content (mg.100 g^{-1})	8.93	48.00	17.50	5.72	32.68
Protein Content (%)	0.53	3.52	2.16	0.44	20.59
Flower Color	0.00	4.00	0.62	0.98	156.93
Tuber Shape	1.00	6.00	2.76	1.46	53.07
Eye Depth	2.00	9.00	6.36	1.26	19.80
Tuber Skin Color	0.00	5.00	1.60	1.12	69.87
Tuber Flesh Color	0.00	5.00	0.92	0.88	96.40

SD, standard Deviation; CV, Coefficient of Variance.

variance of 71.3%. The eigenvalue of the first principal component was 1.504, accounting for 18.97% of the total variation, and the eigenvalue of the second component was 1.333, accounting for 14.92% of the total variation in the population (Supplementary Table S5). Three main groups were detected on the cluster dendrogram in accordance with phenotypic trait values. From up (above) to down (below), the first cluster contained 7 cultivars, the second 76 cultivars, and the third 66 cultivars (Supplementary Figure S1).

Seven potato cultivars were frequently used as parents by word cloud analysis (Figure 2A). Among them, cv. Schwalbe was used the most frequently used (n = 7 times), followed by the cultivars Zhongshu 3 (n = 6), Epoka (n = 6), C93.154 (n = 5), Katahdin (n = 5), Duozibai (n = 4), and Shepody (n = 4) (Figure 2B).

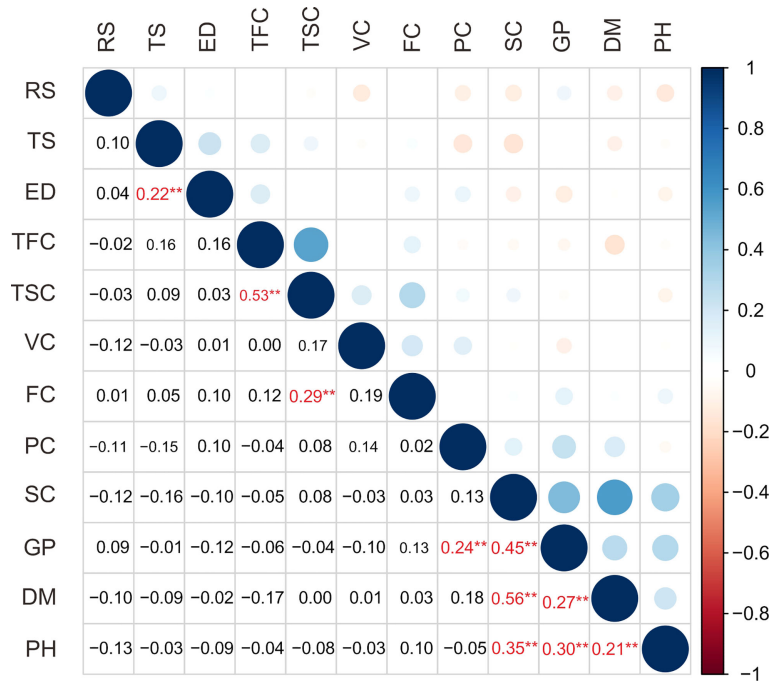


FIGURE 1
Correlation analysis of 12 traits of 149 potato cultivars. RS, reducing sugar content; TS, tuber shape; ED, eye depth; TFC, tuber flesh color; TSC, tuber skin color; VC, vitamin C content; FC, flower color; PC, protein content; SC, starch content; GP, growth period; DM, dry matter content; PH, plant height. The numbers shown in red reach extremely significant correlation and are indicated with ** (P<0.01).

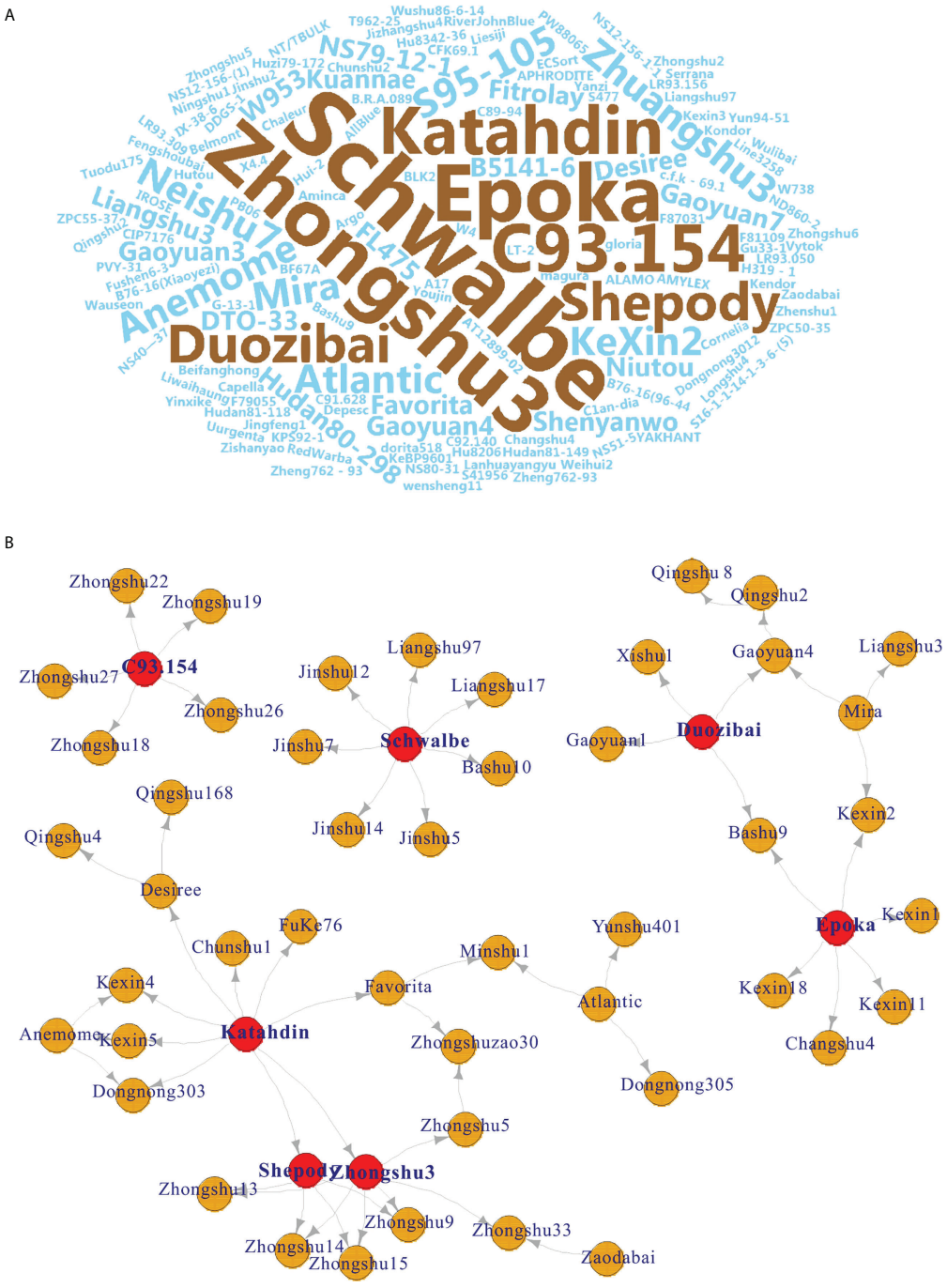


FIGURE 2
Pedigree-based word cloud and network analysis. **(A)** Pedigree-based word cloud analysis of 149 potato varieties; the most frequently used parents are shown in brown. **(B)** Network of seven high-frequency potato varieties (red node) based on the results of the word cloud.

SSR polymorphism

A total of 148 SSR markers, including 77 markers from previous studies (Kishine et al., 2017; Duan et al., 2019; Wang et al., 2019; Duan et al., 2021) and 71 new trinucleotide and tetranucleotide core motif SSR markers, were used for primer screening with sixteen cultivars based on polyacrylamide gel electrophoresis (data not shown). Finally, twenty-four SSR markers from all 12 chromosomes (Supplementary Table S2) were screened out to evaluate the genetic diversity of 149 potato

cultivars. A total of 131 alleles were recorded. The number of alleles per locus (*Na*) varied from 2 to 9, with a mean of 5.46 ± 1.50. The effective number of alleles per locus (*Ne*) ranged from 1.64 to 5.29, with a mean of 3.30 ± 1.09. The value of observed heterozygosity (*Ho*) changed from 0.46 to 0.994 with an average value of 0.797, while the value of expected heterozygosity (*He*) varied from 0.392 to 0.811 with a mean value of 0.66. The mean of polymorphism information content (*PIC*) was 0.702 ± 0.087, and Shannon's information index (*I*) varied from 0.733 to 1.76 (mean 1.397 ± 0.298). The value of the inbreeding coefficient (F_{IS}) ranged from − 0.359 to 0.551 (mean -0.026 ± 0.253) (Table 2).

Population genetic diversity

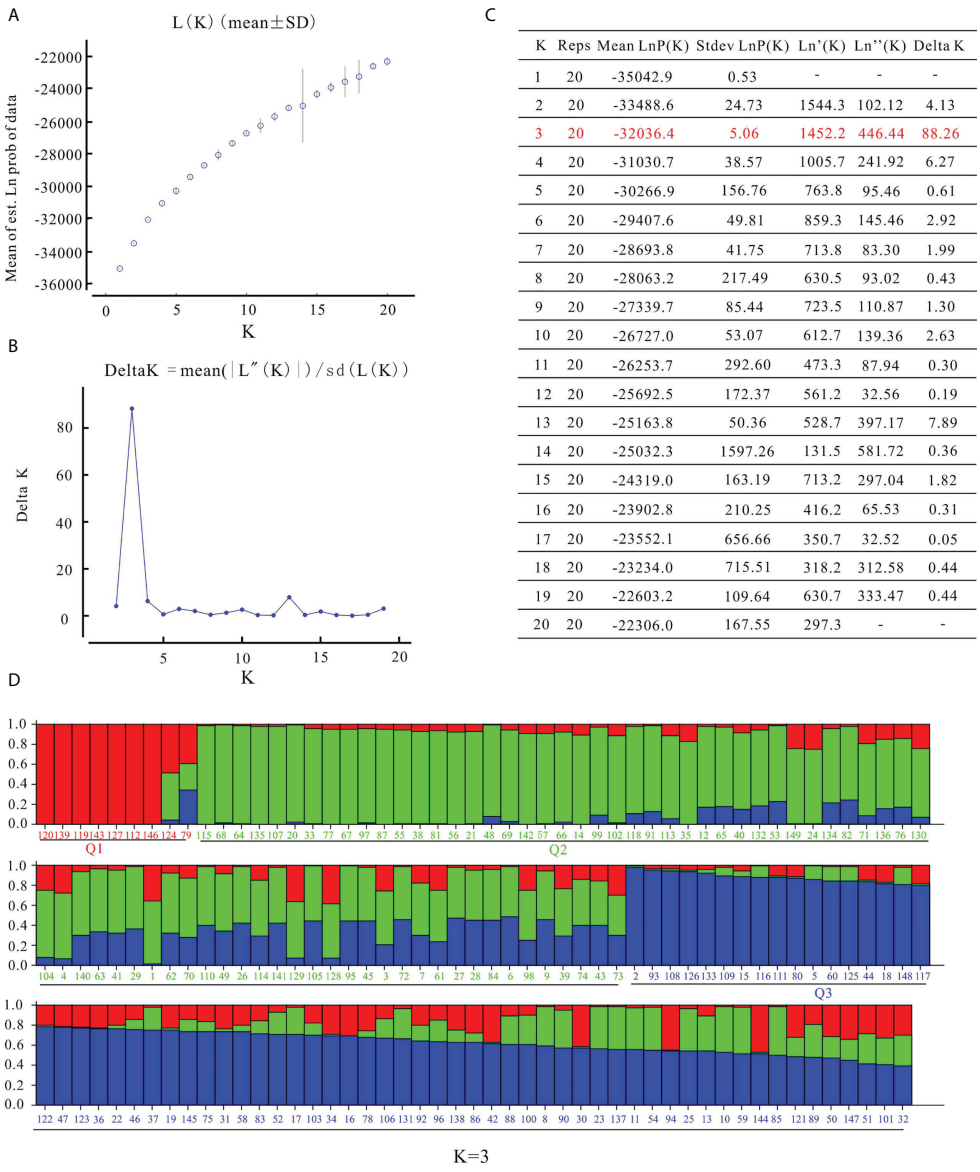
STRUCTURE software was used to characterize the population structure. The log-likelihood value LnP(D) continued to increase without clear inflection as K increased from 1 to 20 (Figure 3A). When using the ΔK method (Evanno et al., 2005), the peak value appeared at K = 3 (Figures 3B, C). In the results of Struture analysis (Figure 3D), there were 9 varieties in the Q1 group, 74 varieties in the Q2 group, and 66 varieties in the Q3 group. According to phylogenetic trees based on the neighbor-joining method, the 149 varieties were classified into three clusters (Figure 4). The first group (Cluster I) consisted of 14 varieties, including 8 varieties from the Q1 group, such as 'Zhongshu 3', 'Zhongshu 8, 'Zhongshu 12', 'Zhongshuzao 30', 'Zhongshu 33', 'Yunshu 101', 'Zhengshu 5' and 'Zhengshu 6' (Figure 4). Cluster II (green) consisted of 85 genotypes (Figure 4), including 63 varieties from the Q2 group. There were 50 varieties in cluster III (blue), including 42 varieties from the Q3 group. The genetic diversity of 149 varieties was also analyzed by PCoA. The results also classified the 149 potato varieties into three major groups (Supplementary Figure S2), which agreed with the results of STRUCTURE. The cultivars in Q1 were grouped in the upper-left corner of the plot, and the genotypes in Q2 were distributed in the lower-left part of the plot. The genotypes in the Q3 subgroup were distributed on the right side of the plot. Molecular variance analysis indicated that the major proportion of the variance was due to variation within groups, and the remaining 10% was due to variation among groups (Supplementary Table S6).

TABLE 2 Diversity information parameters of 149 cultivated potato genotypes using 24 SSR markers.

Marker	*Na*	*Ne*	*He*	*Ho*	*I*	*PIC*	F_{IS}
31924	6	4.45	0.776	0.969	1.596	0.780	-0.226
43016	7	2.57	0.611	0.516	1.441	0.680	0.449
S118	5	3.39	0.705	0.783	1.387	0.715	0.157
S151	6	3.93	0.746	0.913	1.577	0.767	-0.152
S170	6	4.74	0.789	0.988	1.629	0.790	-0.238
S182	5	3.66	0.727	0.907	1.397	0.697	-0.144
S187	5	3.15	0.683	0.888	1.356	0.707	-0.164
S189	9	5.29	0.811	0.994	1.760	0.813	-0.218
S192	4	2.35	0.574	0.733	1.175	0.619	-0.034
S7	7	4.66	0.786	0.925	1.706	0.794	-0.118
SSR08337	4	2.29	0.563	0.702	1.096	0.614	0.070
STG0025	2	1.99	0.498	0.857	1.726	0.654	-0.359
STG0026	4	2.39	0.582	0.795	1.072	0.622	-0.126
STI0012	7	4.25	0.765	0.969	1.641	0.779	-0.223
STI017	5	3.79	0.736	0.988	1.632	0.764	-0.320
STI032	6	4.42	0.774	0.969	1.678	0.790	-0.210
STM0037	6	5.19	0.807	0.988	1.711	0.809	-0.212
STM1049	4	1.79	0.441	0.571	0.786	0.655	0.137
STM1104	6	2.53	0.604	0.658	1.271	0.649	0.252
STM1106	8	2.82	0.645	0.497	1.573	0.713	0.496
STM2022	6	2.99	0.666	0.708	1.447	0.742	0.006
STM3012	5	2.8	0.643	0.857	1.241	0.678	-0.173
STM5121	4	2.11	0.526	0.460	0.908	0.547	0.551
STPoAc58	4	1.64	0.392	0.491	0.733	0.462	0.183
Mean	5.46	3.30	0.660	0.797	1.397	0.702	-0.026

Na, the number of alleles per locus (Na); Ne, the effective number of alleles per locus; Ho, observed heterozygosity; He, expected heterozygosity; I, Shannon's information index; PIC, polymorphism information content; F_{IS}, inbreeding coefficient.

10.3389/fpls.2022.954162



K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	Ln''(K)	Delta K
1	20	-35042.9	0.53	-	-	-
2	20	-33488.6	24.73	1544.3	102.12	4.13
3	20	-32036.4	5.06	1452.2	446.44	88.26
4	20	-31030.7	38.57	1005.7	241.92	6.27
5	20	-30266.9	156.76	763.8	95.46	0.61
6	20	-29407.6	49.81	859.3	145.46	2.92
7	20	-28693.8	41.75	713.8	83.30	1.99
8	20	-28063.2	217.49	630.5	93.02	0.43
9	20	-27339.7	85.44	723.5	110.87	1.30
10	20	-26727.0	53.07	612.7	139.36	2.63
11	20	-26253.7	292.60	473.3	87.94	0.30
12	20	-25692.5	172.37	561.2	32.56	0.19
13	20	-25163.8	50.36	528.7	397.17	7.89
14	20	-25032.3	1597.26	131.5	581.72	0.36
15	20	-24319.0	163.19	713.2	297.04	1.82
16	20	-23902.8	210.25	416.2	65.53	0.31
17	20	-23552.1	656.66	350.7	32.52	0.05
18	20	-23234.0	715.51	318.2	312.58	0.44
19	20	-22603.2	109.64	630.7	333.47	0.44
20	20	-22306.0	167.55	297.3	-	-

FIGURE 3
Analysis of population structure of 149 potato varieties. **(A)** Mean log-likelihood [Ln(K) ± SD], **(B)** △K values, **(C)** each K value based on the model reported in the article Evanno et al. (2005), **(D)** population structure of 149 potato varieties on K values of 3. The colored bar grouped the varieties in the corresponding populations, red, Q1 group; green, Q2 group; blue, Q3 group.

Identification of marker–trait associations

Based on the phenotypic traits and genotype data, associations between markers and traits were analyzed through TASSEL software using a mixed linear model (MLM). We performed 1572 (131 SSR alleles × 12 traits) marker-trait association tests. Thirty-three associations (2.09%) reached a significant level ($P < 0.01$), and the number was reduced to five (0.32%) after the false discovery rate (FDR) method test ($Q < 0.05$). Three loci (STI032_4, STI032_6 and SSR08337_1) were associated with starch content traits, explaining 7.31% to 12.56% of the variation. Two loci (STI032_4 and STI0012_3) were associated with the growth period, explaining 9.72% to 11.01% of the variation, respectively (Table 3 and Supplementary Figure S3).

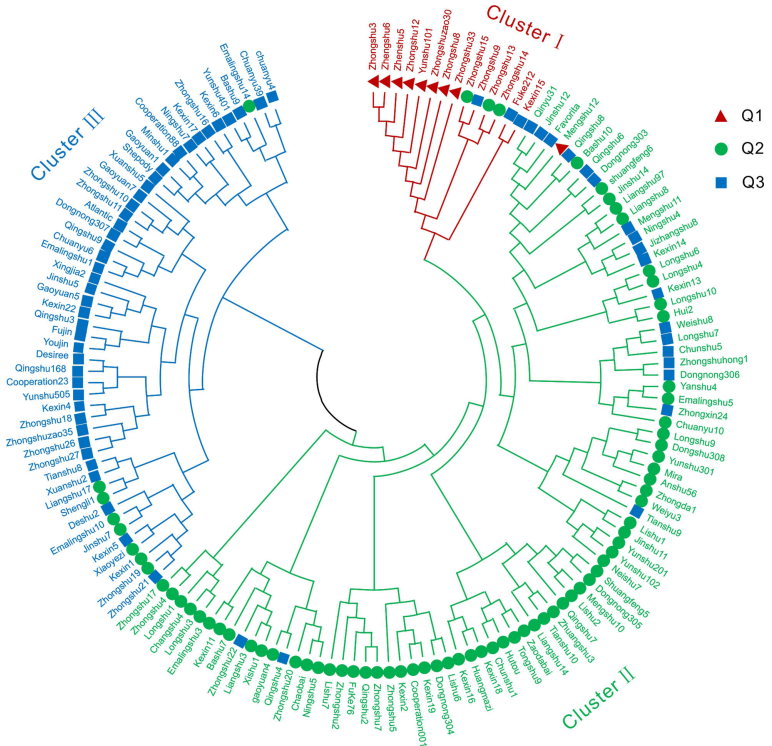


FIGURE 4
Phylogenetic tree of 149 potato varieties. A phylogenetic tree of all 149 potato varieties analyzed using 24 SSR markers was constructed by the neighbor-joining method. Cluster I included 14 varieties shown in red, cluster II included 85 varieties shown in green, and cluster III contained 50 varieties shown in blue. Q1, Q2 and Q3 represent the three groups of classification results of Structure analysis.

Discussion

Evaluation of cultivated potato germplasm resources

Generally, phenotypic traits are considered to be the basic elements for selecting clones from cross-combinations during potato breeding. The traits of plant maturity, tuber starch content, reducing sugar content, tuber shape, skin color and flesh color are widely used to select varieties in potato improvement programs (Jansky et al., 2009; Xu and Jin, 2017). Twelve phenotypic traits were evaluated in the study to stress the interaction between phenotypic variability and genetic diversity (Table 1, Figure 1).

Cluster analysis has been widely used for classifying plant genetic material. Dendrograms of the cluster analysis based on morphological and physiological traits for rice (Verma et al., 2019), tomato (Ayenan et al., 2021) and sweet potato (Meng et al., 2021). In this study, most traits of 149 cultivars showed significant differences. The dendrogram exhibited three large cultivar groups with significantly different values for several traits (Supplementary Figure S1). Our findings provide useful

TABLE 3 Marker–trait associations identified for agro-morphological traits of potato using MLM approaches in the software program TASSEL.

Traits	Locus	Chr.	P value	Q-value	PVE (%)	Estimate
Starch Content	STI032_4	5	9.24E-06	1.21E-03	12.56	-1.77
	STI032_6	5	2.27E-04	1.49E-02	8.86	1.56
	SSR08337_1	4	8.54E-04	3.73E-02	7.31	1.35
Growth Period	STI032_4	5	3.56E-05	4.67E-03	11.01	-13.18
	STI0012-3	4	1.08E-04	7.07E-03	9.72	-17.14

P value, the significance level of marker–trait associations using TASSEL software; Q-value, the significance level of false discoery rate (FDR) analysis in R software; PVE(%), phenotypic variation explained; Estimate, the value of gene effect estimated using TASSEL software.

information for selecting parents for hybrid combinations and identifying similar varieties for market supervision.

SSR markers have been widely used to estimate genetic diversity in many potato populations. The results of our study indicated that the average number of alleles per marker was 5.46 and the mean value of *PIC* was 0.702, demonstrating a high level of heterozygosity in tetraploid cultivated potato accessions. These results were consistent with those reported in previous studies (Kishine et al., 2017; Wang et al., 2019) for markers 31924, 43016, STG0025 and STM0037. However, lower allele numbers for markers S118, S170, S182, S187, S189, S192 and S7 were observed compared with those reported by Duan et al. (2019). Duan et al. (2021) reported a gene diversity of 0.258 in 189 potato genotypes, including 61 wild *Solanum* species genotypes, 32 *S. tuberosum Andienum* genotypes, 87 *S. tuberosum Chilotanum* genotypes, and 9 complex *Solanum* hybrids of diploids. Wang et al. (2019) reported a gene diversity of 0.309 in 292 potato germplasms of foreign elite lines, local landraces and cultivars in China. Research on the Colombian Central Collection of *S. tuberosum group Andigenum* showed a genetic diversity of 0.252-0.319 (Berdugo-Cely et al., 2017). A total of 214 advanced potato clones analyzed by the Infinium 22 K Potato Array revealed an overall average gene diversity of 0.59 (Pandey et al., 2021). Moreover, genotyping of 288 potato (*Solanum tuberosum L.*) accessions using SSR and AFLP markers revealed a genetic diversity of 0.311-0.367 (Wang et al., 2017), which is lower than that observed in the present study. The differences in the number of alleles might be due to the high genetic diversity of materials and highly polymorphic markers in the study.

Potato is a tetraploid outcrossing crop, and its heterozygosity is usually higher than expected (Machida-Hirano, 2015; Meirmans et al., 2018). In the present study, we detected a high level of genetic diversity in the population, and the expected and observed heterozygosity varied from 0.392 to 0.811 and 0.460 to 0.988, respectively. The mean value of observed heterozygosity (*Ho*) was higher than that of expected heterozygosity (*He*) in the study (Table 2). Additionally, the *PIC* varied from 0.462 to 0.813 with an average of 0.702, and the population F_{IS} changed from − 0.359 to 0.551 with an average of -0.025, of which 15 SSRs were negative, showing that there was significant heterozygous redundancy in the experimental potato varieties. This may be related to self-incompatibility introgression and heterozygosity of cultivated potato germplasm (Xu and Jin, 2017), which agrees with previously reported results (Meirmans et al., 2018). In addition, mutation positive selection and heterosis in the evolution process are also important reasons for the high heterozygosity of potato (Machida-Hirano, 2015).

Genetic relationship among the population

The detection of genetic structure is an important part of population genetic studies. Multiple factors, such as population size, natural selection, genetic drift and evolutionary history, can affect the genetic structure of potato germplasm. Wang et al. (2019) reported that based on collection sites, two main groups were subdivided into seven groups through Structure analysis. Wang et al. (2017) reported that global potato accessions could be classified into seven subgroups and an admixture group through Bayesian analysis. Pandey et al. (2021) also reported three groups of advanced clones based on potato marker classes detected by a potato breeding program in the USA. 189 potato genotypes were divided into five subgroups by STRUCTURE analysis (Duan et al., 2021). In this study, we compared the results of Structure analysis (Figure 3D) and phylogenetic tree clustering (Figure 4). Clustal I contains 8 varieties from the Q1 group, accounting for 88.9% of all Q1 varieties; Clustal II contains 63 varieties from the Q2 group, accounting for 85.1% of all Q2 varieties; and Clustal III contains 42 varieties from the Q3 group, accounting for 63.7% of Q3 varieties (Figure 4). If the similarity ancestry threshold is set at 60% (Zhao et al., 2010), 7 cultivars belong to the Q1 group, and Zhongshu 33 and Mengshu 12 are admixtures. Forty-eight cultivars belong to the Q2 group, 45 cultivars belong to the Q3 group, and the rest are mixtures. This result is consistent with the phylogenetic tree clustering. The difference between Structure analysis and phylogenetic tree clustering may be due to the different algorithms. The cluster results of Structure analysis (Figure 3D) and PCoA in this study were similar to those of previous reports (Duan et al., 2019; Duan et al., 2021), except for 'Longshu 7', 'Yanshu 4' and 'Fujin'. The grouping patterns or numbers were different from those in other studies, which may be caused by the use of different markers and populations.

The results of word cloud and network analysis indicated that seven potato varieties (cvs. Schwalbe, Zhongshu 3, Epoka, C93.154, Katahdin, Duozibai and Shepody) were highly frequently parental in these potato accessions (Figure 2). Similar results have been reported previously, demonstrating that the parents of Schwalbe, Katahdin, Epoka, Zhongshu 3, Duozibai and Shepody are often used extensively (Li et al., 2018a; Lee et al., 2021). Combined with the Structure data, Zhongshu 9 (No. 144) is the offspring of the cross combination of Zhongshu 3 (No. 139) × Shepody (No. 96), and Zhongshu 33 (No. 124) is the progeny of the cross combination of Zhongshu 3 (No. 139) × Zaodabai (No. 118). The introgression of the hybrids can be confirmed in the study (Figure 3D). In addition, five out of seven offspring of Schwalbe, five out of six offspring of Epoka,

and three offspring of Mira were all in the second green group (Figure 3D). Four of the five descendants of C93.154 and the three descendants of Desiree belong to the third blue group (Figure 3D). The AMOVA results indicated 90% variation within the groups and 10% variation among groups, which is similar to previous results (Wang et al., 2019; Duan et al., 2021).

Markers associated with starch content and growth period traits

Many studies have focused on QTLs for maturity and starch content in different populations in recent years. Li et al. (2019) identified eleven QTLs for tuber starch content from seven chromosomes and six QTLs for plant maturity from five chromosomes. A QTL was detected at 84 cM on chromosome 5, contributing to 33.55% of the variation in plant maturity (Li et al., 2018b). A major QTL on chromosome V for plant maturity was identified (Bradshaw et al., 2008). Researchers cloned the gene *StCDF1*, which encodes a DOF transcription factor that regulates tuberization and plant life cycle length (Kloosterman et al., 2013). A candidate SNP marker for plant maturity near the *StCDF1* gene was identified through SNP genetic mapping (Massa et al., 2015). In this study, we detected 4 SSRs located on chromosomes IV and V that showed significant genetic effects on maturity and starch content traits based on a false discovery rate (FDR) method test (Table 3), which enhanced the reliability of the association results.

As in previous reports, we also found a significant correlation between growth period and tuber starch content traits in potato (Figure 1) (Schönhals et al., 2016; Li et al., 2018b; Li et al., 2019). The starch content in late-maturing genotypes was higher than that in early-maturing cultivars, probably because more time was available to accumulate starch. Interestingly, the allele tagged by STI032_4 was negatively associated with both growth period and starch content. In contrast, the allele tagged by STI032_6 was positively associated with starch content traits. In contrast, the marker STI032 was only associated with plant maturity in 192 genotypes from the Longshu 8 × Zaodabai cross in a previous report (Li et al., 2019).

In conclusion, we analyzed the genetic diversity of 149 cultivated potato varieties in China using phenotypic traits and molecular SSR markers, and alleles of the STI032 marker with a significant correlation with starch content and growth period were detected. The study provided a useful tool for genetic analysis, varietal identification, breeding improvement and more extensive studies in potato.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Materials. Further inquiries can be directed to the corresponding author.

Author contributions

LJ designed the experiment; JH and MM performed the SSR analysis, MM, JX, SD, and CB provided the samples; JH, MM, and FJ analyzed the data and prepared the manuscript, LJ, GL, XW, and JH edited the manuscript. All authors read and confirmed the manuscript.

Funding

This research was funded by Agricultural Breeding Project of Ningxia Hui Autonomous Region, China, grant number 2019NYYZ01-1, Science and Technology Project of Inner Mongolia Autonomous Region, China, grant number 2021ZD0005, China Agriculture Research System, grant number CARS-9.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.954162/full#supplementary-material

References

Ayenan, M. A. T., Danquah, A., Agre, P. A., Hanson, P., Asante, I. K., and Danquah, E. Y. (2021). Genomic and phenotypic diversity of cultivated and wild tomatoes with varying levels of heat tolerance. *Genes* 12, 503. doi: 10.3390/genes12040503

Berdugo-Cely, J., Valbuena, R. I., Sánchez-Betancourt, E., Barrero, L. S., and Yockteng, R. (2017). Genetic diversity and association mapping in the Colombian central collection of *Solanum tuberosum l. Andigenum* group using SNPs markers. *PloS One* 12, e0173039. doi: 10.1371/journal.pone.0173039

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308

Bradshaw, J. E., Hackett, C. A., Pande, B., Waugh, R., and Bryan, G. J. (2008). QTL mapping of yield, agronomic and quality traits in tetraploid potato (*Solanum tuberosum subsp. tuberosum*). *Theor. Appl. Genet.* 116, 193–211. doi: 10.1007/s00122-007-0659-1

Chandra, A., Grisham, M. P., and Pan, Y. B. (2014). Allelic divergence and cultivar-specific SSR alleles revealed by capillary electrophoresis using fluorescence-labeled SSR markers in sugarcane. *Genome* 57, 363–372. doi: 10.1139/gen-2014-0072

D'Hoop, B., Keizer, P., Paulo, M. J., Visser, R., Eeuwijk, F. A., and Eck, H. J. (2014). Identification of agronomically important QTL in tetraploid potato cultivars using a marker–trait association analysis. *Theor. Appl. Genet.* 127, 731–748. doi: 10.1007/s00122-013-2254-y

Domański, L. (2001). Assessment of morphological characters of potato tubers. *Monografie i Rozprawy Naukowe IHAR Radzików. Poland* 10, 92–95.

Duan, Y., Duan, S., Xu, J., Zheng, J., Hu, J., Li, X., et al. (2021). Late blight resistance evaluation and genome-wide assessment of genetic diversity in wild and cultivated potato species. *Front. Plant Sci.* 12, 710468. doi: 10.3389/fpls.2021.710468

Duan, Y., Liu, J., Xu, J., Bian, C., Duan, S., Pang, W., et al. (2019). DNA Fingerprinting and genetic diversity analysis with simple sequence repeat markers of 217 potato cultivars (*Solanum tuberosum l.*) in China. *Am. J. Potato Res.* 96, 21–32. doi: 10.1007/s12230-018-9685-6

Earl, D. A., and Vonholdt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the evanno method. *Conserv. Genet. Resour.* 4, 359–361. doi: 10.1007/s12686-011-9548-7

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x

FAOSTAT (2020) *Statistical data*. Available at: www.fao.org/faost.

Hardy, O. J., and Vekemans, X. (2002). SPAGeDi: A versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol. Ecol. Notes* 2, 618–620. doi: 10.1046/j.1471-8286.2002.00305.x

Jansky, S. H., Jin, L. P., Xie, K. Y., Xie, C. H., and Spooner, D. M. (2009). Potato production and breeding in China. *Potato Res.* 52, 57. doi: 10.1007/s11540-008-9121-2

Kishine, M., Tsutsumi, K., and Kitta, K. (2017). A set of tetra-nucleotide core motif SSR markers for efficient identification of potato (Solanum tuberosum) cultivars. *Breed. Sci.* 67, 544–547. doi: 10.1270/jsbbs.17066

Kloosterman, B., Abelenda, J. A., Gomez, M., d., M. C., Oortwijn, M., de Boer, J. M., et al. (2013). Naturally occurring allele diversity allows potato cultivation in northern latitudes. *Nature* 495, 246–250. doi: 10.1038/nature11912

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Lee, K.-J., Sebastin, R., Cho, G.-T., Yoon, M., Lee, G.-A., and Hyun, D.-Y. (2021). Genetic diversity and population structure of potato germplasm in RDA-genebank: Utilization for breeding and conservation. *Plants* 10 (4), 752. doi: 10.3390/plants10040752

Liang, C., Wan, T., Xu, S., Li, B., Li, X., Feng, Y., et al. (2018). Molecular identification and genetic analysis of cherry cultivars using capillary electrophoresis with fluorescence-labeled SSR markers. *3 Biotech.* 8, 16. doi: 10.1007/s13205-017-1036-7

Li, L., Paulo, M. J., Strahwald, J., Lübeck, J., Hofferbert, H. R., Tacke, E., et al. (2008). Natural DNA variation at candidate loci is associated with potato chip color, tuber starch content, yield and starch yield. *Theor. Appl. Genet.* 116, 1167–1181. doi: 10.1007/s00122-008-0746-y

Liu, K., and Muse, S. V. (2005). PowerMarker: An integrated analysis environment for genetic marker analysis. *Bioinformatics* 21, 2128–2129. doi: 10.1093/bioinformatics/bti282

Li, J., Wang, Y., Wen, G., Li, G., Li, Z., Zhang, R., et al. (2019). Mapping QTL underlying tuber starch content and plant maturity in tetraploid potato. *Crop J.* 7, 261–272. doi: 10.1016/j.cj.2018.12.003

Li, X., Xu, J., Duan, S., Bian, C., Hu, J. , Shen, H., et al. (2018a). Pedigree-based deciphering of genome-wide conserved patterns in an elite potato parental line. *Front. Plant Sci.* 9, 690. doi: 10.3389/fpls.2018.00690

Li, X., Xu, J., Duan, S., Zhang, J., Bian, C., Hu, J., et al. (2018b). Mapping and QTL analysis of early-maturity traits in tetraploid potato (*Solanum tuberosum l.*). *Int. J. Mol. Sci.* 19, 3065. doi: 10.3390/ijms19103065

Machida-Hirano, R. (2015). Diversity of potato genetic resources. *Breed. Sci.* 65, 26–40. doi: 10.1270/jsbbs.65.26

Massa, A. N., Manrique-Carpintero, N. C., Coombs, J. J., Zarka, D. G., Boone, A. E., Kirk, W. W., et al. (2015). Genetic linkage mapping of economically important traits in cultivated tetraploid potato (*Solanum tuberosum l.*). *G3 Genes|Genomes|Genetics* 5, 2357–2364. doi: 10.1534/g3.115.019646

McCord, P. H., Sosinski, B. R., Haynes, K. G., and Clough, M. E. (2011). Linkage mapping and QTL analysis of agronomic traits in tetraploid potato (*Solanum tuberosum subsp. tuberosum*). *Crop Sci.* 51, 771–785. doi: 10.2135/cropsci2010.02.0108

Meirmans, P. G., Liu, S., and van Tienderen, P. H. (2018). The analysis of polyploid genetic data. *J. Hered.* 109, 283–296. doi: 10.1093/jhered/esy006

Meng, Y., Su, W., Ma, Y., Liu, L., Gu, X., Wu, D., et al. (2021). Assessment of genetic diversity and variety identification based on developed retrotransposon-based insertion polymorphism (RBIP) markers in sweet potato (*Ipomoea batatas (L.) lam.*). *Sci. Rep.* 11, 17116–17116. doi: 10.1038/s41598-021-95876-w

Pandey, J., Scheuring, D. C., Koym, J. W., Coombs, J., Novy, R. G., Thompson, A. L., et al. (2021). Genetic diversity and population structure of advanced clones selected over forty years by a potato breeding program in the USA. *Sci. Rep.* 11, 1–18. doi: 10.1038/s41598-021-87284-x

Peakall, R., and Smouse, P. E. (2012). GenAlEx 6.5: Genetic analysis in excel. population genetic software for teaching and research–an update. *Bioinformatics* 28, 2537–2539. doi: 10.1093/bioinformatics/bts460

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959. doi: 10.1093/genetics/155.2.945

R Core Team (2019). *R: A language and environment for statistical computing* (Vienna: R Foundation for Statistical Computing). Available at: https://www.R-project.org/.

Schönhals, E. M., Ortega, F., Barandalla, L., Aragones, A., Ruiz de Galarreta, J. I., Liao, J. C., et al. (2016). Identification and reproducibility of diagnostic DNA markers for tuber starch and yield optimization in a novel association mapping population of potato (*Solanum tuberosum l.*). *Theor. Appl. Genet.* 129, 767–785. doi: 10.1007/s00122-016-2665-7

Tiwari, J. K., Ali, N., Devi, S., Kumar, V., Zinta, R., and Chakrabarti, S. K. (2018). Development of microsatellite markers set for identification of Indian potato varieties. *Sci. Hortic.* 231, 22–30. doi: 10.1016/j.scienta.2017.11.027

van Eck, H. J. T. (2007). "Genetics of Morphological and Tuber Traits." In Potato Biology and Biotechnology. ed. D. Vreugdenhil, J. Bradshaw, C. Gebhardt, F. Govers, D. K. L. Mackerron, M. A. Taylor and H. A. Ross (Amsterdam: Elsevier Science B.V.), 91–115.

Verma, H., Borah, J. L., and Sarma, R. N. (2019). Variability assessment for root and drought tolerance traits and genetic diversity analysis of rice germplasm using SSR markers. *Sci. Rep.* 9, 16513–16513. doi: 10.1038/s41598-019-52884-1

Wang, J., Hou, L., Wang, R.-Y., He, M.-M., and Lliu, Q.-C. (2017). Genetic diversity and population structure of 288 potato (*Solanum tuberosum l.*) germplasms revealed by SSR and AFLP markers. *J. Integr. Agr* 16, 2434–2443. doi: 10.1016/S2095-3119(16)61619-2

Wang, F., Li, X., Yang, Y., Yi, H., Jiang, B., Zhang, X., et al. (2018). SSR analyser-a special software suitable for SSR fingerprinting of plant varieties. *Scientia Agricultura Sin.* 51, 2248–2262. doi: 10.3864/j.issn.0578-1752.2018.12.003

Wang, Y., Rashid, M. A. R., Li, X., Yao, C., Lu, L., Bai, J., et al. (2019). Collection and evaluation of genetic diversity and population structure of potato landraces and varieties in China. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00139

Xu, J., and Jin, L. (2017). Advances and perspectives in research of potato genetics and breeding. *Scientia Agricultura Sin.* 50, 990–1015. doi: 10.3864/j.issn.0578-1752.2017.06.003

Yang, H. (1989). National Potato Variety Resources Catalogue. Harbin, China: Heilongjiang Science and Technology Press.

Zhao, K., Wright, M., Kimball, J., Eizenga, G., McClung, A., Kovach, M., et al. (2010). Genomic diversity and introgression in *O. sativa* reveal the impact of domestication and breeding on the rice genome. *PloS One* 5, e10780. doi: 10.1371/journal.pone.0010780

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 29 November 2022
DOI 10.3389/fpls.2022.1059817

Check for updates

OPEN ACCESS

EDITED BY
Roland Pieruschka,
Helmholtz Association of German Research Centres (HZ), Germany

REVIEWED BY
Michael Benjamin Kantar,
University of Hawaii, United States
Fiona R. Hay,
Aarhus University, Denmark

*CORRESPONDENCE
Rainer Vollmer
r.vollmer@cgiar.org

SPECIALTY SECTION
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

RECEIVED 02 October 2022
ACCEPTED 14 November 2022
PUBLISHED 29 November 2022

CITATION
Vollmer R, Villagaray R, Castro M, Cárdenas J, Pineda S, Espirilla J, Anglin N, Ellis D and Rennó Azevedo VC (2022) The world's largest potato cryobank at the International Potato Center (CIP) – Status quo, protocol improvement through large-scale experiments and long-term viability monitoring.
Front. Plant Sci. 13:1059817.
doi: 10.3389/fpls.2022.1059817

COPYRIGHT
© 2022 Vollmer, Villagaray, Castro, Cárdenas, Pineda, Espirilla, Anglin, Ellis and Rennó Azevedo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

The world's largest potato cryobank at the International Potato Center (CIP) – Status quo, protocol improvement through large-scale experiments and long-term viability monitoring

Rainer Vollmer[1]*, Rosalva Villagaray[1], Mario Castro[1], José Cárdenas[1], Sandra Pineda[1], Janeth Espirilla[1], Noelle Anglin[1,2], Dave Ellis[1] and Vânia Cristina Rennó Azevedo[1]

[1]Genebank, International Potato Center (CIP), Lima, Peru, [2]Small Grains and Potato Germplasm Research unit, US Department of Agriculture (USDA), Aberdeen, ID, United States

Long-term conservation of Plant Genetic Resources (PGR) is a key priority for guaranteeing food security and sustainability of agricultural systems for current and future generations. The need for the secure conservation of genetic resources collections *ex situ* is critical, due to rapid and extreme climatic changes which are threatening and reducing biodiversity in their natural environments. The International Potato Center (CIP) conserves one of the most complete and diverse genetic resources collections of potato, with more than 7500 accessions composed of 4900 cultivated potato and 2600 potato wild relative accessions. The clonal conservation of cultivated potato, principally landraces, through *in vitro* or field collections is indispensable to maintain fixed allelic states, yet it is costly and labor-intensive. Cryopreservation, the conservation of biological samples in liquid nitrogen (-196°C), is considered the most reliable and cost-efficient long-term *ex-situ* conservation method for clonal crops. Over the last decade, CIP has built one of the largest potato cryobanks worldwide, cyopreserving more than 4000 cultivated potato accessions which represents 84% of the total cultivated potato collection currently conserved at CIP. In approximately, four years the entire potato collection will be cryopreserved. The development of an applied, robust cryopreservation protocol for potato, serves as a model for other clonally maintained crop collections. The CIP cryobank designs experiments with a high number of genetically diverse genotypes (70-100 accessions, seven cultivated species), to obtain reliable results that can be extrapolated over the collection as genotypes can often respond variably to the same applied conditions. Unlike most published reports on cryopreservation of plants, these

large-scale experiments on potato are unique as they examine the acclimatization process of *in vitro* plants prior to, as well as during cryopreservation on up to ten times the number of genotypes conventionally reported in the published literature. As a result, an operational cryopreservation protocol for potato has advanced that works well across diverse potato accessions, not only with reduced processing time and costs, but also with an increased average full-plant recovery rate from 58% to 73% (+LN) for routine cryopreservation. The present article describes the composition of CIP's cryobank, the cryopreservation protocol, methodology for the dynamic improvement of the operational protocol, as well as data collected on regeneration from long term cryopreserved potatoes.

KEYWORDS

potato, *Solanum*, cryopreservation, cryoconservation, cryobank, viability, recovery, PVS2

Introduction

Potato is the third most important staple crop worldwide, with a total production of $437x10^6$ tons on $20.7x10^6$ hectares, and an average yield of 20.9 tons/hectare (Devaux et al, 2014; FAOSTAT, 2022). Potatoes are rich in carbohydrates, protein, vitamin C, vitamin B6, magnesium, potassium, and fiber and frequently used as a primary nutrient source in countries where people cannot afford high energy diets (Zaheer and Akhtar, 2016).

The agrobiodiversity of important food crops is predominately conserved as dried botanical seed at cold temperatures. However, for 50% of the ten most important crops for human livelihoods (Goldshein, 2011), potato, cassava, sweet potato, yam, and banana, conserving the unique allelic combination in an individual (clone) is desired and this cannot be conserved through botanical seed. Thus, different conservation strategies, such as *in situ* conservation, conservation as an *ex situ* collection in the field or greenhouse, conservation of collections of vegetative propagules (e.g. tuber collections), *in vitro* conservation, or cryopreservation are utilized. At the International Potato Center, potatoes were mostly maintained as *ex situ* field collections of tubers which needed to be regenerated annually; however, in the 1980's an *in vitro* laboratory was established, and potato accessions were introduced and conserved *in vitro*. As cultivated potato is a clonally propagated crop, it is principally conserved in genebanks as tubers (at 4°C) or *in vitro* collections (at 7-22°C).

While the tubers have the advantage that stored germplasm can be directly planted in the field, infection by pathogens or insects, short storage periods, along with challenges due to space, cost and labor needed for regeneration and maintenance, are limiting factors for long-term conservation of tubers. In contrast, *in vitro* potato germplasm collections can be conserved as plants in test tubes, which can be maintained long-term as pathogen-free material, facilitating distribution of the collections for use. Although the *in vitro* storage period can be extended to 2-3 years between subcultures or propagation cycles, the plant material needs to be renewed periodically and any handling of the genebank material exposes the collection to the risk of human error (mix up of accessions), fungi or bacteria contamination, or somaclonal variation sometimes produced from tissue culture. As the renewal process occurs repeatedly, the probability of changes, due to human error or biology of the conserved material, can add up over time (Anglin et al., 2021). Many of the above-mentioned disadvantages can be lessened, mitigated, or eliminated using cryopreservation for clonal crop maintenance. This is one of the reasons the CIP genebank chose to begin a cryopreservation program.

Cryopreservation is defined as the conservation of biological material at ultra-low temperatures, generally in the liquid (-196° C) or vapor phase of liquid nitrogen (LN) and is considered a cost- and space-efficient long-term conservation method for clonal crops (Vollmer et al., 2021). *"Cryopreserved plant materials can theoretically remain alive for centuries, after which they can be removed from the frozen conditions and regenerated into healthy, growing plants"* (Food and Agriculture Organization (FAO), 2021). The genetic and morphological stability of cryopreserved material was confirmed in multiplies studies with diverse species, e.g. *Vanilla planifolia*, *Solanum tuberosum, Hedeoma todsenii*, *Passiflora pohlii*, Rubus *grabowskii, Malus*, etc. (Volk et al., 2008; Castillo et al., 2010; Merhy et al., 2014; Li et al., 2017; Pence et al., 2017; Gonzáles-Arnao et al, 2022)

In the late 1990s, CIP initiated research into cryobanking as a long-term conservation method for potato and other clonal crops. Many factors contributed to the decision to begin cryobanking and the reasons included the long-term security and stability of the collection, the reduction of annual

maintenance costs for *in vitro* conservation of potato and the ongoing challenge of needing skilled staff to phytosanitary clean and conserve a collection of 4850 potato *in vitro* accessions. While the cost of placing a sample into cryo is high (~400 US $/acc.), the annual maintenance cost once it is in cryo is ~7 US $/acc. (includes periodic replacement of cryotanks and a LN generator after 20 years of use) which overall is much lower than the annual cost of routine *in vitro* conservation (~65 US$/acc.).

Cryopreservation protocols for plants are typically developed and optimized with a small number of accessions, with the assumption that the results obtained can be successfully applied to large cryo-collections of hundreds or thousands of diverse genotypes. This is usually not the case because germplasm collections consist of high diversity and numerous different species. Therefore, variation in responses to virtually any manipulation exists, leading to the need for tailoring of protocols to adjust for variability in responses. Thus, at CIP, cryopreservation experiments are set up with >70 accessions including all cultivated species to help ensure robust application of results (Vollmer et al., 2018). Further, clear minimum *a priori* criteria have been established for accessions to qualify undergoing cryopreservation (viability rate, phytosanitary state, verification of genetic identity, representation of diversity, etc.), as well as the application of high-quality standards for all steps of the cryopreservation process (Reed, 2008; Food and Agriculture Organization (FAO), 2014; Vollmer et al., 2016).

There are very few reports in the literature on large-scale, long-term monitoring of viability of cryopreserved plant material in cryobanks (Vollmer et al., 2017). For such long-term monitoring of viability of a cryo collection through time, one needs to have a long-term vision of more than 100 years (more than the lifetime of a single scientist), and therefore, the number of samples placed into cryopreservation needs to be large enough to permit periodic future assessment of the viability of cryobanked accessions, without compromising the cryobank stock for long-term genebank conservation.

Lessons learned through the experience of operational-scale potato cryopreservation and adjustments made to the protocol to improve the efficiency of cryopreservation over time have produced a rigorous quality management system (QMS) which includes a viability monitoring program with extensive data collected on recovery and survivability of potato accessions after cryopreservation. Details from experiments and knowledge gained are presented to benefit others cryopreserving large plant genetic resources collections.

Materials and methods

Routine cryopreservation

Nodal stem segments from *in vitro* potato plants maintained in CIP's *in vitro* medium-term collection (7 ± 2°C with a 16h photoperiod and light intensity of 5-20 $\mu mol.m^{-2}.s^{-1}$) were maintained under active growth conditions and subcultured in 25x150 mm glass test tubes every 3-4 weeks (genotype-dependent), with 4-5 explants per tube, on solid Murashige and Skoog medium (MS) [Murashige and Skoog, 1962], supplemented with standard vitamins (MSP09, Caisson Laboratories, East Smithfield, PA), 25 $g.L^{-1}$ of sucrose and 3.0 $g.L^{-1}$ of Phytagel™ (P8169, Merck Sigma-Aldrich®, St. Louis, MO). The pH of the culture medium was adjusted to 5.60 ± 0.02 with NaOH or KOH (1-2 M) and HCl (1-2 M), prior to autoclaving at 121°C for 20 min. The multiplied plant material was incubated at 20 ± 2°C, with a 16h photoperiod and light intensity of 80–100 $\mu mol.m^{-2}.s^{-1}$ (fluorescent tubes, 36W, cool day light). *In vitro* plants were multiplied from one single tube (medium-term storage), to three, six, and fourteen tubes, in three subsequent subculture cycles within a period of 9-12 weeks.

The final, or fourth, multiplication cycle was performed in deep Petri dishes. In this case, uninodal stem segments were subcultured at a high density (80-90 explants per Petri dish), and incubated for one week under normal conditions (20 ± 2°C, 16h photoperiod, light intensity of 80–100 $\mu mol.m^{-2}.s^{-1}$), followed by a cold acclimatization period for two to three weeks (at 7 ± 2° C, 16h photoperiod, light intensity of 10-20 $\mu mol.m^{-2}.s^{-1}$), depending on the genotype-specific growth pattern of the accessions (Figure 1A).

Shoot tips, 0.8-1.2 mm long and 0.4-0.7 mm wide (genotype-dependent) containing 3-4 leaf primordia were excised from the cold-acclimated *in vitro* plants. Per accession, a total of 150 shoot tips were excised and placed on a small sterile piece of filter paper (size: 1x1 cm; 10 shoot tips per filter paper) that was supported on solid MS medium of the same composition as previously described, but supplemented with 2.8 $g.L^{-1}$ of Phytagel™ (culture medium is contained in a standard Petri dish) [Figure 1B].

Each set of ten shoot tips was treated with 2 mL of Loading Solution (LS) at room temperature (20-25°C) for 20 minutes (Figure 1C). LS is composed of liquid MS medium (MSP09, Caisson Lab.), supplemented with 2.0 M glycerol (G5516, Merck Sigma-Aldrich®) and 0.4 M sucrose. The pH of the LS was adjusted to 5.80 ± 0.02. Loading solution, Plant Vitrification Solution 2 (PVS2), and Rewarming Solution (RS) were prepared with milli-Q water, while for culture media deionized-distilled water was used. Treatments with LS, PVS2, and RS were performed in sterile 15 mL glass screw top test tubes. The time that each treatment ends was noted on small labels and stuck onto the test tube caps.

After the loading phase, LS was removed from the test tube with a Pasteur pipette and replaced with 2 mL of pre-chilled PVS2 (0°C). Shoot tips were treated for exactly 50 minutes with PVS2 (on ice) (Figure 1D). PVS2 is composed of liquid MS medium (MSP09, Caisson Lab.), supplemented with 3.28 M glycerol, 2.42 M ethylene glycol (324558, Merck Sigma-Aldrich®), 1.9 M dimethyl sulfoxide (D4540, Merck

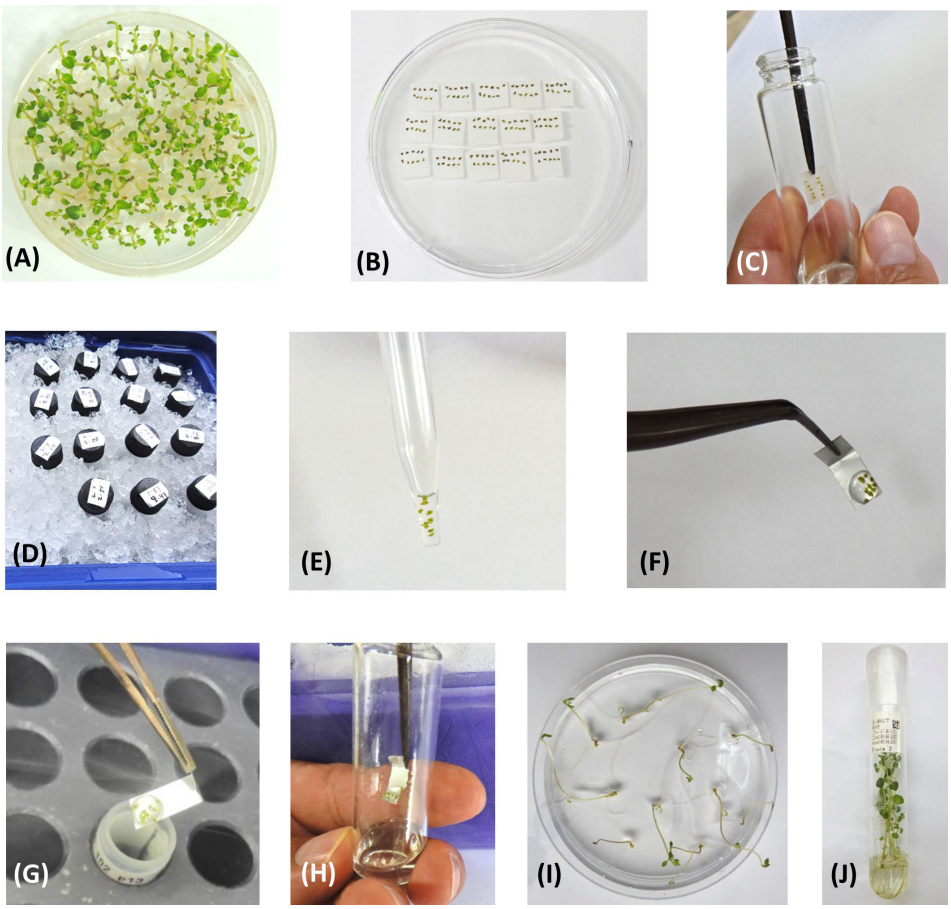


FIGURE 1
Steps in the CIP potato PVS2-droplet vitrification cryopreservation protocol. **(A)** *In vitro* potato plants grown from uninodal axillary stem segments after incubation for one week at 20 ± 2°C ("bud breaking"), followed by a cold acclimatization period of 2-3 weeks at 7 ± 1°C. Plants shown are ready for shoot tip excision. **(B)** Excised shoot tips (0.8-1.2 mm of size; genotype-dependent), placed on small filter papers, prior to cryoprotection treatment. **(C)** Introduction of shoot tips into vial containing 2 mL of Loading Solution (LS). Shoot tips are treated for 20 min in LS. **(D)** Treatment with Plant Vitrification Solution 2 (PVS2) [on ice, for 50 minutes]. **(E)** Ten shoot tips contained in the tip of a glass Pasteur pipette during the final minute in PVS2. **(F)** Shoot tips contained in a droplet of PVS2 placed onto a small aluminum foil strip (5x20 mm). **(G)** Aluminum strip is quickly plunged into liquid nitrogen (LN) ("Fast freezing"). **(H)** Thawing of shoot tips in 4 mL of Rewarming Solution (RS) at room temperature. Shoot tips are rewarmed for 20 minutes in RS. **(I)** Plants recovered from LN, 30 days after thawing. **(J)** *In vitro* plants 60 days after thawing, ready for further *in vitro* multiplication or direct transplant to the greenhouse.

Sigma-Aldrich®), and 0.4 M sucrose. The pH was adjusted to 5.80 ± 0.02.

One minute before the PVS2 treatment has concluded, shoot tips were sucked up with a Pasteur pipette (Figure 1E) and placed in a small droplet (20-25 µl) of pre-chilled PVS2, on the non-folded end of a small sterile aluminum foil strip (5 x 20 mm) [heavy duty quality, Boardwalk® BWK 7136, Essendant Co., Deerfield, IL]. The aluminum foil strip was previously folded at one end (fold of ~ 3-4 mm) to facilitate handling with forceps. The ten shoot tips were contained within one single droplet of uniform shape and volume (Figure 1F). The aluminum foil strip was quickly plunged ("fast freezing") into LN (Figure 1G) in a 1.8-mL cryovial (internal thread Nunc™, Thermo Fisher Scientific™, Waltham, MA) contained in a freezing box (CoolBox™ 30 Systems, Biocision, Larkspur, CA). Each cryovial contained one aluminum foil strip with ten shoot tips.

30 of the 150 shoot tips (three cryovials) from each cryopreservation run were thawed after a minimum 24h in LN to obtain the initial viability or recovery rate. For thawing, the aluminum foil strip was taken out of the cryovial and quickly plunged into 4 mL of recovery solution (RS) composed of liquid MS medium, supplemented with 0.6 M sucrose (pH: 5.80 ± 0.02) [Figure 1H]. The shoot tips were rewarmed in RS at room temperature for 20 minutes.

Thawed shoot tips were recovered with a Pasteur pipette and placed for ~30 seconds on a stack of 2-3 sterile filter papers to remove excess RS, and then transferred slipping the filter paper with the shoot tips facing down onto filter papers (2.5x2.5 cm) supported on MS culture media (as previously described), in a

standard Petri dish (10x150 mm). The culture medium was supplemented with 25 $g.L^{-1}$ sucrose, 20 $mL.L^{-1}$ coconut water (C5915, Merck Sigma-Aldrich®), 0.1 $mg.L^{-1}$ gibberellic acid (G7645, Merck Sigma-Aldrich®), 0.4 $mg.L^{-1}$ of kinetin (K0753, Merck Sigma-Aldrich®), and 2.8 $g.L^{-1}$ of Phytagel. The pH was adjusted to 5.60 ± 0.02. Three filter papers, each containing ten shoot tips (total: 30 shoot tips/Petri dish) were contained in each Petri dish. Petri dishes were wrapped completely in aluminum foil and incubated in the dark at 20 ± 2°C for nine days. After nine days, the ten shoot tips contained on each filter paper were removed from the filter paper and transferred directly onto individual Petri dishes with fresh culture medium of the same composition (each Petri dish contained 10 shoot tips). Samples were incubated at 20 ± 2°C, under diffuse light conditions (covering the top of the Petri dishes with a sheet of aluminum foil), with a 16h photoperiod and light intensity of 80–100 $\mu mol.m^{-2}.s^{-1}$. After four days, the aluminum foil was removed, and the samples were incubated for 17 additional days under normal light conditions. Samples that developed into complete *in vitro* plants or showed the potential to do so (bud breaking) (Figure 1I), were transferred to 25x150 mm glass test tubes containing the same culture medium, with a maximum of five plants per tube. Samples were incubated in the same environmental conditions for 30 more days.

Full plant recovery was assessed 60 days after thawing (Figure 1J). A sample was considered as recovered when it developed into a complete normal looking *in vitro* plant, with a functional apex, elongated stem, leaf formation, and rooting. Plants coming from each of the three Petri dishes were assessed independently (three replicates of ten shoot tips per accession).

Large-scale experiment – Effect of sucrose in recovery cycle

A set of 73 diverse potato landraces (Table 1) selected from the CIP genebank were cryopreserved with the above-described routine cryopreservation protocol (Vollmer et al., 2021). After thawing, shoot tips were placed on culture media with four different sucrose concentrations (0.07 M, 0.1 M, 0.2 M, and 0.3 M), in the dark, for nine days before they were transferred to culture medium with the standard sucrose concentration (0.07 M). As a control treatment, the previously used routine protocol for recovery was used, i.e. shoot tips were maintained on 0.3 M sucrose-rich medium for three days, and then transferred every three days to medium with a step-wise reduction in sucrose (0.2 M and then 0.1 M sucrose, total time period: nine days), before being transferred to culture media with standard sucrose concentration (0.07 M sucrose). More than 3000 potato accessions were cryopreserved between 2010 and 2019 applying the routine control treatment.

The experiment was set up as a completely randomized factorial design (73x5), and repeated three times, with a sample size of n=15 per repetition (three replicates of five shoot tips); hence, the average recovery rates shown in the figures are based on 45 shoot tips. As data was not distributed normally (Anderson-Darling test: $p<0.005$), nor any adequate data transformation could be applied (e.g. arcsine square root transformation) [Velleman and Hoaglin, 1981], the data was analyzed with the non-parametric Kruskal Wallis test for multiple samples ($\alpha=0.05$). Data analysis was performed with SPC for Excel (V 6.0, BPI Consulting) and Minitab™ 17.1.0. Python software (V3.11).

Long-term viability monitoring experiment

In 2013, a long-term viability monitoring experiment was initiated (completely randomized design), cryopreserving each year a higher number of shoot tips per accession (240) for 8-14 additional accessions annually than the routine protocol. After a minimum of 24 hours in LN, a sample of 30 shoot tips was thawed and recovered with the current available routine protocol. After 2, 4, and 8 years in LN additional these same samples (30 shoot tips each) were pulled out from LN and recovered with the same protocol for thawing and recovery that was used for the original sample to evaluate the effect of time on these accessions in cryopreservation. The long-term design of the experiment is to remove additional samples after 16, 32, and 64 years and process them as previously described. To date, 101 potato accessions have been included in this experiment, and the 2-year, 4-year and 8-year data is based on 76, 52 and 17 accessions, respectively. The experiment includes 40 of 45 accessions of CIP's diverse mini-core collection which includes seven cultivated potato species. Data was analyzed with the Minitab™ 17.1.0.

Results

Composition of CIP's potato cryobank – Routine cryopreservation

The potato cryobank at CIP currently holds 4086 accessions, belonging to seven cultivated potato species based on the taxonomy of Hawkes (1990), representing 83.7% of the *in vitro* collection and 83.0% of the total potato clonal collection in the CIP genebank. Most of the accessions are *S. tuberosum* subsp. *andigenum* (2797 acc.), which accounts for 68.4% of all cryobanked accessions, followed by *S. stenotomum* subsp. *stenotomum* (264 acc.), *S. phureja* (165 acc.), *S. tuberosum* subsp. *tuberosum* (151 acc.), *S.* ×*chaucha* (120 acc.), and *S. stenotomum* subsp. *goniocalyx* (90 acc.). The bitter potatoes, *S.* ×*curtilobum*, *S.* ×*ajanhuiri and S.* ×*juzepczukii*, have less

TABLE 1 List of 73 potato landraces that were used in the large-scale experiment study.

Accession identifier	ITPGRFA GLIS-DOIs	Species/subspecies	Country of origin
CIP 704178	10.18730/AAV5	*S. tuberosum* subsp. *tuberosum*	Chile
CIP 705068	10.18730/B40~	*S. tuberosum* subsp. *tuberosum*	Chile
CIP 706679	10.18730/CH3~	*S. tuberosum* subsp. *tuberosum*	Chile
CIP 706713	10.18730/CJ0S	*S. tuberosum* subsp. *andigenum*	Argentina
CIP 704087	10.18730/A8A=	*S. tuberosum* subsp. *andigenum*	Bolivia
CIP 704767	10.18730/AVFK	*S. tuberosum* subsp. *andigenum*	Bolivia
CIP 704980	10.18730/B1HW	*S. tuberosum* subsp. *andigenum*	Bolivia
CIP 705127	10.18730/B5RF	*S. tuberosum* subsp. *andigenum*	Colombia
CIP 704057	10.18730/A7G9	*S. tuberosum* subsp. *andigenum*	Ecuador
CIP 705352	10.18730/BC0Y	*S. tuberosum* subsp. *andigenum*	Ecuador
CIP 705407	10.18730/BDF3	*S. tuberosum* subsp. *andigenum*	Mexico
CIP 700921	10.18730/91RP	*S. tuberosum* subsp. *andigenum*	Peru
CIP 701561	10.18730/96BN	*S. tuberosum* subsp. *andigenum*	Peru
CIP 702152	10.18730/9ATG	*S. tuberosum* subsp. *andigenum*	Peru
CIP 702190	10.18730/9B1Q	*S. tuberosum* subsp. *andigenum*	Peru
CIP 702439	10.18730/9D9N	*S. tuberosum* subsp. *andigenum*	Peru
CIP 702343	10.18730/9CJ=	*S. tuberosum* subsp. *andigenum*	Peru
CIP 703971	10.18730/A4X0	*S. tuberosum* subsp. *andigenum*	Peru
CIP 704406	10.18730/AGR9	*S. tuberosum* subsp. *andigenum*	Peru
CIP 704501	10.18730/AKKT	*S. tuberosum* subsp. *andigenum*	Peru
CIP 704541	10.18730/AMRT	*S. tuberosum* subsp. *andigenum*	Peru
CIP 703244	10.18730/9J9~	*S. stenotomum* subsp. *goniocalyx*	Bolivia
CIP 701611	10.18730/96W1	*S. stenotomum* subsp. *goniocalyx*	Peru
CIP 702961	10.18730/9H2Z	*S. stenotomum* subsp. *goniocalyx*	Peru
CIP 703282	10.18730/9K3P	*S. stenotomum* subsp. *goniocalyx*	Peru
CIP 703352	10.18730/9MV4	*S. stenotomum* subsp. *goniocalyx*	Peru
CIP 703777	10.18730/9ZNH	*S. stenotomum* subsp. *goniocalyx*	Peru
CIP 704243	10.18730/ACJQ	*S. stenotomum* subsp. *goniocalyx*	Peru
CIP 705458	10.18730/BEYD	*S. stenotomum* subsp. *goniocalyx*	Peru
CIP 705575	10.18730/BJGG	*S. stenotomum* subsp. *goniocalyx*	Peru
CIP 706036	10.18730/BZW0	*S. stenotomum* subsp. *goniocalyx*	Peru
CIP 702287	10.18730/9C0H	*S. stenotomum* subsp. *stenotomum*	Bolivia
CIP 702587	10.18730/9EQY	*S. stenotomum* subsp. *stenotomum*	Bolivia
CIP 703473	10.18730/9Q98	*S. stenotomum* subsp. *stenotomum*	Bolivia
CIP 704141	10.18730/A9T9	*S. stenotomum* subsp. *stenotomum*	Bolivia
CIP 704771	10.18730/AVKQ	*S. stenotomum* subsp. *stenotomum*	Bolivia
CIP 705952	10.18730/BXK1	*S. stenotomum* subsp. *stenotomum*	Bolivia
CIP 706250	10.18730/C5G*	*S. stenotomum* subsp. *stenotomum*	Bolivia
CIP 706845	10.18730/CNTU	*S. stenotomum* subsp. *stenotomum*	Bolivia
CIP 700362	10.18730/8X4P	*S. stenotomum* subsp. *stenotomum*	Peru
CIP 701676	10.18730/97HP	*S. stenotomum* subsp. *stenotomum*	Peru
CIP 702353	10.18730/9CM0	*S. stenotomum* subsp. *stenotomum*	Peru
CIP 702834	10.18730/9G63	*S. stenotomum* subsp. *stenotomum*	Peru
CIP 703288	10.18730/9K9W	*S. stenotomum* subsp. *stenotomum*	Peru
CIP 703312	10.18730/9KWA	*S. stenotomum* subsp. *stenotomum*	Peru
CIP 703314	10.18730/9KYC	*S. stenotomum* subsp. *stenotomum*	Peru
CIP 703318	10.18730/9M2G	*S. stenotomum* subsp. *stenotomum*	Peru

(Continued)

TABLE 1 Continued

Accession identifier	ITPGRFA GLIS-DOIs	Species/subspecies	Country of origin
CIP 703709	10.18730/9XRY	*S. stenotomum* subsp. *stenotomum*	Peru
CIP 704043	10.18730/A72*	*S. stenotomum* subsp. *stenotomum*	Peru
CIP 705477	10.18730/BFGZ	*S. stenotomum* subsp. *stenotomum*	Peru
CIP 705569	10.18730/BJBB	*S. stenotomum* subsp. *stenotomum*	Peru
CIP 703294	10.18730/9KD*	*S. phureja*	Colombia
CIP 703506	10.18730/9R4=	*S. phureja*	Colombia
CIP 705802	10.18730/BSFH	*S. phureja*	Colombia
CIP 704203	10.18730/ABKX	*S. phureja*	Ecuador
CIP 706764	10.18730/CKD~	*S. phureja*	Ecuador
CIP 706825	10.18730/CN8J	*S. phureja*	Ecuador
CIP 701025	10.18730/92PF	*S. phureja*	Peru
CIP 703654	10.18730/9W7J	*S. phureja*	Peru
CIP 704859	10.18730/AY0T	*S. ×chaucha*	Bolivia
CIP 702208	10.18730/9B8Y	*S. ×chaucha*	Peru
CIP 704047	10.18730/A76U	*S. ×chaucha*	Peru
CIP 707129	10.18730/CS5*	*S. ×chaucha*	Peru
CIP 707136	10.18730/CS9U	*S. ×chaucha*	Peru
CIP 706211	10.18730/C4FU	*S. ×ajanhuiri*	Bolivia
CIP 706213	10.18730/C4G0	*S. ×ajanhuiri*	Bolivia
CIP 703810	10.18730/A0J9	*S. ×ajanhuiri*	Peru
CIP 704234	10.18730/ACEK	*S. ×juzepczukii*	Bolivia
CIP 706777	10.18730/CKT9	*S. ×juzepczukii*	Bolivia
CIP 706050	10.18730/C09D	*S. ×juzepczukii*	Peru
CIP 706776	10.18730/CKS8	*S. ×curtilobum*	Bolivia
CIP 702455	10.18730/9DET	*S. ×curtilobum*	Peru

The table shows accessions identifier (CIP number), International Treaty for Plant Genetic Resources (ITPGRFA) Global Information System Digital Object Identifier (GLIS-DOI) of the accession, species/subspecies, and country of origin.

representation in the cryobank compared to the other species, with only 7-27 acc. per species (Table 2).

Within the cryobank, 3757 of 4086 accessions are landraces (92.0%), that were collected or obtained from Andean countries, the primary center of diversity of cultivated and wild potato. 61% of the accessions are originally from Peru, followed by Bolivia (11%), Ecuador (8%), Colombia (6%), Argentina (4%) and Chile (3%). On the species-level, one or two countries of origin are predominant, e.g. 80% of the accessions of *S. tuberosum* subsp. *tuberosum* are derived from Chile, while for *S. tuberosum* susp. *andigenum*, *S. stenotomum* subsp. *stenotomum, S. stenotomum* subsp. *goniocalyx, S. ×chaucha*, and *S. ×curtilobum*, 68-97% of these accessions are originally from Peru. *S. ×ajanhuiri* has its origin mainly in Bolivia (91% of accessions), while *S. ×juzepczukiii* accessions were collected from Bolivia (52%) and Peru (44%). *S. phureja* has its primary center of diversity in Colombia and Ecuador, with 52% and 39% of the accessions coming from these two countries, respectively (Table 2).

The ploidy levels of the cryobanked potato species are 2x, 3x, 4x, and 5x. The diploid and pentaploid species showed a lower average recovery rate of 54.4 to 60.8%, compared to the tetraploid species (63.8 – 64.0%). Interestingly, a triploid *S. xchaucha* species showed the highest average recovery rate of all species (67.1%). The improved cultivars showed a high average full-plant recovery rate of 65.6%. When evaluating the recovery rate within each species, the accession-specific recovery rate determined to have a wide range from 20 to 100% (Table 2).

Large scale experiment - Effect of sucrose concentration in recovery medium

Placing shoot tips post-thaw on recovery medium with the normal sucrose concentration (0.07 M) for nine days significantly increased the average survival rate (SR: 77.4%) and recovery rate (RR: 71.5%), compared to the routine control cryopreservation protocol used at CIP (SR: 69.7%, RR: 59.5%), involving a step-wise decrease in sucrose concentration (three days on 0.3 M, 0.2 M and 0.1 M, respectively). Samples

TABLE 2 Composition of the potato cryobank of the International Potato Center (CIP), classified by species/subspecies (based on the taxonomy of Hawkes, 1990), ploidy level (PL), number of accessions (NoA), countries of origin, and average recovery rate (ARR) and range of recovery rates (RoRR).

Species/subspecies	PL	NoA	Countries of origin	ARR (%)	RoRR (%)
S. tuberosum subsp. *tuberosum* L.	4x	151	Chile (121); Argentina (9); Peru (3); Mexico (3); New Zealand (3); Colombia (2); Bhutan (2); India (2); Philippines (2); Russia (2); Venezuela (1); Guatemala (1)	63.8%	25.0 – 100.0%
S. tuberosum subsp. *andigenum* (Juz. & Bukasov) Hawkes	4x	2797	Peru (1892); Bolivia (317); Ecuador (224); Argentina (138); Colombia (130); Venezuela (32); Guatemala (27); Mexico (26); Bangladesh (5); Russia (5); Philippines (1)	64.0%	20.0 – 100.0%
S. stenotomum subsp. *stenotomum* (Juz. & Bukasov) Hawkes	2x	264	Peru (185); Bolivia (72); Colombia (2); Ecuador (2); Russia (2); Argentina (1)	59.2%	23.3 – 100.0%
S. stenotomum subsp. *goniocalyx* (Juz. & Bukasov) Hawkes	2x	90	Peru (87); Bolivia (1); Chile (1); Costa Rica (1)	60.8%	25.0 - 100.0%
S. ×chaucha Juz. & Bukasov	3x	120	Peru (89); Bolivia (23); Ecuador (8)	67.1%	26.7 – 100.0%
S. phureja Juz. & Bukasov	2x	165	Colombia (86); Ecuador (65); Peru (12); Russia (2)	58.3%	26.0 – 100.0%
S. ×ajanhuiri Juz. & Bukasov	2x	11	Bolivia (10); Peru (1)	54.4%	25.0 – 83.3%
S. ×juzepczukii Juz.	3x	27	Bolivia (14); Peru (12); Argentina (1)	58.6%	23.3 – 90.0%
S. ×curtilobum Juz. & Bukasov	5x	7	Peru (5); Bolivia (1); Argentina (1)	59.1%	30.0 – 76.7%
Unclassified landraces	?	125	Peru (73); Bangladesh (12); Bolivia (11); Ecuador (11); Sweden (7); Argentina (4); Bhutan (3); Guatemala (2); Colombia (1); New Zealand (1);	62.3%	25.0 – 100.0%
Improved cultivars	4x	304	Peru (117); United Kingdom (14); United States of America (23); Mexico (20); Colombia (16); India (16); Czech Republic (13); Netherlands (12); Argentina (10); Germany (9); Poland (8); Chile (7); Ecuador (7); Brazil (6); Russia (3); Rwanda (3); Bolivia (2); China (2); Japan (2); Kenya (2); Canada (2); South Korea (1); Philippines (1); Belgium (1); Cameroon (1); El Salvador (1); Turkey (1); Tanzania (1); Uzbekistan (1); Vietnam (1); Democratic Republic of the Congo (1)	65.6%	28.3 – 100.0%
Research material	TBD	25	United Kingdom (10); Peru (7); Netherlands (5); Germany (2); Mexico (1)	64.2%	30.0 – 100.0%
TOTAL	**4086**		**63.5%**		

The number of accessions per country is indicated in parenthesis next to each country. TBD, To be determined.

showing green tissue were recorded as survived, and those that developed into complete normal looking *in vitro* plants were considered as recovered. A stable sucrose concentration of 0.1 M resulted in a significantly higher recovery rate (62.5%) compared to higher sucrose concentrations of 0.2 M (RR: 53.8%) and 0.3 M (RR: 48.8%). The differential between SR and RR was higher with sucrose concentrations of 0.2 M and 0.3 M (13.8-16.5%), compared to lower concentrations of 0.07 M, 0.1 M, or the routine control treatment (5.9-10.2%) [Figure 2]. As the experiment was performed with 73 accessions representing seven potato species and four different ploidy levels, this represents a sample size of 1.5% in relation to the size of the complete *in vitro* potato collection at CIP.

43 of 73 assessed accessions (59%) showed the highest RR on culture medium with a sucrose concentration of 0.07 M, while 12 and six of 73 accessions had its highest RR at a sucrose concentration of 0.1 M and the routine control treatment, respectively. Surprisingly, five of the 73 tested accessions had the highest RR on a sucrose concentration of 0.2 M, which is a 2.7-times higher sucrose concentration compared to the standard culture medium of potato (Table 3).

Six species/subspecies, *S. tuberosum* subsp. *andigenum*, *S. tuberosum* subsp. *tuberosum*, *S. x ajanhuiri*, *S. stenotomum* subsp. *goniocalyx*, *S. stenotomum* subsp. *stenotomum* and *S. phureja*, showed a similar pattern, i.e. a pronounced decrease in RR as the sucrose concentration of the culture medium increased (0.07 M vs 0.1 M vs 0.2 M vs. 0.3 M). In absolute values of RR, the routine control showed a very similar pattern to the 0.1 M sucrose treatment (Figure 3A).

In contrast, *S. xchaucha*, *S. xcurtilobum* and *S. xjuzepczukii*, showed a relative stable RR with increasing sucrose concentration. In the case of *S. xcurtilobum* accessions, the highest RR was observed at a sucrose concentration of 0.2 M (Figure 3B).

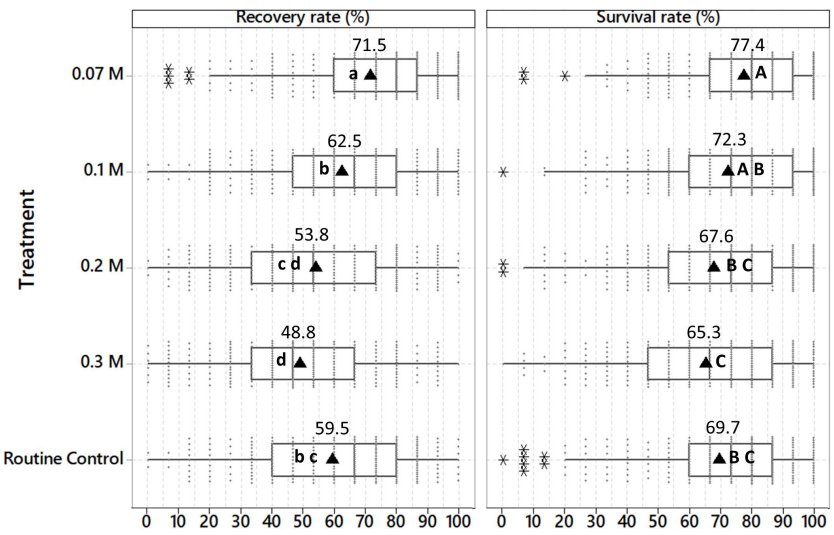


FIGURE 2
Average survival and recovery rates of 73 potato accessions cryopreserved with the PVS2-droplet vitrification method (+ LN), 60 days after thawing. A sample was considered as recovered when it developed into a complete normal looking *in vitro* plant, with a functional apex, elongated stem, leaf formation, and rooting. Plants that showed only leaf formation, greening, deformation or hyperhydration were considered as survival only. Shoot tips were excised from cold-acclimated *in vitro* plants (at 7°C) and cryopreserved with the PVS2-droplet vitrification method. Thawed shoot tips were placed for 9 days in darkness on culture media with different sucrose concentrations (0.07 M, 0.1 M, 0.2 M, and 0.3 M) and then transferred to culture medium with standard sucrose concentration (0.07 M). After thawing the shoot tips of the routine control treatment were maintained for 3 days on each of the 0.3 M, 0.2 M, and 0.1 M sucrose-rich culture medium (total period of 9 days, in darkness.), and then transferred to culture medium with standard sucrose concentration (0.07 M). The experiment was repeated three times (n = 15 per repetition and treatment). Different lower-case and upper-case letters indicate significant differences for the multiple Kruskal Wallis test of recovery and survival rate, respectively ($p < 0.05$). Filled circle: individual values; black triangle: mean value; asterisk: outlier value.

The experimental results were confirmed during routine operations, with a substantial increase in the average RR from 57.9% (3067 accessions; routine control) to 73.2% (1019 accessions; 0.07 M sucrose). The mode of RR distribution, shifted from 60-70% to the 80-90% interval, i.e. the distribution is now skewed to the left with higher frequencies of RR close to the 80-100% ranges. In contrast, the distribution of RR of the routine control (3067 acc.), was close to normal, with 1694 and 1373 accessions showing recovery rates in the ranges of 20-60% and 61-100%, respectively (Figure 4).

Long-term viability monitoring experiment

After two, four, and eight years in LN, the accessions showed no significant differences in full-plant recovery rates for a hypothesized mean difference of less than 10% between years (paired t-test; confidence level of 95%). The two-year set (76 acc.) had an average recovery rate of 67.1% after two years in LN, compared to the original rate of 67.7% (after minimum 24h in LN) [Figure 5A]. The accessions of the four-year set (52 acc.) showed average recovery rates of 63.3% and 64.5%, after two and four years in LN, compared to the original rate of 64.0% (Figure 5B). Finally, after eight years in LN (17 acc.), the accessions showed a stable average recovery rate of 51.3%, compared to the original recovery rate (50.0%), and the average rates after two (52.5%) and four years (52.7%) in LN (Figure 5C). The average recovery rate of the 8-year set was lower than the other two time periods, as during this earlier time period, the potato cryopreservation protocol was still in development.

Discussion

To the best of the authors knowledge CIP conserves the largest cryo-collection in the world belonging to one single genus (*Solanum*). Over ten years, CIP has built up its potato cryobank with ~4100 potato accessions for long-term storage, with an average full-plant recovery rate of 63.5%. The improvement of the first phase of the post-thawing recovery cycle (reduction in sucrose concentration in the post-thaw recovery medium), resulted in a substantial increase of the average recovery rate of approximately 15%, which was confirmed during routine cryopreservation with >1000 potato accessions (seven species). These results suggest that the long-held dogma that has existed in plant cryopreservation that a high sucrose concentration was

10.3389/fpls.2022.1059817

TABLE 3 Percent of 73 cryopreserved potato accessions showing the highest recovery rate on each of the different post-thaw sucrose concentrations 60 days after thawing.

	Sucrose concentration in recovery medium during the first nine days of the recovery cycle							
	0.3 M	0.2 M	0.1 M	0.07 M	RC	0.07 M or 0.1 M	0.07 M or RC	0.1 M or RC
Number of accessions	0	5	12	43	6	2	2	3
Percent	0%	7%	16%	59%	8%	3%	3%	4%

Seven accessions showed the highest recovery rate with more than one treatment (three columns to the right, i.e. 0.07 M/0.1 M, 0.07 M/RC, and 0.1 M/RC). The treatments are described in the caption of Figure 1. RC, Routine Control.

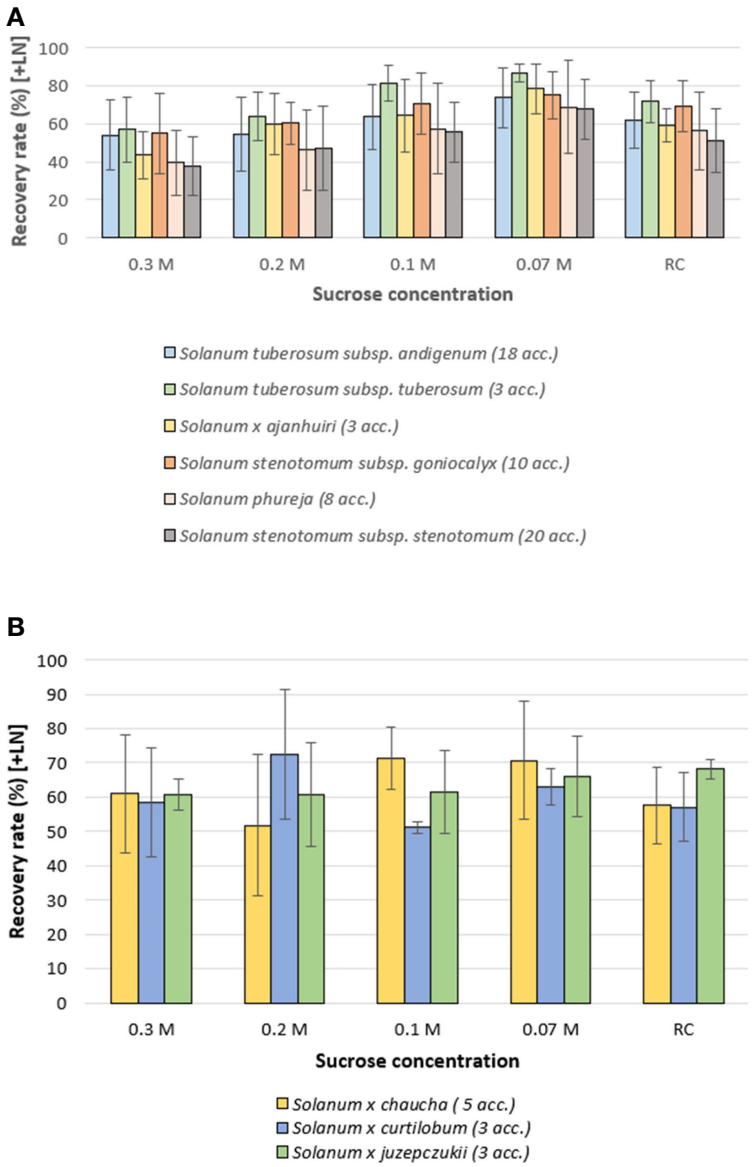


FIGURE 3
Average recovery rates of 73 cryopreserved potato accessions (+ LN), 60 days after thawing, classified by taxa. Thawed shoot tips were recovered on culture media with different sucrose concentrations (for details see caption Figure 1; RC, Routine Control). **(A)** Six species/subspecies of potato with similar patterns of recovery rate vs decreasing sucrose concentration in the recovery medium. **(B)** Three potato taxa that maintained stable recovery rates on culture medium with higher sucrose concentrations. Error bars: standard deviation.

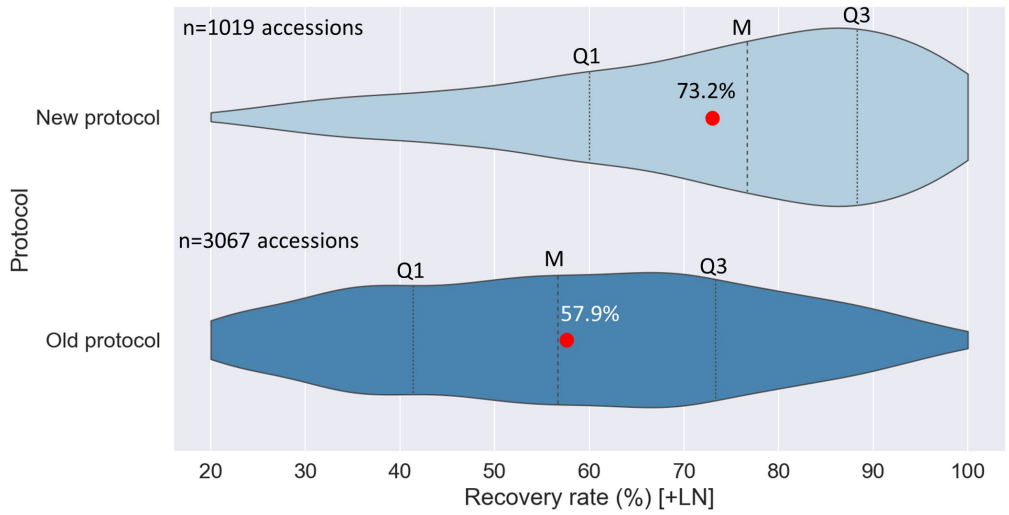


FIGURE 4
Violin plots of the distribution of recovery rate ranges for potato accessions cryopreserved with the old (3067 acc.) and new cryopreservation protocol (1019 acc.) [60 days after thawing]. Formerly, thawed/rewarmed potato shoot tips were recovered on culture media with a stepwise decrease in sucrose concentration, i.e., three days on 0.3 M sucrose-rich medium, three days on 0.2 M, three days on 0.1 M, all in darkness, then standard sucrose concentration of 0.07 M and normal light conditions. While for the new protocol, thawed/rewarmed shoot tips are placed for 9 days on culture medium with standard sucrose concentration (0.07 M; in darkness), and then incubated under normal light conditions. Q1: First quarter; Q3: Third quarter; M: median value; red circle: mean value.

needed to reduce the osmotic shock coming off the high osmotic levels in PVS2 may not be needed in all plants and indeed, could be detrimental to survival as evidenced in potato. These data are important as they further confirm the importance of the need for further investigation and refinement of post-thaw treatments for plant cryopreservation success.

It would be interesting to apply similar post-thaw experiments to other important clonal crop collections, such as cassava, yam and sweetpotato, as little information is available regarding collection-wide responses to cryopreservation. A critical point is that a relatively modest sampling with a sample size of 1.5% of the total collection (> 72 acc.), yielded results that reliably could be extrapolated to the entire collection [Figure 4]. This could be critical for development of robust operational cryopreservation protocols for other large clonal collections such as the cassava collections at The International Center for Tropical Agriculture (CIAT) and The International Institute of Tropical Agriculture (IITA) with 6155 and 4359 cassava accessions, respectively (CIAT, 2022; IITA, 2022), sweetpotato (5054 acc.) at CIP and yam (5839 acc.) at IITA (Genebank Platform, 2022). Although crop specific responses to cryopreservation are expected, the results presented here with reduced sucrose in the recovery medium could benefit these other collections as well.

Large cryopreserved plant genetic resources collections, have been established in only a small number of organizations around the world, e.g. the National Center for Genetic Resources Preservation (US, Fort Collins, CO), International Musa Germplasm Transit Centre (Europe, Belgium, Leuven), Leibniz Institute of Plant Genetics and Crop Plant Research (IPK, Europe, Germany, Gatersleben), International Potato Center (South America, Peru, Lima), National Institute of Agrobiological Sciences (Asia, Japan, Tsukuba), and National Agrobiodiversity Center (Asia, South Korea, Suwon) [Jenderek and Reed, 2017; Panis et al., 2020; Vollmer et al., 2021; Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), 2022]. The second-largest potato cryobank after CIP is located at IPK (1845 accessions).

Further, improvements in the post-thawing steps of the cryopreservation protocol, should increase the recovery rate of those potato accessions that were previously introduced into the cryobank using the higher sucrose recovery media which was the standard at the time. Accessions belonging to the *S. xcurtilobum* and *S. xjuzepczukii* taxa seem to prefer a modified recovery cycle with a higher sucrose concentration (Figure 3B; Table 3). Specially those cryobanks that already have a significant number of accessions cryopreserved, could prioritize research on improvements of the recovery cycle. This will not only improve the general recovery potential of its cryo-collection, but also reduce the number of cryovials required to be removed in future thawing events (e.g. viability monitoring, distribution).

Modifications of other post-thaw variables, like oxidative stress, growth regulators (auxins, cytokinins), and

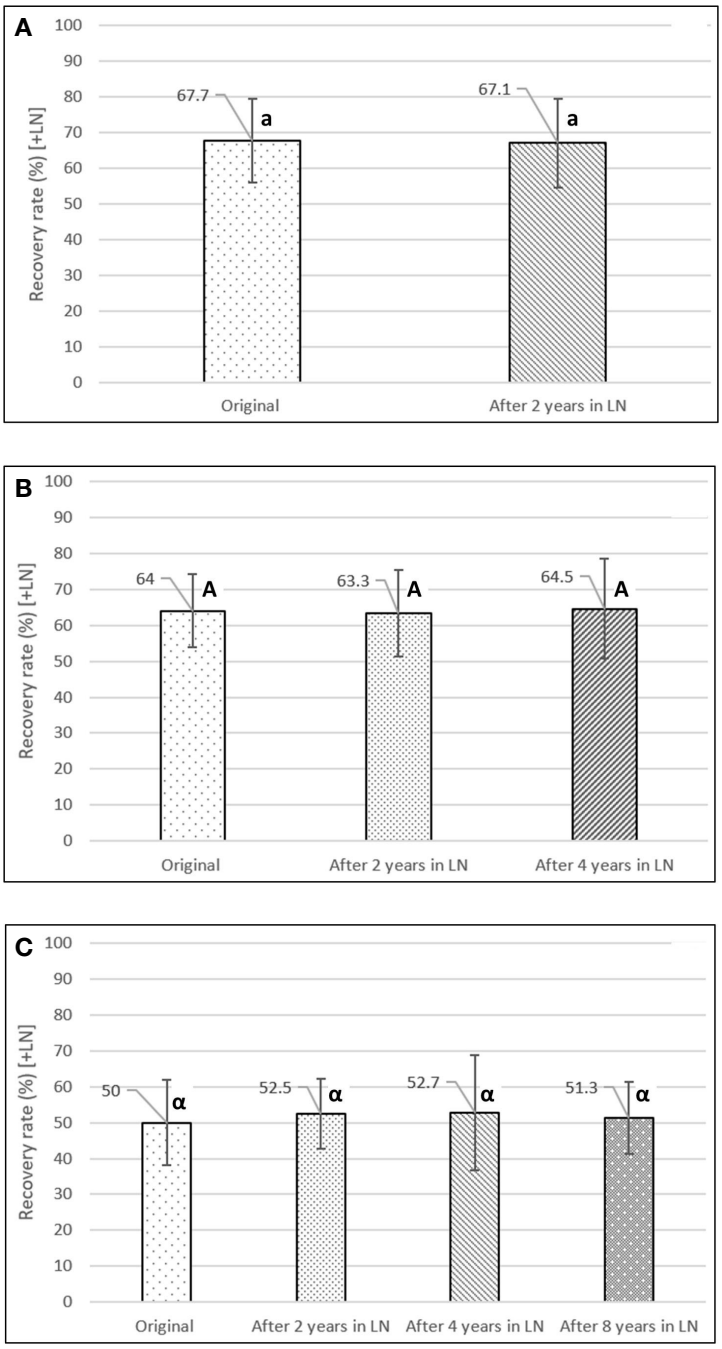


FIGURE 5
Long-term viability experiment of cryopreserved potato accessions. Full-plant recovery rates were assessed and compared after 0 (original), 2, 4, and 8 years in LN. **(A)** After two years in LN (76 accessions), **(B)** after four years in LN (52 accessions), **(C)** after eight years in LN (17 accessions). Different lower-, upper-case, and Greek letters indicate significant differences for the paired t-test, at a confidence level of 95%, for a hypothesized mean difference of less than 10% between years. Error bars: standard deviation.

environmental conditions have the potential to further increase the recovery rate of cryobanked material. Independent variables resulting in a low significance level for the interaction between genotype x treatment could be prioritized in larger-scale applications. The strategy employed in the present study where randomized and replicated experiments were done with a higher number of accessions and followed up during routine cryopreservation provides a very workable solution for large

germplam collections (Figures 2 , 4). As a side effect, valuable evaluation data for a large and diverse number of genotypes are obtained, which can be used for genetic (Ellis et al., 2018), physiological, or other studies.

Additionally, the transitory results of ongoing long-term viability monitoring experiments show that the post-thaw viability rate of cryopreserved potato accessions is stable after two, four, and eight years in liquid nitrogen demonstrating the effectiveness and security of cryopreserving clonal accessions.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

RVo and DE designed and directed the study. RVo managed the data analysis, prepared figures and tables and drafted the manuscript. RVi, MC, SP, and JC performed the cryopreservation experiments, long-term viability monitoring study, and routine cryopreservation work, and reviewed and edited the manuscript. JE assisted with drafting and editing the manuscript. DE, NA, and VR supervised, reviewed and edited the manuscript. All authors contributed to the article and approved the submitted version.

Acknowledgments

The authors express thanks for the financial support to the German Agency for International Cooperation (GIZ), the Global Crop Diversity Trust (GCDT) and the CGIAR Cooperative Research Program for Genebanks without their continuous support this work would not have been possible. We also acknowledge the staff in the cryobank for their hard work and dedication to germplasm conservation without which none of this would be possible.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Anglin, N. L., Robles, R., Rossel, G., Alagon, R., Panta, A., Jarret, R. L., et al. (2021). Genetic identity, diversity, and population structure of CIP's sweetpotato (I. batatas) germplasm collection. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.660012

Castillo, N. R. F., Bassil, N. V., Wada, S., and Reed, B. M. (2010). Genetic stability of cryopreserved shoot tips of rubus germplasm. *In Vitro Cell. Dev. Biol.-Plant* 46, 246–256. doi: 10.1007/s11627-009-9265-z

Devaux, A., Kromann, P., and Ortiz, O. (2014). Potatoes for sustainable global food security. *Potato Res.* 57:3, 185–199. doi: 10.1007/s11540-014-9265-1

Ellis, D., Chavez, O., Coombs, J., Soto, J., Gomez, R., Douches, D., et al. (2018). Genetic identity in genebanks: application of the SolCap 12K SNP array in fingerprinting the global in trust potato collection. *Genome* 61, 523–537. doi: 10.1139/gen-2017-0201

Food and Agriculture Organization (FAO) (2014). *Genebank standards for plant genetic resources for food and agriculture. rev. ed* (Rome: FAO). Available at: http://www.fao.org/3/i3704e/i3704e.pdf

Food and Agriculture Organization (FAO) (2021) *International expert panel - cryopreservation: A long-term strategy for hard-to-conserve PGRFA collections in a post-COVID*. Available at: https://www.fao.org/plant-treaty/meetings/meetings-detail/en/c/1492833/ (Accessed September 10, 2022).

Food and Agriculture Organization (FAO) (2022) *FAOSTAT – value of agricultural production (years 2018-2020)*. Available at: https://www.fao.org/faostat/en/#data/QV (Accessed September 10, 2022).

Genebank Platform (2022) *International institute of tropical agriculture (IITA)*. Available at: https://www.genebanks.org/genebanks/iita/ (Accessed November 8, 2022).

Goldschein, E. (2011). "The 10 most important crops in the world," in *Business insider*. (Businessinsider) Available at: https://www.businessinsider.com/10-crops-that-feed-the-world-2011-9.

González-Arnao, M. T., Cruz-Cruz, C. A., Hernández-Ramírez, F., Alejandre-Rosas, J. A., and Hernández-Romero, A. C. (2022). Assessment of vegetative growth and genetic integrity of vanilla planifolia regenerants after cryopreservation. . *Plants* 11, 1630. doi: 10.3390/plants11131630

Hawkes, J. G. (1990). *The potato: evolution, biodiversity and genetic resources* (London: Belhaven Press), 259p.

International Center for Tropical Agriculture (CIAT) (2022) *Cassava*. Available at: https://ciat.cgiar.org/what-we-do/breeding-better-crops/rooting-for-cassava/ (Accessed November 8, 2022).

International Institute of Tropical Agriculture (IITA) (2022) *Our genetic resources*. Available at: https://www.iita.org/research/genetic-resources/ (Accessed November 8, 2022).

Jenderek, M. M., and Reed, B. M. (2017). Cryopreserved storage of clonal germplasm in the USDA national plant germplasm system. *In Vitro Cell.Dev.Biol.-Plant* 53, 299–308. doi: 10.1007/s11627-017-9828-3

Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) (2022) *Cryo and stress biology*. Available at: https://www.ipk-gatersleben.de/en/research/genebank/cryo-and-stress-biology (Accessed September 10, 2022).

Li, J.-W., Chen, H.-Y., Li, X.-Y., Zhang, Z., Blystad, D.-R., and Wang, Q.-C. (2017). Cryopreservation and evaluations of vegetative growth, microtuber production and genetic stability in regenerants of purple-fleshed potato. *Plant Cell Tissue Org. Cult.* 128, 641–653. doi: 10.1007/s11240-016-1142-y

Merhy, T. S. M., Vianna, M. G., Garcia, R. O., Pacheco, G., and Mansur, E. (2014). Cryopreservation of *Passiflora pohlii* nodal segments and assessment of gentic stability of regenerated plants. *CryoLetters* 35, 204–215 doi: 10.14295/rbho.v17i1.721.

Panis, B., Nagel, M., and Van den houwe, I. (2020). Challenges and prospects for the conservation of crop genetic resources in field genebanks, in *In vitro* collections and/or in liquid nitrogen. *Plants* 9, 12. doi: 10.3390/plants9121634

Pence, V. C., Philpott, M., Culley, T. M., Plair, B., Yorke, S. R., Lindsey, K., et al. (2017). Survival and genetic stability of shoot tips of *Hedeoma todsenii* R.S. Irving

after long-term cryostorage. *In Vitro Cell. Dev. Biol.-Plant* 53, 28–338. doi: 10.1007/s11627-017-9854-1

Reed, B. M. (2008). "Cryopreservation–practical considerations," in *Plant cryopreservation: A practical guide*. Ed. B. M. Reed (New York: Springer). doi: 10.1007/978-0-387-72276-4_1

Velleman, P. F., and Hoaglin, D. C. (1981). *Applications, basics, and computing of exploratory data analysis* (Boston, Massachusetts, US: Duxbury Press), 354 p Available at: https://ecommons.cornell.edu/bitstream/handle/1813/78/A-B-C_of_EDA_040127.pdf.

Volk, G. M., Waddell, J., Bonnart, R., Towill, L., Ellis, D., and Luffman, M. (2008). High viability of dormant *Malus* buds after 10 years of storage in liquid nitrogen vapor. *CryoLetters* 29, 89–94.

Vollmer, R., Espirilla, J., Villagaray, R., Cárdenas, J., Castro, M., Sánchez, J. C., et al. (2021). "Cryopreservation of potato shoot tips for long-term storage," in *Solanum tuberosum. methods in molecular biology, vol 2354*. Eds. D. Dobnik, K. Gruden, Ž Ramšak and A. Coll (Humana, New York: Springer Press). doi: 10.1007/978-1-0716-1609-3_2

Vollmer, R., Villagaray, R., Cárdenas, J., Castro, M., Chávez, O., Anglin, N. L., et al. (2017). A large-scale viability assessment of the potato cryobank at the international potato center (CIP). *In Vitro Cell.Dev.Biol.-Plant* 53, 309–317. doi: 10.1007/s11627-017-9846-1

Vollmer, R., Villagaray, R., Castro, M., Anglin, N. L., and Ellis, D. (2018). Cryopreserved potato shoot tips showed genotype-specific response to sucrose concentration in rewarming solution (RS). *Plant Cell Tissue Organ Culture (PCTOC)* 136, 353–363. doi: 10.1007/s11240-018-1520-8

Vollmer, R., Villagaray, R., Egúsquiza, V., Espirilla, J., García, M., Torres, A., et al. (2016). The potato cryobank at the international potato center (CIP): A model for long term conservation of clonal plant genetic resources collections of the future. *CryoLetters* 37 (5), 318–329.

Zaheer, K., and Akhtar, M. H. (2016). Potato production, usage and nutrition – a review. *Crit. Rev. Food Sci. Nutr.* 56, 711–721. doi: 10.1080/10408398.2012.724479

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 10 January 2023
DOI 10.3389/fpls.2022.1045894

Check for updates

OPEN ACCESS

EDITED BY
Alfonso Del Rio,
University of Wisconsin-Madison,
United States

REVIEWED BY
Cinzia Montemurro,
University of Bari Aldo Moro, Italy
Joseph Kuhl,
University of Idaho, United States

*CORRESPONDENCE
Anita Behn
anita.behn@uach.cl
Derie Fuentes
derie.fuentes@unab.cl

SPECIALTY SECTION
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

RECEIVED 16 September 2022
ACCEPTED 16 December 2022
PUBLISHED 10 January 2023

CITATION
Behn A, Lizana C, Zapata F, Gonzalez A, Reyes-Díaz M and Fuentes D (2023) Phenolic and anthocyanin content characterization related to genetic diversity analysis of *Solanum tuberosum* subsp. *tuberosum* Chilotanum Group in southern Chile. *Front. Plant Sci.* 13:1045894. doi: 10.3389/fpls.2022.1045894

COPYRIGHT
© 2023 Behn, Lizana, Zapata, Gonzalez, Reyes-Díaz and Fuentes. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Phenolic and anthocyanin content characterization related to genetic diversity analysis of *Solanum tuberosum* subsp. *tuberosum* Chilotanum Group in southern Chile

Anita Behn[1]*, Carolina Lizana[1], Felipe Zapata[2], Alvaro Gonzalez[2], Marjorie Reyes-Díaz[3] and Derie Fuentes[2,4]*

[1]Instituto de Producción y Sanidad Vegetal, Facultad de Ciencias Agrarias y Alimentarias, Universidad Austral de Chile, Valdivia, Chile, [2]Biocomputing and Applied Genetics, Center for Systems Biotechnology, Fraunhofer Chile Research Foundation, Santiago, Chile, [3]Departamento de Ciencias Químicas y Recursos Naturales, Facultad de Ingeniería y Ciencias, Universidad de La Frontera, Temuco, Chile, [4]Centro de Biotecnología de Sistemas, Facultad de Ciencias de la Vida, Universidad Andres Bello, Santiago, Chile

The potato (*Solanum tuberosum* L) is one of the four most important crops worldwide in production and consumption. It originated from South America along the Andes, where six hotspots of diversity known as subcenters of origin are described from Venezuela to Chiloe Island in Chile, and where the greatest diversity of potatoes in the world is found. Today, the use of ancestral genetic resources has gained significant relevance, recovering and producing foods with a greater nutrient content and beneficial to human health. Therefore, native potatoes possess a set of characteristics with great potential for use in potato breeding guided primarily to produce better feed, especially potatoes of the *Chilotanum* Group that are easily crossed with conventional varieties. The primary objective of this study was to evaluate 290 accessions of *S. tuberosum* subsp *tuberosum* belonging to the Chilotanum Group using a set of molecular markers and correlate them to its phenotypic traits for future use in breeding programs. For this purpose, 290 accessions were analysed through 22 specific microsatellites described previously, correlating them with flesh and skin colour, total phenolic content, and anthocyanin content. A division into groups considering all the 290 accessions resulted in two clusters using

STRUCTURE analysis and seven different genetic clusters using UPGMA. The latter exhibited common phenotypic characteristics as well as anthocyanin content, strongly supporting a correlation between phenotypic traits and the genetic fingerprint. These results will enable breeders to focus on the development of potatoes with high polyphenol and anthocyanin content.

KEYWORDS

SSR, native potatoes, anthocyanin, diversity analysis, total phenolic content, *Solanum tuberosum*, Chilotanum Group

Introduction

The potato (*Solanum tuberosum*) is currently the fourth most important crop after rice, wheat, and corn. It plays a relevant role in global nutrition and for the sustainable development goals of the United Nations related to zero hunger and feed security (FAO, 2008; Ahmadu et al., 2021; FAOSTAT, 2022). The crop has become increasingly relevant due to population growth and the need to increase the supply of healthy food especially in developing regions such as Africa and Latin America. However, in developed countries, an important global trend in recent years is related to achieving a healthier diet for the population that allows curbing the deleterious effects of obesity and other metabolic diseases through the use and selection of food enriched with beneficial molecules such as antioxidants and anti-inflammatory agents. In this regard, the use of various ancestral genetic agro-resources has become increasingly important, both for direct use and for breeding programs of commercial varieties (Zhang et al., 2017).

Potatoes are a staple food and a relevant component of daily diet. Its high nutritional value and high yield potential (Koch et al., 2019) make this crop substantial even in a food shortage. Potatoes have been considered a highly nutritious contribution to human health. The resistant starch or slowly digestible starch of potatoes exerts a healthy impact under habitual consumption, and it is complemented with its high nutritional value through antioxidants, minerals, and vitamins, including potassium, folate, magnesium, and zinc (Lizana et al., 2021). The superior satiating effect of potato compared to that of rice and pasta in a mixed meal was found to be consistent with its lower energy density (Zhang et al., 2017), exerting a positive effect also for persons conscious of healthy alimentation. Moreover, colored native potatoes are especially rich in polyphenols and anthocyanins, with antioxidant properties, in a broad range that is generally associated with tuber pigmentation (Ah-Hen et al., 2012; Valiñas et al., 2017; Inostroza-Blancheteau et al., 2018).

Pigmented potatoes generally contain antioxidant compounds as phenols (Rasheed et al., 2022) that have a molecular structure characterised by the presence of one or more phenolic rings. Phenolic compounds protect plants against biotic stress caused by herbivores, insects, and pathogens and also against abiotic stress such as tissue damages caused by excessive UV radiation and free radicals (Friedman, 2006; Pourcel et al., 2007). These compounds are synthesised in large quantities as a product of the secondary metabolism of plants (Quiñones et al., 2012). Within phenols, there are flavonoid pigments known as anthocyanins, which are present in almost all plants. Within the phenolic acids, potatoes contain the following anthocyanins: petunidin, delphinidin, malvidin, pelargonidin, and peonidin glycosides. These pigments have a suitable chemical structure to function as antioxidants, donating hydrogen ions or electrons to free radicals, displacing them with their aromatic structure (Kuskoski et al., 2004; Dudek et al., 2022). Diets rich in antioxidants are associated with lower risk of incidence of diseases such as atherosclerotic heart disease, certain cancers, and macular degeneration (Hertog et al., 1993). Therefore, the anthocyanins contained in potatoes improve the nutritional value of the tuber (Hertog et al., 1993; Khalid et al., 2020), in addition to being a source of health-promoting antioxidants (Brown, 2005; Khalid et al., 2020). Anthocyanins are efficient scavengers of free radicals possessing anti-inflammatory and antimicrobial properties (Kang and Choung, 2016) and are associated with a reduced risk of developing cardiovascular disease, osteoporosis, and diabetes (Bazzano et al., 2003; Thompson et al., 2009; Speer et al., 2020). They also possess beneficial health properties such as anticancer, hepatoprotective, and antiviral properties (Bontempo et al., 2015; Calderón-Reyes et al., 2020; Fernández and Lizana, 2020; Khalid et al., 2020; Rasheed et al., 2022). The anticancer effects of anthocyanin extract compounds help in suppressing early and advanced cell proliferation and induce apoptosis in colon cancer (Bethke and Jansky, 2008), as well as in human leukemia cells (De Masi et al., 2020). Furthermore, the anthocyanins present in purple-fleshed potatoes decrease the postprandial glycemic response (Stushnoff et al., 2008). Anthocyanins have also been demonstrated to possess significant potential in attenuating oxidative stress in the skin

caused by UVB radiation and inflammatory response (Zhu et al., 2010). To achieve the maximum health benefits in potato consumption by obtaining the maximum concentration of polyphenolic compounds, it is necessary to include the colours and the cooking method of potato cultivars. Moreover, the tuber skin as a by-product differs from other by-products because of the presence of interesting nutritional and pharmaceutical constituents (Schieber and Saldaña, 2009). The temperature related to the cooking method, as well as the presence of oxygen and light, plays an important role in maximizing the preservation of anthocyanins in the tuber for human diet (Sui et al., 2014). The tubers of Chilean potatoes also possess high concentrations of anthocyanins and total phenolic content (TPC), resulting in a wide variability of colours in the skin and flesh (Ah-Hen et al., 2012; Valiñas et al., 2017).

The International Potato Center (CIP) declared six potato hotspots of diversity along the Andes Mountains from Venezuela in the North to Chiloe Island in the South (de Haan et al., 2014). Native potatoes from the northern Andes are primarily *S. tuberosum* subsp. *andigena*, and those from southern Chile are principally the Chilotanum Group (*S. tuberosum* subsp. *tuberosum*). Chile is considered a subcenter of origin of this crop. The Chilean germplasm provides new material for researchers interested in the origin of cultivated potatoes and *S. tuberosum* in Chile (Contreras, 2008). The Potato Genebank at the Universidad Austral de Chile (UACh) possesses a collection of accessions from all over the country that exhibit significant genetic diversity, traits of agronomic and culinary interest, and different degrees of resistance to biotic and abiotic stresses (López et al., 2015). López et al. (2015) revealed the characterisation of native accessions of the Potato Genebank at the UACh in terms of pest resistance using known molecular markers for genotyping.

There has been extensive research to date on potatoes due to the narrow genetic diversity among the known cultivars and the urgency to broaden the genetic pool to enhance breeding programs. Potatoes might be found in different environments, and because of this plasticity, they exhibit a great potential to adapt to the future climatic change caused by global warming and water stress conditions. In addition, more sustainable cultivation methods must introduce more abiotic and biotic resistance genes from wild potatoes. The 22 simple sequence repeat (SSR) markers developed by Ghislain et al. (2004) may be the most used markers to evaluate the diversity of native potatoes. The SSRs were used at the CIP, GLKWS, the Potato Genebank at the UACh, and other potato research centers to characterise them genetically. Other researchers have also worked with SSR markers, such as Wang et al. (2019) who proposed the model-based structure analysis, discriminating the population into two major subgroups, which can be further subdivided into seven groups based on collection sites Berdugo-Cely et al. (2017) analysed 809 Andigenum group accessions from the CCC using 5968 SNPs to determine the genetic diversity and population structure of the Andigenum germplasm, as well as the usefulness of this collection to map qualitative traits across the potato genome, showing that the CCC can be subdivided into two major groups associated with their ploidy level, viz., Phureja (diploid) and Andigena (tetraploid). Tillault and Yevtushenko (2019) fingerprinted 20 potato varieties, including five new genotypes developed in Alberta, Canada, using 10 SSR markers selected for their high discriminatory power. In that study, STM0037, STM1016, and STM1104 were found to be the best SSR markers to detect genetic differences between potato varieties. In addition, Salimi et al. (2016) determined a number of alleles per locus, ranging from 2 (STM1049) to 9 (STM1104), respectively, with an average of 6.22 to 9 alleles per locus. UPGMA dendrogram indicated that American genotypes exhibit higher expected heterozygosity than European genotypes, concluding that SSRs are appropriate markers for evaluating genetic diversity within and among potatoes from different geographical regions. Antonova et al. (2020) found that a variety of potatoes from different countries were combined into mixed groups, probably indicating the intensive exchange of breeding material, using the set of alleles in the 14 examined SSR loci. Due to the increased number of the genotyped accessions, the resolving power of only these 14 SSR markers was not sufficient because of the low values of bootstrap coefficients. In contrast, Italian accessions were subclustered into two groups and were found to be genetically highly similar to the South American germplasm, a finding that was also supported by the morphological and chemical measurements affecting their principal qualitative traits (Palumbo et al., 2019). A large number of commercial varieties in Europe and the United States have their ancestors in the potato germplasm of southern Chile. Therefore, the aims of the present study were to genetically characterise accessions from the Potato Genebank at the UACh belonging to the Chilotanum Group (*S. tuberosum* subsp. *tuberosum*) using SSR markers, chemically in terms of TPC and anthocyanin content and phenotypically in terms of tuber coloration in the flesh and skin, and to determine the correlations between them.

Materials and methods

Germplasm management and phenotypic trait screening

The tubers of 290 accessions from the Potato Genebank at the UACh were screened in two seasons, according to the scale for potato pigmentation from the CIP characterization parameters (Gómez, 2000). The definition of primary and secondary coloration of skin and flesh of tubers was based on the percentage of each colour, where primary colour is the one

that is present in more than 50% of the skin or flesh respectively.

All accessions were grown in 2015/2016 and 2016/2017 at the Austral Experimental Station of the UACh located in Valdivia (39°47′ LS, 73°14′ W, 19 m a.s.l.). Two rows (each 4 m long and separated by 0.75 m) per accession were planted. The plant density was 5.3 plants/m^2, which were fertilised with N (100 kg/ha), P_2O_5 (180 kg/ha), and K_2O (60 kg/ha) at the planting time and N (50 kg/ha) plus K_2O (60 kg/ha) at ridging, according to previous soil analyses (data not shown). Insecticides and fungicides were used to control biotic damage, according to manufacturer's recommendations. Sprinkler irrigation was implemented to satisfy water requirements of plots. The harvest of tubers was done by hand when leaves and stem were dead (BBCH scale 97), according to each potato genotype. A total of 10 potato tubers were randomly selected from each accession for morphological characterisation, as well as for TPC and anthocyanin content determination. In total, 100 g of fresh potatoes (skin and flesh) was obtained from 10 different tubers, sliced, and lyophilized for chemical analysis two weeks after the harvest.

TPC and specific anthocyanin content measurements

Total polyphenols were determined by a colorimetric assay using the Folin–Ciocalteu reagent according to the modified method of Singleton and Rossi (1965). For extraction, 0.2 g of lyophilized powder (skin + flesh) was mixed with 2 ml of methanol (80% v/v) and vortexed for 1 min. Then, the extracts were filtered through Whatman no. 5 filter paper. For quantification, 100 µL of supernatant was mixed with 750 µL of Folin–Ciocalteu reagent for 1 min and incubated at 22°C for 90 min in a water bath. Absorbance was read at 750 nm, and the TPC was determined from the standard curve of gallic acid.

The anthocyanin content was determined as described previously (Abdel-Aal and Hucl, 1999) Samples were lyophilized, and 1 g of powder was homogenised in 25 ml of acidified ethanol (pH 1), incubated for 30 min at 50°C, and centrifuged at 5400 g for 15 min at 4°C. The absorbance of the supernatant was measured at 530 and 700 nm and results were expressed in ppm of cyanidin 3-glucoside (Cy3Glu).

Anthocyanin-specific metabolites were also quantified by HPLC-DAD in the flesh of different native potato accessions. The extracts were prepared as described by Ruhland and Day (2000) with modifications made by Ribera et al. (2010), with a flow rate of 1 ml min^{-1}. The anthocyanin compounds were measured through the content of anthocyanidins (aglycone anthocyanins) using the method proposed by Nyman and Kumpulainen (2001). Delphinidin, malvidin, petunidin, cyanidin, and peonidin were used as standards (Sigma Chemical Co. St. Louis, MO) and quantified using the same HPLC method described earlier. Signals were detected at 320 nm with the mobile phase acidified water and acetonitrile by HPLC-DAD using a Kromasil reverse phase (RP)-18 column (250 * 4.6 mm) equipped with a photodiode and a DAD (Jasco MD 2015 Plus).

DNA extraction and SSR amplification

Total DNA from the 290 accessions was obtained from leaf tissue using the PureLink Genomic DNA kit (ThermoFischer, USA) according to manufacturer's instructions. The quality of each purification product was evaluated using the QuBit 3.0 fluorometer and Infinite F200/M200 NanoQuant spectrophotometer at wavelengths of 260 and 280 nm. Then, these purified DNA samples were used for allele amplification in a volume of 25 µL using the KAPA2G Robust HotStart ReadyMix PCR kit.

The sequence of primers for the 22 SSRs were obtained from a previous study (Ghislain et al., 2004). PCR was performed according to the following program: initial denaturation for 3 min at 95°C, 35 cycles of 15 s at 95°C, 30 s annealing temperature, 15 s at 72°C, and final extension of 3 min at 72° C. The final amplification product was detected by capillary electrophoresis using the Fragment Analyzer kit (Advance Analytical) and the small fragment resolution kit (1–500 bp) DNF-905-K1000. The size of each allele was determined using the PROSize 2.0 software.

Kinship, diversity, and population structure analyses

A population genetic analysis was performed with STRUCTURE software (Falush et al., 2003) on the 290 tetraploid potato genotypes using a Bayesian clustering analysis. A burning period of 5000 and 50000 MCMC reps were used as parameters for the length of the simulation. Four repetitions for 6 Ks were performed. Structure Harvester (Earl and von Holdt, 2012) was used for the evaluation. CLUMPP (Jakobsson and Rosenberg, 2007) was used to consolidate the results.

The kinship analysis was performed using the alleles previously described by Ghislain et al. (2004) for the 290 accessions, assigning values of 1 or 0 for the presence or absence of each allele in a binary form (CVS sheet). The PIC was calculated according to the following equation described by Nei (1973):

$$PIC = 1 - \sum \left(fi^2\right)$$

where *fi* is the frequency of the *i*th amplified allele over all individuals of the population (Nei, 1973). The PIC values range between 0 and 1.0, where values close to 1.0 represent a good discriminatory power for that locus.

Genetic similarity and cluster analyses were performed using the DendroUPGMA server (Garcia-Valle et al., 1999), and Dice similarity coefficients were calculated based on presence or absence of alleles using the UPGMA method (unweighted pair group method with arithmetic averages) (Sokal and Michener, 1958). The distance matrix based on Dice coefficients represents an agglomerative hierarchical clustering representation of mean genetic distances within each species. Each calculated matrix of Newick descriptors was built and used to construct a distance-based dendrogram using the SeaView tool (Gouy et al., 2010).

Genetic fingerprint analysis

To identify the unique minimal allelic profile representing each potato accession, a full vector representation was constructed for all the 290 accessions, containing a complete binary allelic representation for each accession (Supplementary Figure 1). To evaluate the genetic fingerprint that represents a unique combination of SSRs for each accession, an algorithm was designed to search a relational database containing all binary allelic representations and the minimum and unique combinations of SSRs. Thus, these vector profiles would allow identifying a specific accession through a unique PCR band profile.

Statistical analysis

Descriptive statistics was used to characterise the TPC and anthocyanin content of the genotypes of the Potato Genebank. All data concerning phenotype features such as the skin and flesh colour and spot distribution were grouped into seven genetic clusters. Their statistical significance was evaluated by two-way ANOVA and Tukey's *post hoc* test to determine statistical differences between the clusters and features.

Results

The 290 native potato accessions belonging to the Potato Genebank at the UACh were cultivated and harvested, and their tubers were evaluated phenotypically using the standards defined by the CIP (Gómez, 2000; CIP, Perú). The traits were evaluated in 10 tubers per accession. Within the 290 accessions, cream (41%), yellow (26%), or white (2%) was the primary flesh color of the tubers, totaling 69% of the screened collection. Pigmented tubers primarily showed purple coloration, where 21 accessions had purple flesh, 10 had pink flesh, and only 1 accession showed a slightly red pigmentation in the flesh. Altogether, they represented 10% of total accessions, whereas 46 accessions showed the same tones as a secondary colour (16%). Tubers with secondary coloration showed white, cream, or yellow as the primary color and pink or violet as the secondary colour, marbled or with circular distribution in the vascular ring, presenting different color distributions in the flesh (Figure 1). In total, 78 accessions (26%) showed some anthocyanin-related colour.

Regarding the skin colours, 164 accessions (57%) exhibited some variation in skin pigmentation from pink to blackish, and 83 accessions showed the same color range as the secondary

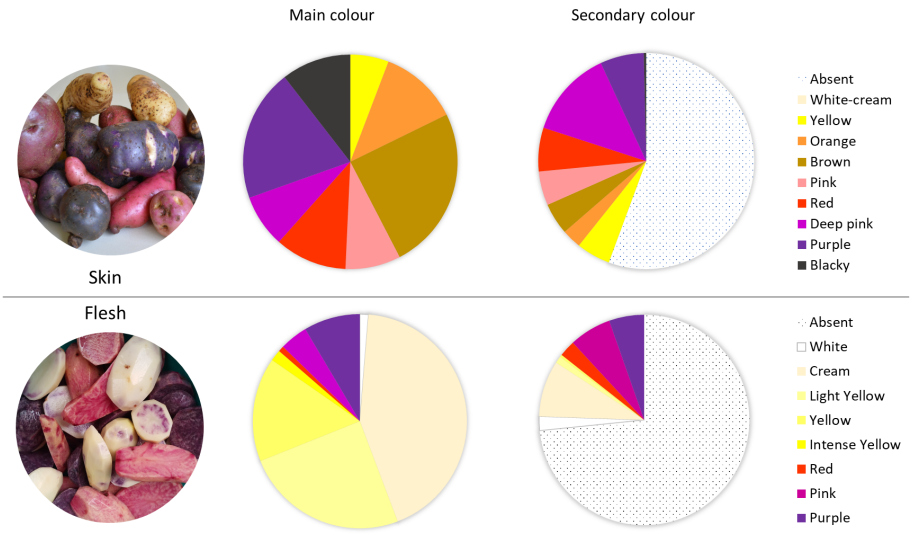


FIGURE 1
Characterization and distribution of accessions in terms of primary (main) or secondary color of the tuber skin (upper panel) and according to primary or secondary flesh color (lower panel).

colour (29%), totaling up to 85% of total accessions, revealing a greater abundance of anthocyanins and phenols in the skin than in the flesh of the screened potatoes. A total of 30 accessions (10%) showed blackish colour and 56 accessions showed purple pigmentation (19%) in the skin. Regarding the secondary colour, only 2 accessions exhibited marbled distribution with a blackish pigmentation, 19 accessions showed a purple coloration, and 30 accessions showed a pink coloration. The most frequent primary colour of the skin was yellow (24%), and 59% of the screened accessions showed no secondary colour (Figure 1).

Total phenolic content

The TPC of the accessions ranged from 1070 to 18,103 µg g^{-1} DW, with a mean of 4286 µg g^{-1} DW and a median of 3787 µg g^{-1} DW, and 39% of the genotypes had higher TPC values than the average value (Figure 2A). Moreover, 98% of the potato accessions had<10,000 µg g^{-1} DW, and only 6 accessions exceeded that value. Within the total number of accessions, there were two highly recognizable native genotypes with purple flesh and skin, whose TPC values were >15,000 µg g^{-1} DW. In general, potatoes with low TPC had white flesh and brown skin, and those with high TPC had purple flesh and blackish skin, showing an 18-fold difference in TPC between them.

To evaluate the relationship between the colour and TPC of tubers, the genotypes were grouped according to their skin and flesh colour and the average TPC per group was calculated according to each colour category (Figure 2A). In terms of flesh colour level, the analyses showed that genotypes with purple flesh as the primary colour had 1.6 times more TPC of up to 6383 µg g^{-1} DW than genotypes with pink (3937 µg g^{-1} DW) and cream (3999 µg g^{-1} DW) flesh (Figure 2A). Because of the difference in the proportion of genotypes with purple and pink marbled flesh when the accessions were compared, the tubers in those accessions with purple flesh as the secondary colour showed slightly less TPC of 5792 µg g^{-1} DW than the tubers with pink marbled flesh whose TPC reached 5386 µg g^{-1} DW.

Anthocyanin content

The anthocyanin content in the tubers showed a broader range of variation than that of TPC, ranging from 0.3 to 920 µg

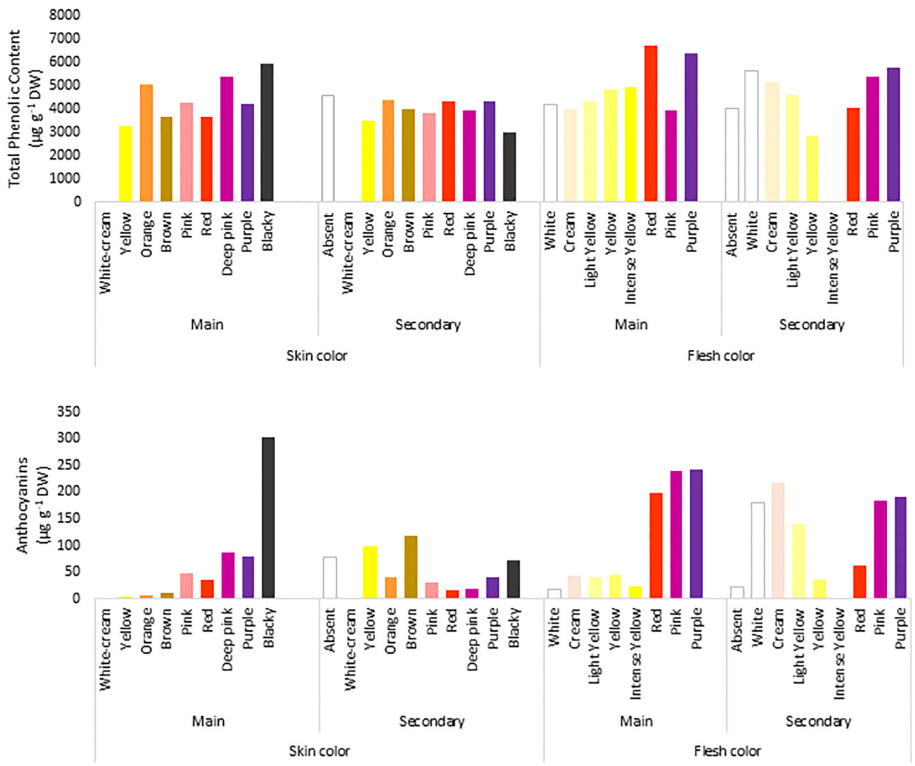


FIGURE 2
Concentration of total phenolic compounds (µg Gallic acid/g DW) determined by Folin–Ciocalteu method (upper panel) and total anthocyanin content (µg/g of cyanidin 3-glucoside) were extracted in ethanol acidified using Abdel-Aal and Hucl (1999) method (lower panel) in the 290 accessions of the Potato GenBank at the UACh.

g^{-1} DW. The average concentration was only 61.3 μg g^{-1} DW, and the median was 17.7 μg g^{-1} DW. In total, 80% of all genotypes had anthocyanin content below the average, and only 10 genotypes exceeded 400 μg g^{-1} DW. Three of them showed anthocyanin content between 600 and 700 μg g^{-1} DW, whereas one genotype showed the highest content of 920 μg g^{-1} DW. Potato tubers with less anthocyanin content had cream flesh and brown skin (0.3 μg g^{-1} DW), and those with high anthocyanin content had purple flesh and blackish skin (920 μg g^{-1} DW), showing a 3066-fold difference in the content between them.

Similar to TPC, the genotypes with purple flesh as the primary color had the highest concentration of anthocyanins, and those with pink flesh had 1.2 times more anthocyanin content than tubers with red flesh (Figure 2B). The highest concentrations of anthocyanins were found in accessions with flesh colours ranging from red to purple, which were 7-fold higher than those found in tubers with flesh color ranging white to intense yellow (Figure 2B). The TPC and anthocyanin contents were higher in accessions with higher flesh scores than in accessions with red, pink, and purple flesh colours (Figure 3).

The specific anthocyanins were analysed by HPLC-DAD in the whole of potato tubers, which revealed higher concentrations of delphinidin of up to 553 μg g^{-1} DW, followed by cyanidin (2–205 μg g^{-1} DW) and petunidin (2–236 μg g^{-1}DW), and lower concentrations of peonidin (1–18 μg g^{-1} DW) and malvidin (7–67 μg g^{-1} DW). Delphinidin was the most frequent anthocyanin found in the screened potato tubers, closely followed by cyanidin and petunidin; these anthocyanins were related to purple, magenta, and dark red flesh colours, respectively. All anthocyanins, except malvidin, showed a significant relationship ($p< 0.01$) between their specific concentrations and the total anthocyanin content of each accession (Figure 4). Specific anthocyanins were associated with different flesh colours of tubers as the primary or secondary colour (Figure 4). Higher concentrations and a more diverse profile of anthocyanins were detected in accessions with pink or purple flesh as the primary colour, supporting the relationship between the more intense colour and the higher anthocyanin content.

Regarding the anthocyanin profile obtained for each accession, there were differences between pink and purple accessions, where both the anthocyanin concentration and the proportion of predominant anthocyanin type differed (Figure 4).

Genetic analysis

Regarding the genotypic characterization, the genetic relationship among all the 290 accessions was evaluated based on the presence or absence of SSRs within the potato genomes described by Ghislain et al. (2004). The number of alleles per SSR ranged from 2 to 13 with an average of 7.36 alleles per SSR, and the PIC ranged from 0.43 to 0.91 with an average of 0.77, considering that the highest value was 1. The discriminatory power of the 22 SSR markers is represented in Table 1, which includes the PIC values calculated for each SSR. The most discriminatory SSRs corresponded to STM3012 and STM3023a with a PIC value of 0.91, followed by STM0037 (PIC value of 0.89) and the following three SSRs with a PIC value of 0.87: STM1104, STM0030, and STM1053. The largest number of alleles per SSR was observed in STM3023a, STM0037, and STM0030, and the lowest number of alleles per SSR was observed in STM1017 and STM1049 (2 and 3, respectively). STM1106 and STM2013 did not amplify any product in any reaction and were thus eliminated from further calculations.

Hierarchical structure or genotype clusters

The genetic analysis performed using STRUCTURE and Structure Harvester software with 4 repetitions for 6 Ks, showed that K=4 was the best value for K (Figure 5A).

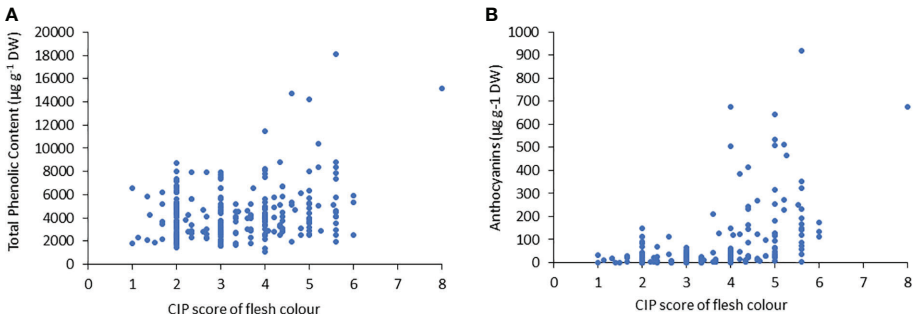


FIGURE 3
Correlation between **(A)** the total phenolic content (TPC) and **(B)** anthocyanin content, related to tuber flesh colors with the CIP score.

10.3389/fpls.2022.1045894

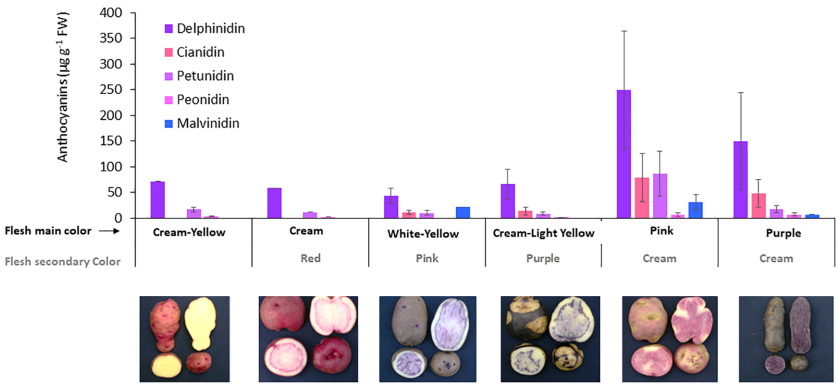


FIGURE 4
Anthocyanin profile in tubers for accession groups with different flesh color. The number of accessions for each group represented in the figure is cream-yellow = 2; cream/red = 1; white-yellow/pink = 6; cream-light yellow/purple = 4; pink/cream = 7; and purple/cream = 5. The X axis shows the combinations of primary and secondary colors in the flesh according to how accessions were grouped. Images show one accession representative of each group.

TABLE 1 SSRs, number of alleles, and PIC.

Locus	Size	Number of alleles	Alleles	Polymorphism information content (PIC)
STM1049	190–203	3	190 ± 2, 200 ± 2, 203	0.58
STM2022	188–217	4	188 ± 2, 205 ± 2, 245 ± 2, 217 ± 2	0.68
STM1053	172–184	5	172 ± 1, 175 ± 1, 178 ± 1, 181 ± 1, 184 ± 1	0.87
STM3023a	180–204	13	180+1, 182 + 1, 184 + 1, 186 + 1, 188 + 1, 190 + 1, 192 + 1, 194 + 1, 196 + 1, 198 + 1, 200 + 1, 202 + 1, 204 + 1	0.91
STM1031	271–293	4	271 ± 1, 274 ± 1, 288 ± 1 or 285 ± 1, 291 ± 1 or 293	0.73
STPoAc58	237–261	6	237 ± 2, 251 ± 1, 245 ± 1, 258 ± 1, 261 ± 1, 254 ± 1	0.74
STM0019a	146–235	8	190 ± 2, 196 ± 2 or 199, 204 ± 1 or 207 ± 2, 211 ± 1 or 213 ± 1, 232 ± 1 or 235 ± 1, 146 ± 1, 171 ± 1, 157	0.84
STM0031	172–197	4	172 ± 1, 190 ± 1, 197 ± 1, 183 ± 1	0.74
STM1052	212–267	7	214 ± 1 or 212, 222 ± 1 or 220, 230 ± 1 or 228 ± 1, 242 ± 1, 255 ± 1 or 253, 259 ± 1 or 257, 267 ± 1	0.83
STM1104	168–188	8	168 ± 1, 171 ± 1, 174 ± 1, 179 ± 1, 182 ± 1, 188 ± 1, 177 + 1, 185 ± 1	0.87
STM1016	235–267	9	247 ± 1, 252 ± 2, 257 ± 1, 260 ± 1, 244 ± 1, 235 ± 2, 267 ± 2, 263 ± 1, 242	0.84
STGBSS	125–145	7	125 ± 2, 131 ± 1, 134 ± 1, 137 ± 1, 140 or 142, 145 ± 1, 128 or 129	0.83
STWAX-2	218–249	9	218 ± 1, 224 ± 1 or 222, 228 ± 1, 231 ± 1, 234 ± 1, 239 ± 1 or 237, 243 ± 1 or 241, 243 ± 1 or 241, 214 ± 1, 249	0.81

(Continued)

10.3389/fpls.2022.1045894

TABLE 1 Continued

Locus	Size	Number of alleles	Alleles	Polymorphism information content (PIC)
STM3012	168–215	8	168 ± 1, 172 ± 2, 199 ± 1, 202 ± 1, 205 ± 1, 215 ± 1, 212 ± 1, 196 ± 1	0.91
STM0037	76–96	10	78+1, 80 + 1, 84 + 1, 86 + 1, 88, 91 + 1, 93 + 1, 96, 82 + 1, 76 + 1	0.89
STM0030	111–164	11	111, 119 + 1, 124, 135, 139 + 1, 141, 143, 146 + 1, 157, 160, 164	0.87
STM2030	180–270	4	180 ± 3, 210 ± 3, 248 ± 3, 270 ± 3	0.66
STM1064	194–201	4	194 ± 1, 197 ± 1, 199 + 1, 201 + 1	0.65
STM1058	114–129	5	120 ± 1, 123 ± 1, 126 ± 1, 129 ± 1, 114 ± 1	0.78
STM1017	135–139	2	135 ± 2, 139 ± 1	0.43

CLUMPP was used to consolidate the results of the 4 repetitions for K=4 and the result was plotted into a bar plot for the STRUCTURE simulation with K=4 subgroups, showing that genotypes formed two different clusters (Figure 5B), where cluster 1 comprises 195 accessions, while cluster 2 contains 95 accessions.

A distance matrix based on Dice coefficients was generated using the 290 native potato accessions in terms of SSR binary fingerprints. Neighbor joining clustering analyses separated them into seven hierarchical groups (clusters), which were arbitrarily named from A to G, and established the genetic kinship or relationship between all of them (Figure 6A). The cophenetic correlation coefficient for each cluster ranged from 0.66 to 0.82, indicating the usefulness of the clustering method, where values close to 1 represented a perfect match (Figure 6B). As shown in Figure 7C, clusters B and F had the highest number of genotypes of 64 and 54, respectively. Supplementary Table 1 presents thedetailed information about the accessions belonging to each cluster.

Using the kinship grouping among the 290 accessions and considering that a major objective was to identify possible genetic markers that differentiate potato accessions with higher TPC and anthocyanin content, we selected those accessions whose TPC and anthocyanin content were above the mean value. Figure 7 shows the histograms of the TPC and anthocyanin content of all the analyzed accessions, including their mean values (anthocyanins = 61.3 μg g^{-1} DW and TPC = 4286 μg g^{-1} DW). Above each histogram, the genetic clusters are shown, considering the accessions with values above the mean in each case. Thus, 58 accessions showed a higher value than the

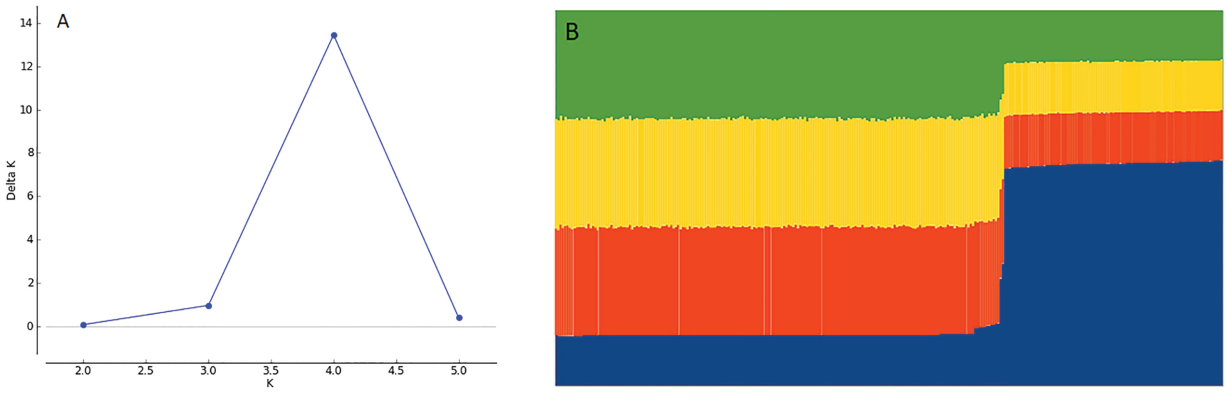


FIGURE 5
(A) Plot for Delta K at different K values. Higher value at K = 4 with Delta K = 14. **(B)** Bar plot of STRUCTURE assuming K = 4 subpopulations showing 2 clusters of genotypes.

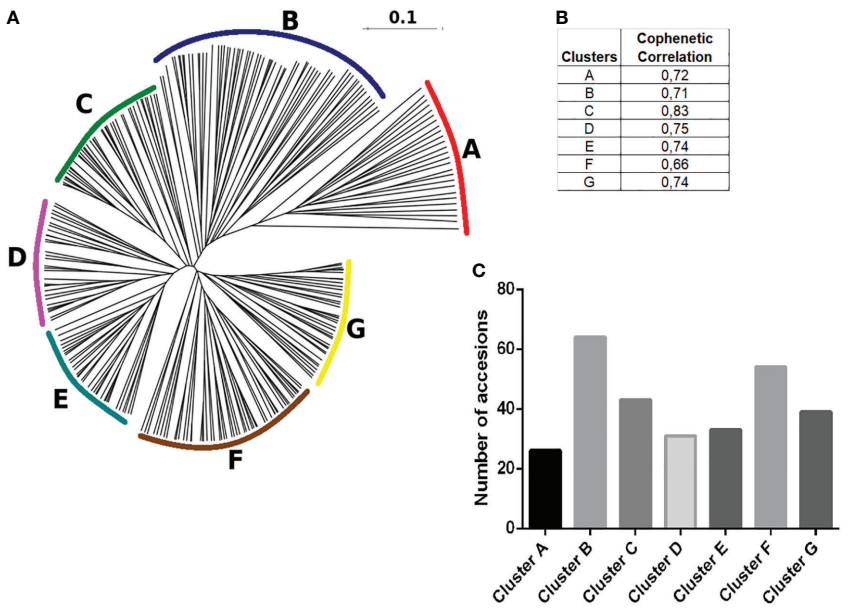


FIGURE 6
Hierarchical clustering of varieties. **(A)** The DendroUPGMA server was used to calculate the Dice similarity matrix coefficients between accessions using the UPGMA method to calculate an agglomerative hierarchical clustering representation of mean genetic relatedness within each species represented in a dendrogram with regard to allelic profiles contained within the 22 SSRs. **(B)** Cophenetic correlation coefficients of UPGMA clusters, values close to 1 represent the usefulness of the clustering method, and values between 0.6 and 1 are those with the highest correlation. **(C)** Total number of accessions distributed in the different genetic subgroups determined in this study.

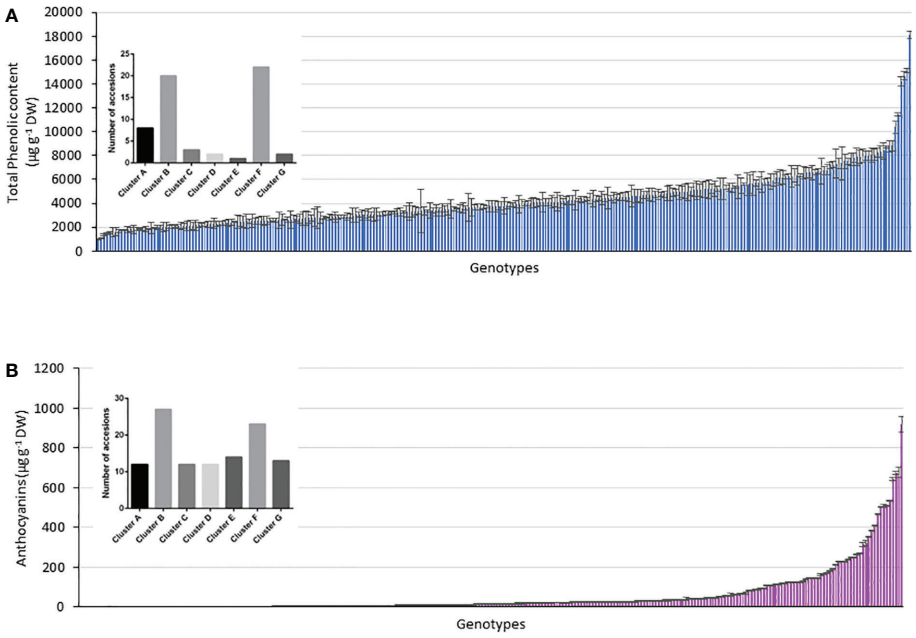


FIGURE 7
Relationship between genotype and phenotype within the accessions considering the genotypes above the respective mean values in **(A)** TPC, **(B)** anthocyanin content and the 290 accessions assigned to clusters A–G.

mean anthocyanin content (61.4 μg g^{-1} DW), which were distributed as follows among the clusters: 8 accessions in cluster A, 20 accessions in cluster B, 3 accessions in cluster C, 2 accessions in cluster D, 1 accession in cluster E, 22 accessions in cluster F, and 2 accessions in cluster G. However, 113 accessions showed values of >4286 μg g^{-1} DW (mean TPC of the population), which were distributed as follows: 12 accessions in cluster A, 27 accessions in cluster B, 12 accessions in cluster C, 12 accessions in cluster D, 14 accessions in cluster E, 23 accessions in cluster F, and 13 accessions in cluster G (Figures 7A, B). The black bars in Figure 7 indicate those clusters with a significant difference in the accessions with high content ($p< 0.05$). In this manner, clusters B and F included the largest number of accessions with high TPC and anthocyanin content (Figures 7A, B). Interestingly, cluster C included accessions with high TPC and anthocyanin content. However, clusters E and D agglomerated accessions with low anthocyanin content (Figure 7B).

Based on the same correlation analysis, we determined the correlation between the tuber genotype and phenotype, considering the primary and secondary colours of the skin and flesh. These colours were classified into nine categories (white cream, yellow, orange, brown, pink, red, deep pink, purple, and blackish) for skin and eight categories (white, cream, light yellow, yellow, intense yellow, red, pink, and purple) for flesh, according to an adapted classification proposed by the CIP, Perú. Considering all groups, the occurrence of these phenotypic traits was evaluated, and the probability of the occurrence of any of these traits in each cluster was quantified. Furthermore, the statistical significance weight for each probability was evaluated to distinguish it as a characteristic trait of the cluster. In relation to the primary skin color, there were no strong correlation between characteristics and clusters (Supplementary Table 2). However, for the secondary colour, cluster F showed a high probability of grouping accessions with secondary deep pink colour (Supplementary Table 2, $p< 0.05$). Regarding flesh colour, clusters B, C and F showed a high probability to agglomerate accessions with cream, and cluster G the accessions with light yellow flesh ($p< 0.05$; Supplementary Table 2). Furthermore, regarding the secondary flesh colour, cluster A showed a high probability to agglomerate accessions with pink flesh, ($p< 0.05$), cluster G agglomerate red flesh potatoes and cluster F showed a high probability to agglomerate accessions with cream flesh ($p< 0.05$; Supplementary Table 2).

Genetic fingerprints

To determine a strategy for differentiating potato accessions by genetic fingerprinting, it is essential to calculate the minimal number of SSR markers required to identify with a unique PCR pattern in one potato accession. Considering the phenotype in the cluster analysis, 113 accessions with TPC above the mean value and 58 accessions with anthocyanin content above the mean value were evaluated. Clusters B and F contained the maximum number of accessions from the total number of accessions, and within the groups, cluster B contained 47% and cluster F contained 59% of accessions with enhanced TPC and anthocyanin content, respectively.

Regarding the number of SSRs present in the seven clusters, STM1016 that was present in all clusters (A–G) and STM3023a, STM1104, STM1016, and STGBSS were polymorphic in a large number of genotypes in clusters A, B, and F. These clusters are important because they have high potential to include genotypes with high levels of anthocyanins and pink-like potato genotypes. These findings suggest that the presence of these SSRs is important in the expression of these phenotypes and should be considered in further investigations.

Discussion

Due to the narrow genetic diversity among the known potato cultivars and the need to broaden the genetic pool to enhance breeding programs, there has been extensive research on potatoes. A growing interest in functional foods has stimulated investigation on pigmented potatoes for their potential effects on human health related to their abundance of phenolic compounds and anthocyanins, as well as a plant protection effect. The biological activities of polyphenols in potatoes might be helpful for breeders in designing new varieties with numerous health benefits, for both pharmaceutical and nutraceutical industries (Rasheed et al., 2022). The inclusion of anthocyanins as a characteristic target in breeding programs can ensure the development of cultivars to satisfy the nutritional requirements in human consumption in the developing world (Mattoo et al., 2022).

The high genetic and phenotypic diversity within the Potato Genebank at the UACh contains the native potato diversity of the subcenter of origin of the *S. tuberosum* subsp. *tuberosum* Chilotanum Group (Spooner et al., 2007). Contreras and Castro (2008) compiled the major source of phenotypic data on the number and diversity of native varieties by UPOV characterization. The present study has complemented the phenotypic information by exploring the diversity in colors, phenols, and anthocyanins, as well as their interactions, and their genetic kinship relationship using molecular markers (SSR), with the aim to elucidate key components required for increased content of nutritional components for human feed and contribute to global food security.

Flesh-colored potatoes can represent an additional source of bioactive compounds, particularly acylated anthocyanins, in the human diet (Bellumori et al., 2017). Anthocyanins play a vital role because of their potential health benefits, and therefore, elucidating their biosynthesis has become a research focus, and one of the most investigated pathways in plants. The antioxidant

intake in humans is focused on TPC and anthocyanin content, which are present in higher concentrations in potato tubers with red, pink, and purple flesh. In our study, there was a positive correlation among the primary color of the flesh, TPC, and anthocyanin content. TPC showed a lower correlation with flesh pigmentation (0.23) than with anthocyanin content (0.44). However, the highest anthocyanin concentrations were found in accessions with flesh colors ranging from red to purple, which were 7-fold higher than those found in tubers with flesh colors ranging from white to intense yellow (Figure 2B). The correlation coefficient between TPC and anthocyanin content was 0.5, showing correlation between the scores of the different genotypes, which are consistent with published data in different plants (Liu, 2013; Diep et al., 2020). The measurements of TPC and anthocyanin content performed on skin and flesh together were also consistent with those reported by Ah-Hen et al. (2012) and Ru et al. (2019) who observed that TPC was much higher in coloured potato tubers than in white, cream, or yellow potato tubers. Ru et al. (2019) found 3- to 4-fold higher TPC in red or purple potatoes than in white or yellow potatoes. Ah-Hen et al. (2012) found 10 times higher TPC in a purple accession than in a white potato tuber. That is consistent with the present investigation, where 16-fold higher TPC was detected in a purple potato accession when compared with white potato tubers. The mean TPC values of all purple flesh accessions were only 1.5-fold higher than those of pink or cream flesh accessions. The anthocyanin content showed no differences between the purple and pink tubers, but it was 5.8-fold higher than that in tubers with cream flesh. Chilean native potatoes with the highest TPC and anthocyanin content possess mostly blackish skin and/or purple flesh, indicating the traditional accessions with blackish skin and dark purple flesh, and these results are consistent with those reported by Ah-Hen et al. (2012). These data are also consistent with those of other studies, which reported that colored potatoes have high concentrations of anthocyanins, reaching 1200 mg kg^{-1} FW (Ruiz et al., 2018; Ercoli et al., 2021; Alarcón et al., 2022).

Colored potatoes had higher TPC and higher antioxidant capacity than potatoes with yellow and white flesh (Ru et al., 2019). In our study, a couple of genotypes were characterized as having purple flesh, but with low TPC, as shown by the mean TPC of 6187 μg g^{-1} DW in potatoes with purple flesh, which varied between the highest content of 18,103 μg g^{-1} DW and the lowest content of 1937 μg g^{-1} DW. The same trend was observed for anthocyanin content in potatoes with purple flesh. These results were not consistent with the results reported by Ru et al. (2019).

In contrast, Ru et al. (2019) observed pelargonidin in potatoes with red flesh, as well as pelargonidin and petunidin in potatoes with purple flesh, wherein the latest presenting a higher total anthocyanin content than potatoes with red flesh. However, in the present study, purple tubers did not necessarily possess higher mean anthocyanin content than pink tubers. Moreover, purple Chilean potatoes with the highest TPC and anthocyanin content contain the highest concentration of delphinidin (829 μg g^{-1} DW) but no petunidin. However, the highest score of petunidin also corresponds to a purple tuber, concluding that both anthocyanins could provide purple color to potato tubers. When a subset of accessions of the Potato Genebank at the UACh was examined using molecular markers of three anthocyanin synthesis genes, a greater allelic diversity was observed at the P locus that is related to the synthesis of blue/purple petunidin-based anthocyanins, were alleles correlated positively with purple pigmentation in most of the accessions (Solís et al., 2022), that is consistent with the CIP scoring in tubers of the Potato Genebank, as most coloured potatoes scored corresponded to purple primary or secondary color. It should be emphasized that the presence of a higher delphinidin content in the screened potato accessions makes them an interesting source of bioactive compounds for human health. The features of delphinidin include 1) its colour that appears as a blue-reddish or purple pigment in plants, 2) stability under acidic conditions, 3) patented for several therapeutic effects, and 4) the maximum inhibitory effect on lipid peroxidation and O_2^- scavenging activity compared with other anthocyanidins (Khoo et al., 2017; Calderón-Reyes et al., 2020; Xu et al., 2020).

As the present results were based on the analysis of raw potatoes, they might change according to the cooking method, as they demonstrated negative effects on the contents of vitamin C, total phenolics, phenolic acids, and DPPH radical-scavenging activity (Fang et al., 2022). Currently, it is reported that cooked potatoes have decreased total anthocyanin concentrations by approximately 3%–59%. Despite these decreases, the potato genotypes had high levels of total phenols as well as high levels of antioxidant activity; hence, even with the drastic process of cooking, these results are remarkable contributors to the antioxidant activity of potato genotypes (Ercoli et al., 2021).

To investigate the genetics and its possible relationship with TPC and anthocyanin content, we performed the kinship analysis of the 290 genotypes and correlated the phenotypic and genotypic data. As the accessions were planted and harvested at the same time in the same field, as well as under the same growing conditions and cultivation techniques, we assumed that the variation between the accessions in terms of tuber color, TPC, and anthocyanin content can be attributed only to the genotypes.

The heterozygous autotetraploid *S. tuberosum* ssp. *tuberosum*, with four homologous sets of chromosomes (2n = 4x = 48), exhibited several heterogeneous genotypes and phenotypes. Even the selection of genotypes for crossing is primarily dependent on morphological traits, and genetic relatedness measured based on molecular markers help predict parental performance and improve heterotic effects. The 22 SSR markers developed by Ghislain et al. (2004) may be the most

used markers to evaluate the diversity of native potatoes, for example at the CIP, GLKWS, the Potato Genebank at the UACh, and other potato research centers. The approach used in this investigation based on the 22 SSR markers (Ghislain et al., 2004) enables merging the data with the named genebanks around the world. It could also be useful for the study of the belonging to different spatiotemporal groups, as done by Spanoghe et al. (2022) by examining a substantial panel of 1219 potato varieties using 35 microsatellite markers (SSR) to evaluate the genetic diversity. Ghislain et al. (2009) developed a potato genetic identification kit to differentiate 93.5% and 98.8% of 742 landraces using 24 and 51 SSR markers, respectively, indicating that SSR markers could efficiently identify potato germplasm at the genetic level. Hence, a high genotypic diversity might be found in the present study because the 290 accessions were different and part of a wide pool in the collection of *S. tuberosum* Chilotanum Group in the Potato Genebank. As the allelic diversity or the mean number of alleles per locus measures the genetic variation, a comparison was made with previous investigations on native potatoes. The mean number of alleles per locus was 7.4, similar to that observed by Muñoz et al. (2016) who reported a mean of 8.0 using only four SSRs, but it was less than that reported by Solano et al. (2013) who showed a mean of 9.16 alleles per locus using seven SSRs. In total, 162 alleles were found, including 20 SSRs, as STM1106 and STM2013 showed no amplification. The PIC of the different SSRs in the present study ranged broadly from 0.43 to 0.91. The mean value (0.77) was comparable with the PIC reported in different investigations that explored the genetic diversity of potato accessions, for example Muñoz et al. (2016) who reported PIC values ranging from 0.77 to 0.86.

While STRUCTURE analysis cluster into two groups with K=4, the hierarchical analysis (UPGMA dendrogram) showed 7 cluster. The reviewing of the accession's distribution between the two approaches showed that the vast majority (> 90%) of the accessions belonging to subgroups A, B, D and F identified through UPGMA analysis were also grouped together in the first STRUCTURE cluster, while the accessions grouped in clusters E and G were mostly (>80%) grouped in the second STRUCTURE cluster. Only the accessions belonging cluster C showed a more equal distribution (55%-45%) among both STRUCTURE clusters. These results suggest that the genetic analysis approach did not resolve the degree of relatedness between the 290 accessions, and the genetic structure in the genebank, which will require further analysis. Notwithstanding the above, the simple observation of the diversity of phenotypes represented in shapes, colours, TPC and anthocyanin content in the genebank hints the high genetic variations that had not been possible to address so far. The result of the STRUCTURE analysis (Figure 5B) showed the need of a more complex and in-depth analysis of the accessions of the Potato Genebank. Selga et al. (2022) suggested that breeding material and the studied Genebank collection were closely related, showing a low degree of population structure between the groups. That result agree with the present results, as only 4 subgroups separated in two clusters were determined on that analysis.

The distribution of accessions in 7 different subgroups (UMPGA analysis, Figure 6A), made sense when correlating the phenotypic differences between them. Duan et al. (2019) developed a UPGMA dendrogram based on 20 polymorphic SSR markers for Chinese breeding lines, indicating that all the 217 cultivars were closely related and lacked the formation of distinct clusters; in contrast to the present kinship analysis that showed the formation of groups clustering all the 290 accessions into seven different genetic clusters, which indicated a number of common phenotypic characteristics as anthocyanin contents, strongly supporting a correlation between the phenotypic traits and the genetic fingerprint, accounting for the high genetic diversity in the Potato Genebank. Considering the SSRs present in the seven clusters with high cophenetic scores, STM1016 was predominant because of its presence in all clusters (A–G), and also the markers STM3023a, STM1104, and STGBSS as they were polymorphic in a large number of genotypes in clusters A, B, and F, appearing to have an importance in the expression of these phenotypes. These clusters are important because they have high potential to accumulate genotypes with high levels of anthocyanins and pink-like potatoes. Muñoz et al. (2016) selected STM1016 among the four SSR markers used that were sufficiently informative to identify 320 different allelic phenotypes and thus 320 potential varieties. They also used STM1106, the SSR that did not amplify in the present study, which was considered to be not diagnostic for the Chilotanum group. Tillault and Yevtushenko (2019) fingerprinted 20 potato varieties, including 5 new genotypes developed in Alberta, Canada, using 10 SSR markers, wherein the number of alleles per locus ranged from two for the SSR marker STPoAc58 to six for STM0030 and STM0037. When compared with the present results, STM0030, STM0037, and STM3023a had the highest number of alleles of 11, 10, and 13 per locus, respectively. Tillault and Yevtushenko (2019) described the markers STM0037, STM1016, and STM1104 as the best SSR markers to detect genetic differences between potato varieties, which is broadly consistent with the present investigation. Another research on potatoes screened a collection of 264 Russet and non-Russet breeding clones and varieties through fingerprinting using 23 SSR markers, resulting in 142 polymorphic alleles. The number of alleles produced per SSR varied from 2 to 10, with an average of 6.2 alleles per marker. The PIC of SSRs ranged from 0.37 to 0.89 with an average of 0.77 (Bali et al., 2018). Despite the narrow-expected diversity in the investigation done by Bali et al., 2018, the scores were similar to those found in the present study, with the number of alleles varying from 2 to 13, and the PIC ranging from 0.43 to 0.91 with the same average of 0.77. However, Jian et al. (2017) detected 190 alleles on 20 SSR loci, and all the SSR alleles were polymorphic among these potato germplasms with an average of 9.5 alleles per SSR locus, ranging from 2 to 23, showing a larger number of alleles than that of the Potato Genebank.

Salimi et al. (2016) used SSRs to determine allelic diversity within and among potatoes from different geographical regions and reported numerous alleles per locus, ranging from two (STM1049) to nine (STM1104), respectively, which is consistent with the present data, where STM1049 showed three alleles and STM1104 showed eight alleles. Antonova et al. (2020) investigated the genetic diversity of potatoes using the set of alleles in the 14 examined SSR loci but indicated that with an increased number of the genotyped accessions, the resolving power of only these 14 SSR markers was not sufficient because of the low values of bootstrap coefficients. Duan et al. (2019) used 16 parental cultivars widely used in breeding to screen 138 SSR markers, where 20 were polymorphic that were used to analyze the genetic diversity of 217 potato cultivars grown in China. Based on the PIC values and the clarity of PCR amplification bands, 11 SSR markers were selected that could differentiate all the 217 cultivars. The 22 SSRs used in the present investigation were diagnostic and sufficient to evaluate genetic diversity and cluster the genotypes in groups. Wang et al. (2019) successfully discriminated the population into two major subgroups using SSRs, which can be further subdivided into seven groups based on collection sites. In the present investigation, seven groups were found, but they have not yet been correlated to the collection site.

Although SSRs are widely used to evaluate genetic diversity in potatoes throughout the world, Campos and Ortiz (2020) described that single nucleotide polymorphisms (SNPs) are increasingly used predominantly due to recent advances in genome sequencing technology, abundance of SNPs in most crop plants, reduced labor required to collect data, and price per data point. Furthermore, Parra-Galindo et al. (2021) found a genomic region in chromosome 10 that harbored SNPs with the strongest association with anthocyanin content in GWAS and underlined the existence of pleiotropic genes or anthocyanin biosynthesis clusters. Seven QTLs were identified to be involved in the genetic control of the anthocyanin content in cooked tubers. These QTLs explained from 31.3% to 44.4% of the phenotypic variance (anthocyanidin content and composition) (erratum: Parra-Galindo et al., 2020; Parra-Galindo et al., 2021). However, Campos and Ortiz (2020) mentioned that it is often challenging to identify SNP markers in polyploids such as the potato, due to separating allelic versus homologous SNPs or determining dosage in autopolyploid, both of which increase the rate of false positives. In the present study, we first decided to work with SSRs, because most analyses conducted to date on genetic diversity in potato genebanks have used microsatellites. Moreover, we achieved our primary objective of correlating the phenotypic and genotypic scores and identifying clusters B and F containing accessions with high amounts of TPC and anthocyanin content and a couple of molecular markers that well described both relevant clusters.

Conclusion

The narrow genetic diversity observed among modern potato varieties necessitates the need for investigations like this one so that potato breeders can enhance genetic diversity in parental clones. The native potatoes of the Chilotanum Group (*S. tuberosum* subsp. *tuberosum*) can be crossed with the common potato varieties and present a source of resistance to disease and stress, as well as high contents of polyphenols and anthocyanins, which make them contributors to breeding for designing new varieties with health benefits to satisfy the nutritional requirements in human consumption and for plant health.

The coloured accessions of the Potato Genebank primarily possessed purple and pink flesh as secondary colors and less than as primary colours. Within the 290 accessions, the primary color of the tubers was cream (41%), yellow (26%), or white (2%), and 10% of the total accessions showed purple, pink, or red flesh, whereas 46 accessions showed the same tone as the secondary color (16%). In total, 78 accessions had some anthocyanin-related colour, corresponding to 26% of total accessions. Regarding the skin colour, 164 accessions (57%) showed some variation in skin pigmentation. ranging from pink to blackish.

The TPC and anthocyanin content showed significant variation, observing genotypes with higher contents than some fruits known for their antioxidant capacity. The genotypes showing purple flesh as the primary color had the highest concentrations of anthocyanins; those with purple flesh and blackish skin had the highest TPC of 18,103 μg g^{-1} DW and the highest anthocyanin content of 920 μg g^{-1} DW. All anthocyanins, except malvidin, showed a significant relationship ($p< 0.01$) between their specific concentrations and the total anthocyanin content of each accession. Similarly, delphinidin was the most frequent specific anthocyanin found in the screened potato tubers, which was different from the published data on the anthocyanins of purple potatoes, indicating that it is a potentially interesting compound for further investigation.

Genetic diversity was evaluated using SSRs, resulting in 146 alleles with an average of 7.36 alleles per SSR and a mean PIC of 0.77; these data were consistent with the published information on other native genotypes. The most discriminatory SSRs corresponded to STM3012 and STM3023a with a PIC value of 0.91, and the largest number of alleles per SSR was observed in STM3023a, STM 0037, and STM 0030.

The STRUCTURE analysis using K=4, formed two different clusters, with 195 and 95 accessions, respectively. The seven clusters obtained using the UPGMA dendrogram showed high cophenetic coefficients. The correlation of genotypes with TPC and anthocyanin content greater than the mean value indicated that clusters B (47%) and F (59%) contained the largest number of genotypes with a high content in each case. Cluster F and A had a high probability of grouping accessions with secondary

deep pink colour of the skin, concluding that the clustering does present a correlation to colours, TPC, and anthocyanin. The possible SSR that differentiated the potato accessions with higher TPC and anthocyanin content was STM1016 that was present in all clusters (A–G), and also STM3023a, STM1104, STM1016, and STGBSS, the markers that were polymorphic in a large number of genotypes in clusters A, B, and F. Finally, the high variability in tuber colors, TPC, and anthocyanin content was consistent with the high genetic diversity screened using genotype-specific SSR markers, which support the use of this genetic material in potato breeding.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Author contributions

AB and CL did the phenotypic screening and chemical analysis, MR-D did the anthocyanins analyses per HPLC. FZ and DF performed the SSR amplifications on the DNA accessions and AG developed the kinship analysis and then the association between the different groups of accessions and their anthocyanin and phenol content. All authors contributed to the article and approved the submitted version.

Funding

This project was funded from JOINT RESEARCH PROJECTS 2014 (Conicyt/BmBf). Molecular markers for the generation of potatoes with enhanced anthocyanin content (MoMaPo) during 2015 - 2018 (BMBF20140035). Universidad Andres Bello (CHL01-0000-334023-90). Universidad La Frontera (GAP DI22-3019).

Acknowledgments

The authors thank Eusebio Miranda for cultivation of the Potato Genebank for tuber screening and Mariela Mora for HPLC analysis.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1045894/full#supplementary-material

References

Abdel-Aal, E. S., and Hucl, P. (1999). A rapid method for quantifying total anthocyanins in blue aleurone and purple pericarp wheats. *Cereal Chem.* 76, 350–354. doi: 10.1094/CCHEM.1999.76.3.350

Ah-Hen, K., Fuenzalida, C., Hess, S., Contreras, A., Vega-Gálvez, A., and Lemus-Mondaca, R. (2012). Antioxidant capacity and total phenolic compounds of twelve selected potato landrace clones grown in southern Chile. *Chil. J. Agric. Res.* 72, 3–9. doi: 10.4067/S0718-58392012000100001

Ahmadu, T., Abdullahi, A., and Ahmad, K. (2021). The role of crop protection in sustainable potato (Solanum tuberosum l.) production to alleviate global starvation problem: An overview. in m. yildiz and y. ozgen (Eds.), solanum tuberosum - a promising crop for starvation problem. *IntechOpen* 354. doi: 10.5772/intechopen.100058

Alarcón, S., Tereucán, G., Cornejo, P., Contreras, B., and Ruiz, A. (2022). Metabolic and antioxidant effects of inoculation with arbuscular mycorrhizal fungi in crops of flesh-coloured solanum tuberosum treated with fungicides. *J. Sci. Food Agric.* 102, 2270–2280. doi: 10.1002/jsfa.11565

Antonova, O., Klimenko, N. S., Rybakov, D. A., Fomina, N. A., Zheltova, V. V., Novikova, L., et al. (2020). SSR analysis of modern Russian potato varieties using DNA samples of nomenclatural standards. *Plant Biotechnol. Breed.* 3, 77–96. doi: 10.30901/2658-6266-2020-4-o2

Bali, S., Patel, G., Novy, R., Vining, K., Brown, C., Holm, D., et al. (2018). Evaluation of genetic diversity among russet potato clones and varieties from breeding programs across the united states. *PloS One* 13, e0201415. doi: 10.1371/journal.pone.0201415

Bazzano, L. A., Serdula, M. K., and Liu, S. (2003). Dietary intake of fruits and vegetables and risk of cardiovascular disease. *Curr. Atheroscler. Rep.* 5, 492–499. doi: 10.1007/s11883-003-0040-z

Bellumori, M., Innocenti, M., Michelozzi, M., Cerretani, L., and Mulinacci, N. (2017). Coloured-fleshed potatoes after boiling: Promising sources of known antioxidant compounds. *J. Food Compos. Anal.* 59, 1–7. doi: 10.1016/j.jfca.2017.02.004

Berdugo-Cely, J., Valbuena, R. I., Sánchez-Betancourt, E., Barrero, L. S., and Yockteng, R.. (2017). Genetic diversity and association mapping in the Colombian Central Collection of Solanum tuberosum L. Andigenum group using SNPs markers. *PloS One* 12 (3), e0173039. doi: 10.1371/journal.pone.0173039

Bethke, P. C., and Jansky, S. H. (2008). The effects of boiling and leaching on the content of potassium and other minerals in potatoes. *J. Food Sci.* 73, H80–H85. doi: 10.1111/j.1750-3841.2008.00782.x

Bontempo, P., De Masi, L., Carafa, V., Rigano, D., Scisciola, L., Iside, C., et al. (2015). Anticancer activities of anthocyanin extract from genotyped solanum tuberosum l. "Vitelotte". *J. Funct. Foods.* 19, 584–593. doi: 10.1016/j.jff.2015.09.063

Brown, C. R. (2005). Antioxidants in potato. *Am. J. Potato Res.* 82, 163–172. doi: 10.1007/BF02853654

Calderón-Reyes, C., Silva Pezoa, R., Leal, P., Ribera-Fonseca, A., Cáceres, C., Riquelme, I., et al. (2020). Anthocyanin-rich extracts of calafate (Berberis microphylla g. forst.) fruits decrease *in vitro* viability and migration of human gastric and gallbladder cancer cell lines. *J. Soil Sci. Plant Nutr.* 20, 1891–1903. doi: 10.1007/s42729-020-00260-8

Campos, H., and Ortiz, O. (2020). *The potato crop: its agricultural, nutritional and social contribution to humankind* (Springer Nature) 518 (Cap4), 128. Available at: https://link.springer.com/content/pdf/10.1007/978-3-030-28683-5.pdf?pdf=button

Contreras, A. (2008). Uso de especies silvestres y cultivadas en el mejoramiento de la papa. *Agrosur* 36, 115–129. doi: 10.4206/agrosur.2008.v36n3-01

Contreras, A., and Castro, I. (2008). *Catálogo de variedades de papas nativas de Chile* (Chile: Universidad Austral de Chile), 234. Available at: www.potatogenebank.cl.

de Haan, S., Polreich, S., Juarez, H., Rodriguez, F., Ccanto, R., Alvarez, C., et al. (2014). "The chirapaq Ñan initiative: Establishment of a long-term on-farm monitoring network for potato landrace diversity," in *Book of abstracts. international conference on enhanced genepool utilization capturing wild relative and landrace diversity for crop improvement. Cambridge (UK).* Eds. S. Dias, E. Dulloo, N. Maxted, S. Kell, E. Thorn, L. Smith, J. Preston and S. Hutchinson (Rome (Italy: Bioversity International), 28.

De Masi, L., Bontempo, P., Rigano, D., Stiuso, P., Carafa, V., Nebbioso, A., et al. (2020). Comparative phytochemical characterization, genetic profile, and antiproliferative activity of polyphenol-rich extracts from pigmented tubers of different solanum tuberosum varieties. *Molecules* 25, 233. doi: 10.3390/molecules25010233

Diep, T., Pook, C., and Yoo, M. (2020). Phenolic and anthocyanin compounds and antioxidant activity of tamarillo (Solanum betaceum cav.). *Antioxid. (Basel)* 9, 169. doi: 10.3390/antiox9020169

Duan, Y., Liu, J., Xu, J., Bian, C., Duan, S., Pang, W., et al. (2019). DNA Fingerprinting and genetic diversity analysis with simple sequence repeat markers of 217 potato cultivars (Solanum tuberosum l.) in China. *Am. J. Potato Res.* 96, 21–32. doi: 10.1007/s12230-018-9685-6

Dudek, A., Spiegel, M., Strugała-Danak, P., and Gabrielska, J. (2022). Analytical and theoretical studies of antioxidant properties of chosen anthocyanins; a structure-dependent relationships. *Int. J. Mol. Sci.* 23, 5432. doi: 10.3390/ijms23105432

Earl, D. A., and von Holdt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the evanno method. *Conserv. Genet. Resour.* 4 (2), 359–361. doi: 10.1007/s12686-011-9548-7

Ercoli, S., Parada, J., Bustamante, L., Hermosín-Gutiérrez, I., Contreras, B., Cornejo, P., et al. (2021). Noticeable quantities of functional compounds and antioxidant activities remain after cooking of colored fleshed potatoes native from southern Chile. *Molecules* 26, 314. doi: 10.3390/molecules26020314

Falush, D., Stephens, M., and Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164 (4), 1567–1587. doi: 10.1093/genetics/164.4.1567

Fang, H., Yin, X., He, J., Xin, S., Zhang, H., Ye, X., et al. (2022). Cooking methods affected the phytochemicals and antioxidant activities of potato from different varieties. *Food Chem.* 14, 100339. doi: 10.1016/j.fochx.2022.100339

FAO (2008). *Año internacional de la papa 2008: Nueva luz sobre un tesoro enterrado. reseña de fin de año*. Available at: https://www.fao.org/3/i0500s/i0500s.pdf

FAOSTAT (2022) *Producción de cultivos*. Available at: http://www.fao.org/faostat/es/#data/QC.

Fernández, R., and Lizana, X. C. (2020). Antocianinas en Solanum tuberosum: Una revisión. *Agrosur* 48 (2), 1–8. doi: 10.4206/agrosur.2020.v48n2-01

Friedman, M. (2006). Potato glycoalkaloids and metabolites: Roles in the plant and in the diet. *J. Agric. Food Chem.* 54, 8655–8681. doi: 10.1021/jf061471t

Garcia-Valle, S., Palau, J., and Romeu, A. (1999). Horizontal gene transfer in glycosyl hydrolases inferred from codon usage in escherichia coli and bacillus subtilis. *Mol. Biol. Evol.* 16, 1125–1134. doi: 10.1093/oxfordjournals.molbev.a026203

Ghislain, M., Núñez, J., Herrera, M., Pignataro, J., Guzman, F., Bonierbale, M., et al. (2009). Robust and highly informative microsatellite-based genetic identity kit for potato. *Mol. Breed.* 23, 377–388. doi: 10.1007/s11032-008-9240-0

Ghislain, M., Spooner, D. M., Rodriguez, F., Villamón, F., Núñez, J., Vásquez, C., et al. (2004). Selection of highly informative and user-friendly microsatellites (SSRs) for genotyping of cultivated potato. *Theor. Appl. Genet.* 108, 881–890. doi: 10.1007/s00122-003-1494-7

Gómez, R. (2000). *Guía para las caracterizaciones morfológicas básicas en colecciones de papas nativas. centro internacional de la papa (CIP), germoplasma de papa, dpto. de mejoramiento y recursos genéticos* (Lima, Perú: CIP).

Gouy, M., Guindon, S., and Gascuel, O. (2010). SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* 27, 221–224. doi: 10.1093/molbev/msp259

Hertog, M. G., Feskens, E. J., Kromhout, D., Hollman, P. C. H., and Katan, M. B. (1993). Dietary antioxidant flavonoids and risk of coronary heart disease: the zutphen elderly study. *Lancet* 342, 1007–1011. doi: 10.1016/0140-6736(93)92876-u

Inostroza-Blancheteau, C., de Oliveira Silva, F. M., Durán, F., Solano, J., Obata, T., Machado, M., et al. (2018). Metabolic diversity in tuber tissues of native chiloé potatoes and commercial cultivars of solanum tuberosum ssp. tuberosum l. *Metabolomics* 14, 138. doi: 10.1007/s11306-018-1428-7

Jakobsson, M., and Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23 (14), 1801–1806. doi: 10.1093/bioinformatics/btm233

Jian, W., Lu, H., Wang, R. Y., He, M. M., and Liu, Q. C. (2017). Genetic diversity and population structure of 288 potato (Solanum tuberosum l.) germplasms revealed by SSR and AFLP markers. *J. Integr. Agric.* 16, 2434–2443. doi: 10.3390/plants10040752

Kang, M. A., and Choung, S. Y. (2016). Solanum tuberosum l. cv hongyoung extract inhibits 2, 4-dinitrochlorobenzene-induced atopic dermatitis in NC/Nga mice. *Mol. Med. Rep.* 14, 3093–3103. doi: 10.3892/mmr.2016.5595

Khalid, W., Khalid, M. Z., Aziz, A., Tariq, A., Ikram, A., Rehan, M., et al. (2020). Nutritional composition and health benefits of potato: A review. *Adv. Food Nutr. Sci.* 5, 7–16. doi: 10.21065/AdvFoodNutriSci.5.7

Khoo, H. E., Azlan, A., Tang, S. T., and Lim, S. M. (2017). Anthocyanidins and anthocyanins: colored pigments as food, pharmaceutical ingredients, and the potential health benefits. *Food Nutr. Res.* 61, 1361779. doi: 10.1080/16546628.2017.1361779

Koch, M., Naumann, M., and Pawelzik, E. (2019). Cracking and fracture properties of potato (Solanum tuberosum l.) tubers and their relation to dry matter, starch, and mineral distribution. *J. Sci. Food Agric.* 99, 3149–3156. doi: 10.1002/jsfa.9530

Kuskoski, E. M., Asuero, A. G., García-Parilla, M. C., Troncoso, A. M., and Fett, R. (2004). Actividad antioxidante de pigmentos antociánicos. *Ciênc. Tecnol. Aliment.* 24, 691–693. doi: 10.1590/S0101-20612004000400036

Liu, R. (2013). Health-promoting components of fruits and vegetables in the diet. *Adv. Nutr.* 4, 384S–392S. doi: 10.3945/an.112.003517

Lizana, X., Sandaña, P., Behn, A., Ávila-Valdés, A., Ramírez, D., Soratto, R., et al. (2021). "Potato," in *Crop physiology case histories for major crops, 1st Edition*. Eds. V. Sadras and D. Calderini (London. Reino Unido: Academic Press is an imprint of Elsevier), 551–588.

López, M., Riegel, R., Lizana, X., and Behn, A. (2015). Identification of viruses and nematodes resistance genes in the chilotanum potato genebank of the universidad austral de Chile. *Chil. J. Agric. Res. (JAR)* 75, 320–327. doi: 10.4067/S0718-58392015000400008

Mattoo, A. K., Dwivedi, S. L., Dutt, S., Singh, B., Garg, M., and Ortiz, R. (2022). Anthocyanin-rich vegetables for human consumption–focus on potato, sweetpotato and tomato. *Int. J. Mol. Sci.* 23, 2634. doi: 10.3390/ijms23052634

Muñoz, M., Folch, C., Rodriguez, F., Kalazich, J., Orena, S., Santos, J., et al. (2016). Genotype number and allelic diversity overview in the national collection of Chilean potatoes. *Potato Res.* 59, 227–240. doi: 10.1007/s11540-016-9329-5

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. U. S. A.* 70, 3321–3323. doi: 10.1073/pnas.70.12.3321

Nyman, N. A., and Kumpulainen, J. T. (2001). Determination of anthocyanidins in berries and red wine by high-performance liquid chromatography. *J. Agric. Food Chem.* 49, 4183–4187. doi: 10.1021/jf010572i

Palumbo, F., Galvao, A. C., Nicoletto, C., Sambo, P., and Barcaccia, G. (2019). Diversity analysis of sweet potato genetic resources using morphological and qualitative traits and molecular markers. *Genes* 10, 840. doi: 10.3390/genes10110840

Parra-Galindo, M. A., Piñeros-Niño, C., Soto-Sedano, J.C., and Mosquera-Vasquez, T. (2020). Chromosomes I and X harbor consistent genetic factors associated with the anthocyanin variation in potato agronomy 2019, 9, 366. *Agronomy* 10, 532. doi: 10.3390/agronomy10040532

Parra-Galindo, M. A., Soto-Sedano, J. C., Mosquera-Vasquez, T., and Roda, F. (2021). Pathway based analysis of anthocyanin diversity in diploid potato. *PloS One* 16 (4), e0250861. doi: 10.1371/journal.pone.0250861

Pourcel, L., Routaboul, J. M., Cheynier, V., Lepiniec, L., and Debeaujon, I. (2007). Flavonoid oxidation in plants: from biochemical properties to physiological functions. *Trends Plant Sci.* 12, 29–36. doi: 10.1016/j.tplants.2006.11.006

Quiñones, M., Miguel, M., and Aleixandre, A. (2012). Los Polifenoles, compuestos de origen natural con efectos saludables sobre el sistema cardiovascular. *Nutr. Hosp.* 27, 76–89. doi: 10.3305/nh.2012.27.1.5418

Rasheed, H., Ahmad, D., and Bao, J. (2022). Genetic diversity and health properties of polyphenols in potato. *Antioxidants* 11, 603. doi: 10.3390/antiox11040603

Ribera, A. E., Reyes-Díaz, M., Alberdi, M., Zuñiga, G. E., and Mora, M. L. (2010). Antioxidant compounds in skin and pulp of fruits change among genotypes and maturity stages on highbush blueberry (Vaccinium corymbosum l.) grown in southern Chile. *J. Soil Sci. Plant Nutr.* 10, 509–536. doi: 10.4067/S0718-95162010000200010

Ruhland, C. T., and Day, T. A. (2000). Effects of ultraviolet-b radiation on leaf elongation, production and phenylpropanoid concentrations in deschampsia antarctica and colobanthus quitensis in Antarctica. *Physiol. Plant* 109, 244–251. doi: 10.1034/j.1399-3054.2000.100304.x

Ruiz, A., Aguilera, A., Ercoli, S., Parada, J., Winterhalter, P., Contreras, B., et al. (2018). Effect of the frying process on the composition of hydroxycinnamic acid derivatives and antioxidant activity in flesh colored potatoes. *Food Chem.* 268, 577–584. doi: 10.1016/j.foodchem.2018.06.116

Ru, W., Pang, Y., Gan, Y., Liu, Q., and Bao, J. (2019). Phenolic compounds and antioxidant activities of potato cultivars with white, yellow, red and purple flesh. *Antioxidants* 8, 419. doi: 10.3390/antiox8100419

Salimi, H., Bahar, M., Mirlohi, A., and Talebi, M. (2016). Assessment of the genetic diversity among potato cultivars from different geographical areas using the genomic and EST microsatellites. *Iran. J. Biotechnol.* 14, 270. doi: 10.15171/ijb.1280

Schieber, A., and Saldaña, M. (2009). Potato peels: A source of nutritionally and pharmacologically interesting compounds–a review. *Food* 3, 23–29. doi: 10.7939/R33T9DM0H

Selga, C., Chrominski, P., Carlson-Nilsson, U., Andersson, M., Chawade, A., and Ortiz, R. (2022). Diversity and population structure of Nordic potato cultivars and breeding clones. *BMC Plant Biol.* 22, 350. doi: 10.1186/s12870-022-03726-2

Singleton, V. L., and Rossi, J. A. (1965). Colorimetry of total phenolics with phosphomolybdic-phosphotungstic acid reagents. *Am. J. Enol. Vitic.* 16, 144–158.

Sokal, R. R., and Michener, C. D. (1958). A statistical method for evaluating systematic relationships. *Univ. Kans. Sci. Bull.* 38, 1409–1438.

Solano, J., Mathias, M., Esnault, F., and Brabant, P. (2013). Genetic diversity among native varieties and commercial cultivars of solanum tuberosum ssp. tuberosum l. present in chile. electron. *J. Biotechnol.* 16, 6. doi: 10.2225/vol16-issue6-fulltext-15

Solís, J. L., Muth, J., Canales, J., Lizana, C., Pruefer, D., Riegel, R., et al. (2022). Allelic diversity of three anthocyanin synthesis genes in accessions of native solanum tuberosum l. ssp. tuberosum at the potato genebank of the universidad austral de Chile. *Genet. Resour. Crop Evol.* 69, 297–314. doi: 10.1007/s10722-021-01230-4

Spanoghe, M., Marique, T., Nirsha, A., Esnault, F., and Lanterbecq, D. (2022). Genetic diversity trends in the cultivated potato: a spatiotemporal overview. *Biology* 11, 604. doi: 10.3390/biology11040604

Speer, H., D'Cunha, N. M., Alexopoulos, N. I., McKune, A. J., and Naumovski, N. (2020). Anthocyanins and human health–a focus on oxidative stress, inflammation and disease. *Antioxid. (Basel)* 9, 366. doi: 10.3390/antiox9050366

Spooner, D. M., Núñez, J., Trujillo, G., Del Rosario, H. M., Guzmán, F., and Ghislain, M. (2007). Extensive simple sequence repeat genotyping of potato landraces supports a major reevaluation of their gene pool structure and classification. *Proc. Natl. Acad. Sci. U.S.A.* 104, 19398–19403. doi: 10.1073/pnas.0709796104

Stushnoff, C., Holm, D., Thompson, M., Jiang, W., Thompson, H., Joyce, N., et al. (2008). Antioxidant properties of cultivars and selections from the Colorado potato breeding program. *Am. J. Potato Res.* 85, 267–276. doi: 10.1007/s12230-008-9032-4

Sui, X., Dong, X., and Zhou, W. (2014). Combined effect of pH and high temperature on the stability and antioxidant capacity of two anthocyanins in aqueous solution. *Food Chem.* 163, 163–170. doi: 10.1016/j.foodchem.2014.04.075

Thompson, M., Thompson, H., McGinley, J., Neil, E., Rush, D., and Holm, D. (2009). Functional food characteristics of potato cultivars (Solanum tuberosum l.): Phytochemical composition and inhibition of 1-methyl-1-nitrosourea induced breast cancer in rats. *J. Food Compos. Anal.* 22, 571–576. doi: 10.1016/j.jfca.2008.09.002

Tillault, A. S., and Yevtushenko, D. P. (2019). Simple sequence repeat analysis of new potato varieties developed in Alberta, Canada. *Plant Direct.* 3, e00140. doi: 10.1002/pld3.140

Valiñas, M., Lanteri, M., ten Have, A., and Andreu, A. (2017). Chlorogenic acid, anthocyanin and flavan-3-ol biosynthesis in flesh and skin of Andean potato tubers (*Solanum tuberosum* subsp. *andigena*). *Food Chem.* 229, 837–846. doi: 10.1016/j.foodchem.2017.02.150

Wang, Y., Rashid, M., Li, X., Yao, C., Lu, L., Bai, J., et al. (2019). Collection and evaluation of genetic diversity and population structure of potato landraces and varieties in China. *Front. Plant Sci.* 10, 139. doi: 10.3389/fpls.2019.00139

Xu, J., Zhang, Y., Ren, G., Yang, R., Chen, J., Xiang, X., et al. (2020). Inhibitory effect of delphinidin on oxidative stress induced by H2O2 in HepG2 cells. *Oxid. Med. Cell Longev.* 20, 4694760. doi: 10.1155/2020/4694760

Zhang, H., Mittal, N., Leamy, L. J., Barazani, O., and Song, B. H. (2017). Back into the wild–apply untapped genetic diversity of wild relatives for crop improvement. *Evol.Appl* 10, 5–24. doi: 10.1111/eva.12434

Zhu, F., Cai, Y., Ke, J., and Corke, H. (2010). Compositions of phenolic compounds, amino acids and reducing sugars in commercial potato varieties and their effects on acrylamide formation. *J. Sci. Food Agric.* 90, 2254–2262. doi: 10.1002/jsfa.4079

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 18 January 2023
DOI 10.3389/fpls.2022.1007104

Check for updates

OPEN ACCESS

EDITED BY
Manuela Nagel,
Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Germany

REVIEWED BY
Muhammad Naeem,
Government of Punjab, Pakistan
Luiz Filipe Protasio Pereira,
Embrapa Café, Brazil

*CORRESPONDENCE
Roxana Yockteng
ryockteng@agrosavia.co

SPECIALTY SECTION
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

RECEIVED 29 July 2022
ACCEPTED 08 December 2022
PUBLISHED 18 January 2023

CITATION
Berdugo-Cely JA, Céron-Lasso MdS and Yockteng R (2023) Phenotypic and molecular analyses in diploid and tetraploid genotypes of *Solanum tuberosum* L. reveal promising genotypes and candidate genes associated with phenolic compounds, ascorbic acid contents, and antioxidant activity.
Front. Plant Sci. 13:1007104.
doi: 10.3389/fpls.2022.1007104

COPYRIGHT
© 2023 Berdugo-Cely, Céron-Lasso and Yockteng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Phenotypic and molecular analyses in diploid and tetraploid genotypes of *Solanum tuberosum* L. reveal promising genotypes and candidate genes associated with phenolic compounds, ascorbic acid contents, and antioxidant activity

Jhon A. Berdugo-Cely[1,2], María del Socorro Céron-Lasso[2] and Roxana Yockteng[2,3]*

[1]Corporación Colombiana de Investigación Agropecuaria-AGROSAVIA, Centro de Investigación Turipaná, Km 13 vía Montería-Cereté, Montería, Córdoba, Colombia, [2]Corporación Colombiana de Investigación Agropecuaria (AGROSAVIA), Centro de Investigación Tibaitatá, Km 13 vía Mosquera-Bogotá, Mosquera, Cundinamarca, Colombia, [3]Institut de Systématique, Evolution, Biodiversité-UMR-CNRS 7205, National Museum of Natural History, Paris, France

Potato tubers contain biochemical compounds with antioxidant properties that benefit human health. However, the genomic basis of the production of antioxidant compounds in potatoes has largely remained unexplored. Therefore, we report the first genome-wide association study (GWAS) based on 4488 single nucleotide polymorphism (SNP) markers and the phenotypic evaluation of Total Phenols Content (TPC), Ascorbic Acid Content (AAC), and Antioxidant Activity (AA) traits in 404 diverse potato genotypes (84 diploids and 320 tetraploids) conserved at the Colombian germplasm bank that administers AGROSAVIA. The concentration of antioxidant compounds correlated to the skin tuber color and ploidy level. Especially, purple-blackish tetraploid tubers had the highest TPC (2062.41 ± 547.37 mg GAE), while diploid pink-red tubers presented the highest AA (DDPH: 14967.1 ± 4687.79 µmol TE; FRAP: 2208.63 ± 797.35 mg AAE) and AAC (4.52 mg ± 0.68 AA). The index selection allowed us to choose 20 promising genotypes with the highest values for the antioxidant compounds. Genome Association mapping identified 58 SNP-Trait Associations (STAs) with single-locus models and 28 Quantitative Trait Nucleotide (QTNs) with multi-locus models associated with the evaluated traits. Among models, eight STAs/QTNs related to TPC, AAC, and AA were detected in common, flanking seven candidate genes, from which four were pleiotropic. The combination in one population of diploid and tetraploid genotypes enabled the identification of more genetic associations. However, the GWAS analysis implemented independently

in populations detected some regions in common between diploids and tetraploids not detected in the mixed population. Candidate genes have molecular functions involved in phenolic compounds, ascorbic acid biosynthesis, and antioxidant responses concerning plant abiotic stress. All candidate genes identified in this study can be used for further expression analysis validation and future implementation in marker-assisted selection pre-breeding platforms targeting fortified materials. Our study further revealed the importance of potato germplasm conserved in national genebanks, such as AGROSAVIA's, as a valuable genetic resource to improve existing potato varieties.

KEYWORDS

antioxidants, phenolic compounds, ascorbic acid content, antioxidant activity, GWAS, potato, diploid, tetraploids

Introduction

Studies on the functional properties of phytochemicals such as polyphenols, carotenoids, and vitamins are extensive because of their beneficial effect on human health (Valcarcel et al., 2016; Fraga et al., 2019; Lourenço et al., 2019). Polyphenols and ascorbic acid (vitamin C) are known to be powerful antioxidants that may scavenge Reactive Oxygen Species (ROS) to prevent DNA damage (Murniece et al., 2014; Silva-Beltrán et al., 2017; Pawlowska et al., 2019). Plants produce polyphenols (*e.g.*, phenolic acids, stilbenes, coumarins, lignins, flavonoids, among others) as secondary metabolites with antioxidative properties, that are involved in defense against ultraviolet radiation, pathogen attack, stress responses, plant growth, development, immunity, and allelopathy. In humans, polyphenols have an effect in the cell proliferation, metabolism, weight, and chronic disease (Cory et al., 2018; Fraga et al., 2019; Lourenço et al., 2019). Ascorbic acid is an essential water-soluble micronutrient. It is one of the main redox components with properties as an enzyme cofactor in many metabolic pathways such as the biosynthesis of organic components, regulatory processes, hormone production, and signal transduction. In humans, this compound supports the immune system by acting in the activation and stabilization of the B and E vitamins and folic acid. In plants, ascorbic acid plays a role in plant development, photosynthesis regulation, cell division, and abiotic stress tolerance (Bozonet and Carr, 2019; Pawlowska et al., 2019; Liu et al., 2020).

Humans must consume polyphenols and vitamins through their diet since most cannot synthesize them (Chambial et al., 2013; Pawlowska et al., 2019). These compounds are produced by several fruits and vegetables, such as potato (*Solanum tuberosum* L.). Potatoes are the fourth most important food crop worldwide after corn, rice, and wheat, with a consumption per capita of 32.4 kg in 2019 (FAOSTAT, 2022). Potato is valuable for food security and nutrition because it contains proteins, carbohydrates, lipids, organic acids, polyphenols, glycoalkaloids, fibers, minerals, and vitamins (Singh and Kaur, 2009; Kowalczewski et al., 2019) and especially because it is by far eaten more than other vegetables (Navarre et al., 2009). A fresh red potato of 100 g without skin can contain 2.06 g of proteins, 16.4 g of carbohydrates, 13.8g of fiber, 5mg of Ca, 0.39mg of Fe and Zn, and 21.3gr of Vitamin C, among others components (USDA, 2022). Potato tubers are the third most important source of phenols after apples (*Malus domestica* Borkh.) and oranges (*Citrus×sinensis*); 50% of phenolic compounds accumulate in the potato peel and tissues nearby (Ah-Hen et al., 2012; Hesam et al., 2012; Murniece et al., 2014; Mishra et al., 2020).

The information about the genetic regulation of the polyphenol biosynthesis in potato tubers is scarce (Valiñas et al., 2015). Then, a few genes related to the production of phenolic compounds in potatoes, such as anthocyanin and flavonoids, have been identified through QTL analyses (Zhang et al., 2009), transcriptome analyses (Stushnoff et al., 2010; Liu et al., 2015; Liu et al., 2018), and Genome-Wide Association Studies (GWAS) (Parra-Galindo et al., 2019; Parra-Galindo et al., 2021). In plants, three pathways, the shikimate, phenylpropanoid, and flavonoid pathways, have a function in the biosynthesis of phenolic compounds (Patil and Masand, 2018). In particular, the genes PAL (*phenylalanine-ammonia-lyase*), C4H (*cinnamate 4-hydroxylase*), 4CL (*4-coumarate-CoA ligase*), CHS (*chalcone synthase*), HCT (*hydroxycinnamoyl-CoA shikimate*), HQT (*hydroxycinnamoyl-CoA quinate*) and transcriptions factors (TFs) as MYB and basic helixloop-helix (bHLH) play a role in the regulation of the synthesis of these compounds (Vogt, 2010; Valiñas et al., 2015; Valcarcel et al., 2016). On the other hand, the ascorbic acid is synthesized at least

by four pathways, the L-Galactose (Wheeler et al., 1998; Wheeler et al., 2003), the L-Glucose (Wheeler et al., 2003; Wolucka and Van Montagu, 2003), the D-Galacturonate (Agius et al., 2003; Wheeler et al., 2003), and the Myo-Inositol pathways (Lorence et al., 2004); where genes as DHA (*ascorbate peroxidase*), GDH (*D-galacturonate reductase*), GGP (*L-galactose dehydrogenase*), GME (*L-galactono-1,4-lactone dehydrogenase*), GMP (*GDP-D-mannose3′, 5′-epimerase*) among others have a role in the synthesis of this component (Caruso et al., 2021; Zheng et al., 2022). Yet, reports about the genes associated with ascorbic acid synthesis and its antioxidant activity in potatoes are not available. Identifying molecular markers that flank gene-related to the production of phenolic compounds and ascorbic acid in potatoes will allow the use of Marker-Assisted Selection (MAS) in breeding programs. MAS could reduce the time and cost of improving biofortified materials.

The origin, diversification, and domestication of *S. tuberosum* occurred in the South American Andean highlands around the Lake Titicaca between Peru and Bolivia (Grun, 1990). The potato´s contemporary landraces are gene pools from south Chile to north Colombia (Hawkes, 1990; De Haan and Rodriguez, 2016). Colombia is one of the secondary diversity centers of potatoes (Spooner et al., 2014), considered essential for the country's food security, in which 3.1 million tons are produced (FAOSTAT, 2022). The Corporación Colombiana de Investigación Agropecuaria (AGROSAVIA) is responsible for conserving and characterizing the Colombian potato genetic resources (Colombian Central Collection - CCC). This collection is one of the largest and most diverse for this species in South America (Berdugo-Cely et al., 2017; Berdugo-Cely et al., 2021; Nagel et al., 2022), with more than 2000 accessions that include landraces or native genotypes (materials developed as a product of selection by Colombian farmers), commercial varieties, and wild-relative species of *S. tuberosum*. In an earlier study (Berdugo-Cely et al., 2017), 809 native genotypes of the CCC collection were analyzed using morphological characters and Single Nucleotide Polymorphism (SNP). Diploid and tetraploid genotypes were separated in two main populations distinguished by ploidy level. Furthermore, this study also found high genetic diversity in shape, flesh, and skin colors in tubers of tetraploid potatoes. The flesh and skin color of potato tubers have been correlated with polyphenols and ascorbic acid content. Tubers with dark colors presented high content of these compounds and an increase in antioxidant activity (Hejtmánková et al., 2009). The high phenotypic diversity in potato tubers conserved in the CCC allows us to believe that this germplasm has a valuable genetic resource for breeding programs to generate new cultivars with high nutritional values.

In consequence, the present study aimed 1) to analyze the natural phenotypic variation of Total Phenol Content (TPC), Ascorbic Acid Content (AAC), and Antioxidant Activity (AA) in potato tubers of diploid and tetraploid genotypes of the CCC to select promising genotypes to be used in future breeding programs, and 2) to associate this phenotypic information with SNP through GWAS to identify market-trait association which may collocate candidate genes with possible implication in the production of polyphenols and ascorbic acid and their antioxidant activity in potato. We hypothesize that the ploidy level affects the concentration and the genetic regulation of polyphenols, ascorbic acid biosynthesis, and their antioxidant activities.

Materials and methods

Plant material and growth conditions

For this study, we used 404 genotypes (accessions) of the potato CCC germplasm bank previously characterized at the molecular level using SNPs by Berdugo-Cely et al. (2017). The analyzed population held 320 tetraploid accessions (2n=4x=48) and 84 diploid accessions (2n=2x=24) genotypes. Of these materials, 335 are Colombian natives, 33 are foreign genotypes from Bolivia, Ecuador, the United States, Netherlands, and Peru, and 33 are of unknown geographic origin (Supplementary Table S1). In May of 2014, sixteen individuals per accession were grown in field conditions in Zipaquirá, Cundinamarca, Colombia (5° 03" 34.36"; N. 74° 03" 29.61 W). The location in Zipaquirá is at 2950 masl. presents clay loam textured soils with a pH of 5.5, relative humidity of 75%, and an average temperature of 15°C. During the harvest period, for each diploid (in September) and tetraploid (in November) type of potatoes, ten randomized tubers with weights between 60 and 80g with high physical and sanitary quality traits per accession were collected. These tubers were washed and sent for chemical analysis to the Laboratorio de Ciencia de los Alimentos of the Universidad Nacional de Colombia situated in Medellín (Antioquia) (https://direcciondelaboratorios.medellin.unal.edu.co/index.php/nuestros-laboratorios/facultad-de-ciencias/22).

Phenotyping

Sample preparation

We cut one section of each tuber per accession collected (10 tubers); these sections were mixed and macerated. From this tissue, five grams (g) were mixed with 25mL of 80% methanol and shook at 500 rpm per 60 min. The mixtures were then centrifuged at 3000 rpm for 15 min. Subsequently, the suspensions were filtered and conserved at four (4) °C until further analysis (approximately between one and five days). Three suspensions per potato accession were independently prepared to measure all traits, constituting three technical replicates.

Total phenol content

The TPC was assayed using a modified protocol for the Folin-Ciocalteu method (Singleton and Rossi, 1965). Fifty microliters (µL) of filtered sample suspension, 425µL of dd H_2O, and 125µL of Folin-Ciocalteu reagent were mixed. After five min of mixing, 400µL of 7.1% Sodium carbonate (Na_2CO_3) was added, and the mixtures were conserved for one hour in dark incubation at room temperature; after this time, the absorbance was measured at 760nm using gallic acid as a standard. The TPC was expressed in terms of milligrams of Gallic Acid Equivalent (GAE) per 100g of fresh weight (FW) (mg GAE/100g FW).

Antioxidant activity

DPPH assay

A modified protocol for DPPH (*2.2-Diphenyl-1-picrylhydrazyl Radical Scavenging Activity*) assay was used to assess the AA (Brand-Williams et al., 1995). Nine hundred and ninety microliters of DPPH solution with 10µL of sample suspension were mixed and incubated for 30 min in the dark. After this time, the absorbance was measured at 517nm using Trolox (*6-hydroxy-2.5.7.8-tetramethylchroman-2-carboxylic acid*) reagent as a standard. The DPPH results (AA_DPPH) were expressed in terms of micromoles of Trolox Equivalent (TE) per 100g of FW (µmol TE/100g FW).

FRAP assay

The second assay to assess the AA was a modified protocol for FRAP (*Ferric Reducing Antioxidant Power*) assay (Benzie and Strain, 1996). Fifty microliters of sample suspensions were mixed with 50µL of 0.3µM acetate buffer (pH 3.4) and 900µL of FRAP solution (10 mM of TPTZ (*2.4.6-tripiridil-s-triazine*), 20 mM of Ferric chloride ($FeCl_36H_2O$), 0.3 µM acetate buffer at pH 3.6). The mix was incubated for 30 min in the dark at room temperature. After this time, the absorbance was measured at 590nm using L-ascorbic acid as a standard. The FRAP results (AA_FRAP) were expressed as milligrams of Ascorbic Acid Equivalent (AAE) per 100g of FW (mg AAE/100g FW).

Ascorbic acid content

The AAC was evaluated through the modified method of Kelebek et al. (2009). Twenty microliters of the sample were filtered with a membrane filter of 0.45µm (Millipore) before being analyzed by High-Performance Liquid Chromatography (HPLC) in a liquid Shimadzu chromatograph (LC-20AD) equipped with an autoinjector SIL-20A/HT and a photodiode array (PDA) detector (SPD-M20A). A LiChrospherR 100 RP-18 column (Merck) (250 x 4 mm, particle size=5 LiChrospherR 100 RP-m) was used for HPLC separation, using a mobile phase of 0.1% of formic acid with flow rate of 0.8 mL/min at 35°C. The AAC concentrations were quantified compared to the L-ascorbic acid reagent as a standard. The AAC results were expressed as milligrams of Ascorbic Acid per 100g of FW (mg AA/100g FW).

Phenotypic statistical analysis

Means and standard deviations of the values obtained for TPC, AAC, and AA (AA_DPPH and AA_FRAP) were calculated. The Shapiro-Wilk test was used to check the normality of the data. One-way analysis of variance (ANOVA) and Tukey's test (p<0.05) were conducted using the shapiro.test, aov, and TukeyHSD functions of *R-stats* package (R Stats, 2022) of R software (R, 2022) to 1) compare the means between and within populations for each variable and 2) compare the variations of TPC, AA, and AAC contents according to skin and flesh tuber colors (Supplementary Table S1). The tuber phenotype color traits were previously obtained (Berdugo-Cely et al., 2017) using the descriptors of the CIP (Centro Internacional de la Papa) to characterize native potatoes (Gómez, 2014). The following statistical analyses using the quantitative traits were independently run in diploid and tetraploid genotypes. Using the *R-stats* package (R Stats, 2022) a Pearson correlation (p<0.05) was used to check the relationship among variables. A Principal Component Analysis (PCA) was performed, and the two first components of the PCA were used to generate a multivariate clustering analysis using the Ward method and the Euclidean distance. These multivariate analyses were implemented in the *R-factoextra* package (Kassambara and Mundt, 2020).

Genotyping

Genotype calling and filtering SNP data

The 404 potato accessions used in this study were previously genotyped in the study of Berdugo-Cely et al. (2017) using the Illumina SolCAP 8K SNP array. This chip array included 8303 SNP markers selected by EST (Expressed Sequence Tag) and transcriptomic data from five cultivars (Bintje, Kennebec, Premier Russet, Shepody, Snowden, and Atlantic) developed in the United States (Hamilton et al., 2011; Felcher et al., 2012). Here, the fluorescence intensities were extracted from Illumina GenomeStudio software (GenomeStudio Software, 2022), and the data was used to determine the genotype calling using *fitTetra* R-package (Voorrips et al., 2011) using with *SaveMarkersModels* function with the parameters of filtering reported by Vos et al. (2015). In the genotype calling, diploids were included allowing the corrected assigned of nulliplex (AAAA), duplex (AABB), and quadruplex (BBBB) clusters. Tetraploid genotypes were coded as: AAAA (0), AAAB (1), AABB (2), ABBB (3), BBBB (4), and diploid genotypes as: AA (0), AB (2), BB (4). Monomorphic markers, genotypes, and markers with missing data up to 20% and data with Minimum

Allele Frequency (MAF) lower than 5% were filtered (Supplementary Table S2).

Population structure, genetic diversity, and linkage disequilibrium analysis

The genetic structure analyses based on the Bayesian clustering simulation in *STRUCTURE* software (Pritchard et al., 2000) were implemented on 1) the mixed population composed of diploid and tetraploid accessions and conducted separately on 2) diploid accessions, and 3) tetraploid accessions to assess the effect of various ploidy levels on the reconstruction of the population stratification. These analyses were simulated from one to ten subpopulations with five replicates for each dataset, assuming an admixture model and correlated allele frequencies with a burn-in of 100000 and 50000 Markov Chain Monte Carlo (MCMC) iterations after the burn-in. The number of subpopulations among the samples was determined using the Evanno method (Evanno et al., 2005) in *STRUCTURE Harvester* (Earl and vonHoldt, 2012). The PCA was conducted in the *adegenet* R-package (Jombart and Ahmed, 2011) using the identified genetic structure, which was confirmed by an analysis of molecular variance (AMOVA´s) and Phi values with 1000 permutations in the *poppr* R-package (Kamvar et al., 2014). The expected heterozygosity (He) was calculated to analyze the genetic diversity in each population using the *adegenet* package. We determined the linkage disequilibrium (LD) of the diploid, tetraploid, and mixed potato populations following the methodology reported by Vos et al. (2017). For the SNPs with a known physical position in the reference genome, all possible pairwise correlation coefficients (r^2) were computed. The LD mean and decay per population were calculated using this information. The LD decay was determined by plotting the r^2 values against the physical distance and visualizing it using the *ggplot2* package in the R program (R, 2022).

Multivariate analysis and selection of promising diploid and tetraploid genotypes using phenotypic and genomic information

A PCA analysis that included the molecular and phenotypic information was implemented; in this case, the molecular traits correspond to the two first components of the PCA implemented with only molecular information and phenotypic information corresponding to the means values generated for each genotype in each feature (TPC, AA_DPPH, AA_FRAP and, AAC). A new cluster analysis was developed, and the genetic groups were described at the phenotypic level. Promising genotypes to TPC, AAC, and AA of the AGROSAVIA collection were identified and selected using an Index Selections (IS) calculated through the *selection index* R-package (Goyani, 2021), adjusting the weights of each character according to their heritability (h^2) that was computed from phenotypic and genotypic data using the *heritability* R-package (Kruijer, 2019).

Genome-wide association studies

The GWAS analyses were conducted using the mean values of TPC, AAC, and AA (AA_DPPH and AA_FRAP) obtained for each genotype. These analyses were implemented in the mixed population and independently in diploid and tetraploid populations because the CCC has a clear population genetic structure in function to the ploidy level (Berdugo-Cely et al., 2017). The complete dataset was analyzed under a tetraploid approach, while the diploid and tetraploid datasets were analyzed under their respective ploidy model approach. This study implemented the GWAS analyses through single-locus and multi-locus methods in the *GWASpoly* (Rosyara et al., 2016) and *mrMLM* (Zhang et al., 2020) R-packages, respectively. Kinship (K) matrices and population structure were included as random and fixed effects to remove associations related to the genetic structure.

A mixed linear model (MLM) implemented in *GWASpoly* was carried out for the single-locus method using the eight available models in this R-package (general, additive, diplo-general, diplo-additive, 1-dom-alt, 1-dom-ref, 2-dom-alt, and 2-dom-ref) (Rosyara et al., 2016). The *p*-values were adjusted with the False Discovery Rate (FDR) method (Benjamini and Hochberg, 1995) at $p<0.001$ to reduce the false-positive rate for the marker-trait association. On the other hand, in the multi-locus GWAS implemented in *mrMLM*, six models (mrMLM, FASTmrMLM, FASTmrEMMA, pLARmEB, pKWmEB, and ISIS EM-BLASSO) were evaluated (Zhang et al., 2020) using the default parameters. The threshold values to determine associated markers are usually established by corrections methods (e.g., Bonferroni), but those are sometimes too strict, missing some essential associations (Yousaf et al., 2021). Some studies (Phan et al., 2018; Riaz et al., 2018; Yousaf et al., 2021) have instead used a threshold of 3.0 or higher to determine the significant SNP-trait associations (STAs) or Quantitative Trait Nucleotide (QTNs). For this study, we established that STAs or QTNs with values of LOD (logarithm of the odds) scores higher than 3.5 were significantly associated. We reported LOD scores identified after evaluating all models for each STA (detected in a single-locus model) or QTN (detected in multi-locus models). The associated SNPs reported in this study were detected for a minimum of two models in each GWAS model. We constructed the Quantile–Quantile (q–q) and Manhattans plots for each evaluated model in R software. We graphed the plots using the model that detected the highest number of STAs in *GWASpoly*, while the graphs reported to *mrMLM* included the QTNs detected in all multi-locus models.

Candidate genes

The STAs and QTNs detected in common by single and multi-locus models (LOD score > 3.5) were searched against the potato genome of *Solanum tuberosum* group Phureja DM1-3 PGSC v4.03 in the Spud database - Potato Genomic Resource (PGSC Data Download; Xu et al., 2011; Sharma et al., 2013). Those genes were proposed as candidate genes associated with TPC, AA (AA_DPPH and AA_FRAP), and ACC.

Results

Total phenol content, ascorbic acid content and antioxidant activity in potato populations

We studied antioxidant properties evaluating the CCC for TPC, AAC, and AA. Diploid and tetraploid genotypes in the CCC varied substantially in the TPC, AA, and AAC traits. These values did not follow a normal distribution ($p<0.05$) except for AAC in the diploid population ($p=0.989$) (Figure 1). Within diploid and tetraploid populations, all traits presented a wide range of phenotypic diversity with highly significant differences ($p<0.0001$) (Figure 1). Differences in TPC values were not significant between diploid and tetraploid populations ($p=0.3302$), but for some tetraploids, TPC was higher. Therefore, the range for tetraploids was wider (144.65 to 3218 mg GAE, median: 838.06) than in diploids (189.4 to 2584.4 mg GAE; median: 826.245) genotypes (Figure 1A). Values for AA using the DPPH assay ranged from 436.37 to 23677.38 µmol TE in diploid genotypes (median: 5287.61), and between 192.85 to 20243.40 µmol TE (median: 4248.82) in tetraploid genotypes (Figure 1B), while values for AA using the FRAP assay ranged from 271.26 to 4086.43 mg AAE (median: 831.81) in diploid genotypes and between 167.74 to 3333.90 mg AAE in tetraploid genotypes (median: 634.5) (Figure 1C). Diploid genotypes presented significantly higher values for AAC ($p<0.0001$) and AA measured by FRAP ($p<0.0001$) and DPPH assay ($p<0.0001$). Finally, AAC values ranged from 1.94 to 6.61 mg AA in diploid genotypes (median: 4.04) and between 1.65 to 4.26 mg AA in tetraploid genotypes (median: 2.54) (Figure 1D). Diploid and tetraploid differed in their antioxidant compounds concentration; diploid potatoes had significantly higher values for AAC and AA, while tetraploid genotypes had higher but non-significant values of TPC.

Correlations among TPC, AAC, and AA in potato populations

To understand which compound could have more influence on antioxidant activity, we calculated pairwise Pearson correlations between traits. The TPC and AA traits presented intermedium and high significant positive correlations ($p<0.0001$) in diploid ($r^2>0.47$ in TPC *vs* AA_DPPH) and tetraploid ($r^2>0.62$ in TPC *vs* AA_DPPH) populations, respectively. AA measured by DPPH and FRAP assays presented high significant positive correlations ($p<0.0001$) in both populations ($r^2>0.75$). In contrast, low or no significant correlations were detected between TPC and AAC (r^2 =0.19 in diploids and $r^2<-0.11$ in tetraploids) and between AAC and AA ($r^2<0.25$ in diploids and $r^2<-0.09$ in tetraploids) in both populations (Figures 2A, B). Based on the results, TPC appeared to be essential in AA in potatoes.

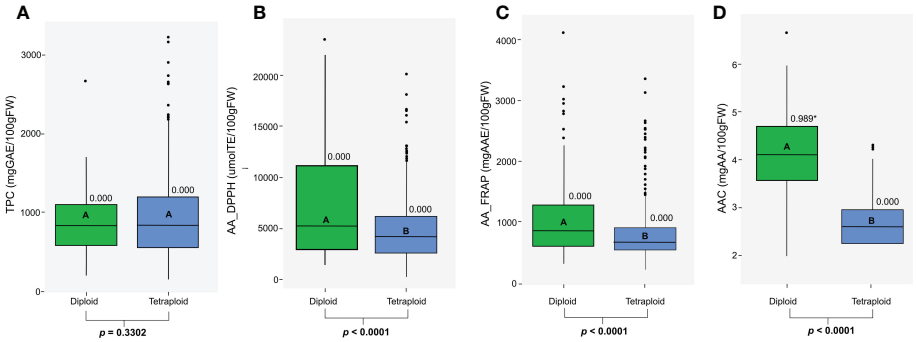


FIGURE 1
Distribution of **(A)** Total Phenol Content (TPC), **(B)** Antioxidant Activity measured by a DPPH assay (AA_DPPH), **(C)** Antioxidant Activity measured by a FRAP assay (AA_FRAP), and **(D)** Ascorbic Acid Content (AAC) in diploid and tetraploid potato populations. The *P-value* for the Shapiro-Wilk normality test is shown, and significant values ($p>0.05$) are specified with an asterisk. The letters and *p-values* in bold boxplots indicate a significant difference between populations based on the ANOVA and the Tukey test ($p<0.05$).

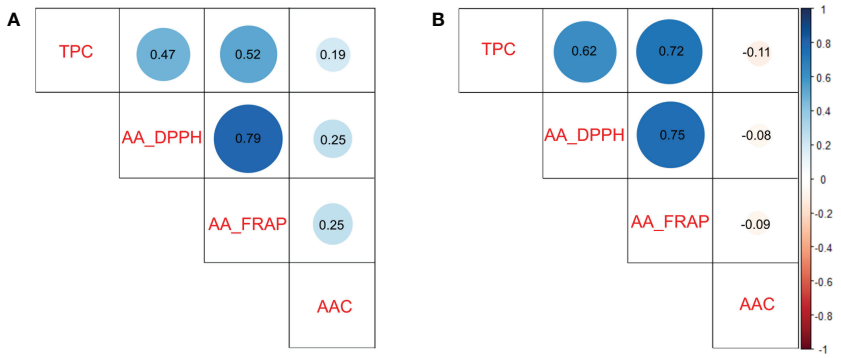


FIGURE 2
Pairwise Pearson correlation between Total Phenols Contents (TPC), Ascorbic Acid Content (AAC), and Antioxidant Activity (AA_DPPH and AA_FRAP) in potato populations. **(A)** Diploid population and **(B)** Tetraploid population.

Phenotypic variation of TPC, AAC, and AA related to tuber color in potato populations

To evaluate if antioxidant content and properties differ depending on skin and flesh tuber color, we conducted an ANOVA and a Tukey Test, independently for tetraploids and diploids. TPC and AA levels presented significant differences (p<0.002) depending on the skin tuber color. In the tetraploid population, purple-blackish tubers had higher levels of TPC and AA (Supplementary Figures S1A–C), and diploid pink-red potatoes presented high and significant values of AA measured by FRAP assay (p<0.05) (Supplementary Figure S1C). In contrast, the AAC did not seem to depend on the skin tuber color in any population (Figure S1D). Moreover, the TPC, AA, and AAC in potatoes did not differ concerning the color of the flesh tuber in the analyzed populations (Supplementary Figure S2). From our results, in potatoes, TPC and AA change depending on skin tuber color, but AAC no.

Phenotypic multivariate analysis in diploid and tetraploid populations

We conducted a PCA to understand which components of the phenotypic data explain most of the variation. The two first components of the PCA explained more than 80% (Dim1: 57.5% and Dim2: 26.3%) of the total variation for the diploid and tetraploid potato populations in the multivariate analyses (Figure 3A). AA and TPC were the most informative traits in the first component of the PCA, explaining between 57-61% and 48-56% of the variation, respectively. AAC was the most informative trait in the second component explaining between 91-99% of the variation. The phenotypic PCA and the cluster analysis confirmed that diploids had higher values of AAC and AA, and tetraploids tend to have higher values of TPC (Figure 3A). The phenotypic traits grouped the diploid and tetraploid samples in three clusters (Supplementary Table S3). In diploids, the genotypes with pink-red potatoes (n=13) with the highest values for TPC (mean=1352.92 ± 486.03 mg GAE), AA_DPPH (mean=16589.91 ± 4336.02 µmol TE), and AA_FRAP (mean=2432.5 ± 800.45 mg AAE) were clustered in the P_D_Cluster_3, while the genotypes with pink-red and purple blackish potatoes (27) with highest values for AAC (mean=4.92 ± 0.55mg AA) conformed the P_D_Cluster_2. In tetraploids, the genotypes with purple-blackish potatoes (52) with the highest TPC contents (mean=1817.94 ± 527,43 mg GAE), AA_DPPH (mean=9319.59 ± 4316.93 µmol TE) and AA_FRAP (mean=1463.31 ± 689.67 mg AAE) values were clustered in P_T_Cluster_1, while the genotypes with pink-red and purple-blackish potatoes (109) that present the highest AAC values (mean=3.09 mg AA ± 0.47) were found in P_T_Cluster_3 (Supplementary Table S3). We could recover clusters that differ in their TPC, AAC, and AA values.

Genetic diversity, population structure, and linkage disequilibrium in potato populations

The potato germplasm's genetic diversity and population structure analyses were analyzed using 4488, 2207, and 4487 polymorphic SNPs to assess the mixed, diploid, and tetraploid populations (Table 1). The mixed population was divided into two main subpopulations, the diploid and tetraploid subpopulations (Figure 3B and Supplementary Figure S3A). These subpopulations had a clear genetic differentiation (Phi=0.405) and a high genetic diversity (He=0.405). Diploid accessions presented medium genetic diversity levels (He=0.177) and were separated into three subpopulations (Supplementary Figure S3B) with a low genetic

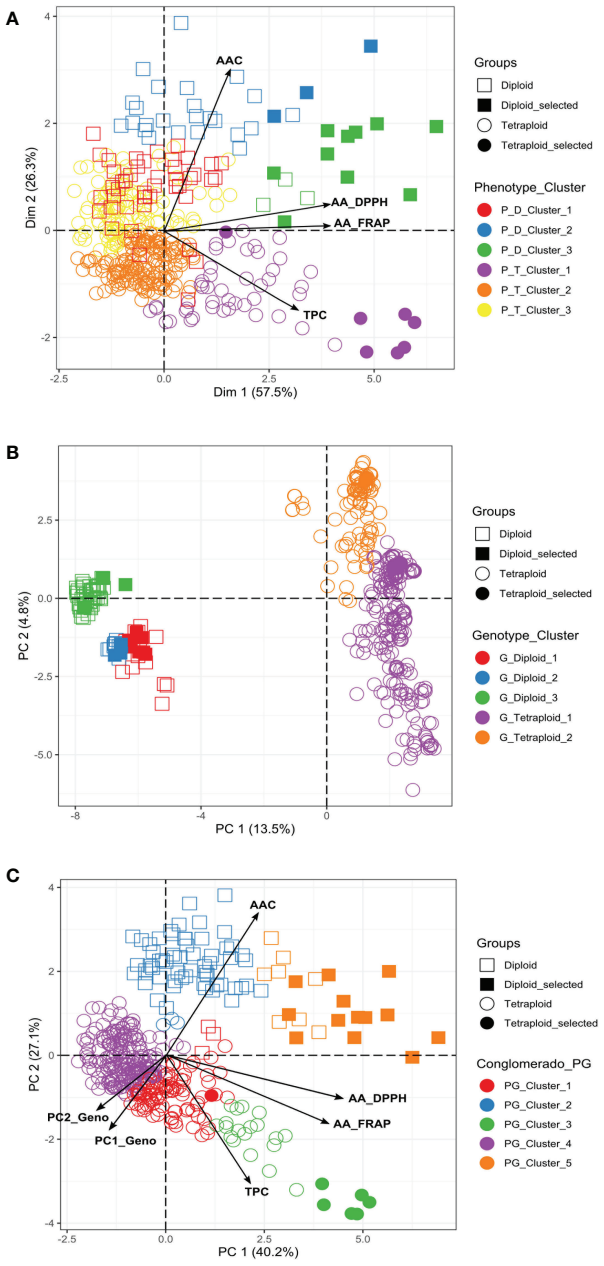


FIGURE 3
Principal Component Analysis (PCA) based on TPC, AA, and AAC phenotypic data and polymorphic molecular markers (4488 SNPs) for diploid and tetraploid accessions. **(A)** Phenotypic PCA **(B)** Genotypic PCA, and **(C)** PCA based on phenotypic and genotypic data. P, Population; D, Diploid; T, Tetraploid; G, Genotypic; PG, Phenotypic/Genotypic. The promising selected genotypes identified in this study were indicated in the figures. Additionally, the clusters showed in the figures were obtained from the phenotype, genotype, and phenotype/genotype multivariate analyses.

differentiation (Phi=0.038) (Table 1). However, these three diploid subpopulations differed in TPC, AAC, and AAC values. The subpopulation G_Diploid_2 regrouped genotypes with yellow-orange and pink-red potatoes with the highest levels of TPC (mean=1126.3 ± 394.52 mg GAE) and AA measured by DPPH (mean=10538.85 ± 5858.06 µmol TE) and FRAP (mean=1248.1 ± 534.33 mg AAE) assays. Moreover, the three subpopulations presented similar values of AAC with means between 3.97 to 4.25 mg AA (Supplementary Table S4). On the other hand, tetraploid genotypes were divided into two highly diversified subpopulations (He=0.422) (Supplementary Figure S3C) with low genetic structure (Phi=0.040) (Table 1). At the phenotype

TABLE 1 Genetic diversity, structure population and linkage disequilibrium statistics in potato populations.

Population	Number of samples	Number of polymorphic markers	Number of subpopulations	AMOVA Source of variation	AMOVA Variation (%)	Genetic diversity (He)	Phi	LD mean (r^2)
Mixed	404	4488	K=2	Among	40.50	0.371	0.405*	0.192
				Within	59.49			
Diploid	84	2207	K=3	Among	3.81	0.177	0.038*	0.403
				Within	96.18			
Tetraploid	320	4487	K=2	Among	4.07	0.422	0.040*	0.179
				Within	95.92			

level, these two tetraploid subpopulations presented similar values for TPC, AAC, and AA (Supplementary Table S4). At the linkage disequilibrium level, all SNPs detected in mixed, diploid, and tetraploid populations were used to calculate the LD mean (r^2) and LD decay. These values were similar in mixed (r^2 =0.192 with a LD decay=12500 base pairs - pb) and tetraploid (r^2 =0.179 with a LD decay=12000 pb) populations, while the LD was higher in the diploid population (r^2 =0.403 with a LD decay=11500 pb) (Table 1 and Figure 4). The populations with high levels of genetic diversity (mixed and tetraploid) presented a minor level of LD. In contrast, the diploid population presenting low genetic diversity presented the highest LD levels. These results suggest that the low genetic diversity and high LD in the diploid population can affect the statistical power to identify SNP associated with phenotype traits in this population. The number of genetic variants necessary to conduct association studies in this subpopulation may be higher than the one used in this study.

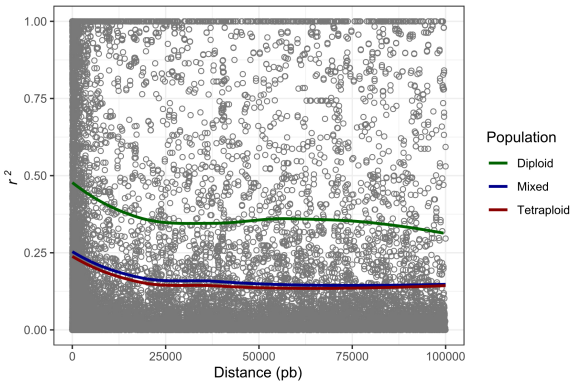


FIGURE 4
Genome-wide LD decay calculated to mixed, tetraploid, and diploid populations. The x-axis represents the physical distance in base pairs (pb) and the y-axis represents the average pairwise correlation coefficient (r^2) of SNPs.

Multivariate analysis and selection of promising genotypes in diploid and tetraploid populations using phenotypic and genomic information

The collection was also analyzed by combining the phenotypic and genomic data. The PCA and clustering analyses that included 67% of the genetic variability explained in the two first components of the PCA (PC1: 40.2% and PC2: 27.1%) suggested the division of these genotypes into five genetic groups, three subgroups for tetraploid and two subgroups for diploid genotypes (Figure 3C). The tetraploid genotypes with purple-blackish tubers (n=25) with the highest levels of TPC (mean=2062.41 ± 547.37 mg GAE), AA_DPPH (mean=12208.7 ± 2954.61 µmol TE), AA_FRAP (mean=1984.8 ± 622.53 mg AAE) and AAC (mean=2.46 mg ± 0.68 AA) were clustered in the PG_Cluster_3, while the diploids genotypes with pink-red tubers (n=21) with the highest levels of TPC (mean=1245.95 ± 426.82 mg GAE), AA_DPPH (mean=14967.1 ± 4687.79 µmol TE), AA_FRAP (mean=2208.63 ± 797.35 mg AAE) and AAC (mean=4.52 mg ± 0.68 AA) were grouped in the PG_Cluster_5 (Supplementary Table S5).

The selection index (SI) for each trait was calculated based on heritability values (h^2: TPC=0.6956; AA_DPPH=0.6505; AA_FRAP=0.466 and AAC=0.6369). The tetraploid accessions, And_424, And_186, And_210, And_18, and And_209, had higher SI values (SI>1845) and concentrations for TPC, AA (DPPH and FRAP), and AAC traits with a concentration of phenolic compounds above 2651 mg GAE. We also selected the tetraploid accessions And_210 and And_290, and diploids Cha_49, Cha_46, and Phu_42 genotypes with SI>11886 and antioxidant activities above 18271 µmol TE measured by DPPH. Tetraploids And_186 and And_209 and diploids Cha_49, Cha_15, and Cha_16 genotypes were selected (SI>1453) for presenting antioxidant activities above 3113 mg AAE measured by FRAP. Diploid accessions Cha_2, Cha_16, Phu_7, Phu_75, and Phu_48 were selected (SI>7) based on their ascorbic acid contents above 5.24 mg AA. The tetraploid accessions And_186, And_209, and And_210 and the diploids Cha_49 and Cha_16 genotypes are particularly interesting for

presenting high values for more than two traits (Figure 3A and Supplementary Table S1). This strategy allowed us to select 11 genotypes as promising. Additionally, a selection index calculated using the joint data of the four traits allowed the addition of nine genotypes to the list of promising materials (SI>11431). In total, we selected 20 genotypes as promising genotypes (Supplementary Table S1).

SNP-trait associations using single-locus GWAS analyses

A single-locus GWAS was performed in mixed, diploid, and tetraploid populations to identify STAs with TPC, AA_DPPH, AA_FRAP, and AAC. We considered an STA when at least two models with LOD score > 3.5 of *GWASpoly* identified it. The analysis revealed fifty-eight (58) STAs, 25 STAs in the mixed population, 16 STAs in diploid, and 17 STAs in tetraploid populations (Supplementary Table S6). Twelve STAs were pleiotropic because they were associated with two or three traits, while 17 were shown in two or three populations. Most STAs were revealed in only one population: 14 STAs in diploid, 17 in mixed, and 11 in tetraploid populations. Manhattans and q-q plots showed significant STAs with normal data distribution (Figure 5 and Supplementary Figures S4-S6). In the mixed population, the analyses showed five STAs (in chromosomes 2, 4, and 9) with TPC with phenotype variation estimated (PVE-r^2) of 0.2-4.4%, 17 STAs (in chromosomes 2-6 and 9) with AA_DPPH (0.1-7.1%), nine STAs (in chromosomes 2-4, 6 and 9) with AA_FRAP (0.1-7.8%) and nine STAs (in chromosomes 2-4, 6 and 9) with AAC (0.1-2.1%) (Figure 5 and Supplementary Table S6). In the diploid population, the analysis revealed two STAs in chromosome 3 and 12 STAs in chromosomes 2, 6, and 9 associated with AA_DPPH (31.1%) and TPC (2.2-15%), respectively (Supplementary Table S6 and Supplementary Figure S5). However, no associations with AA_FRAP and

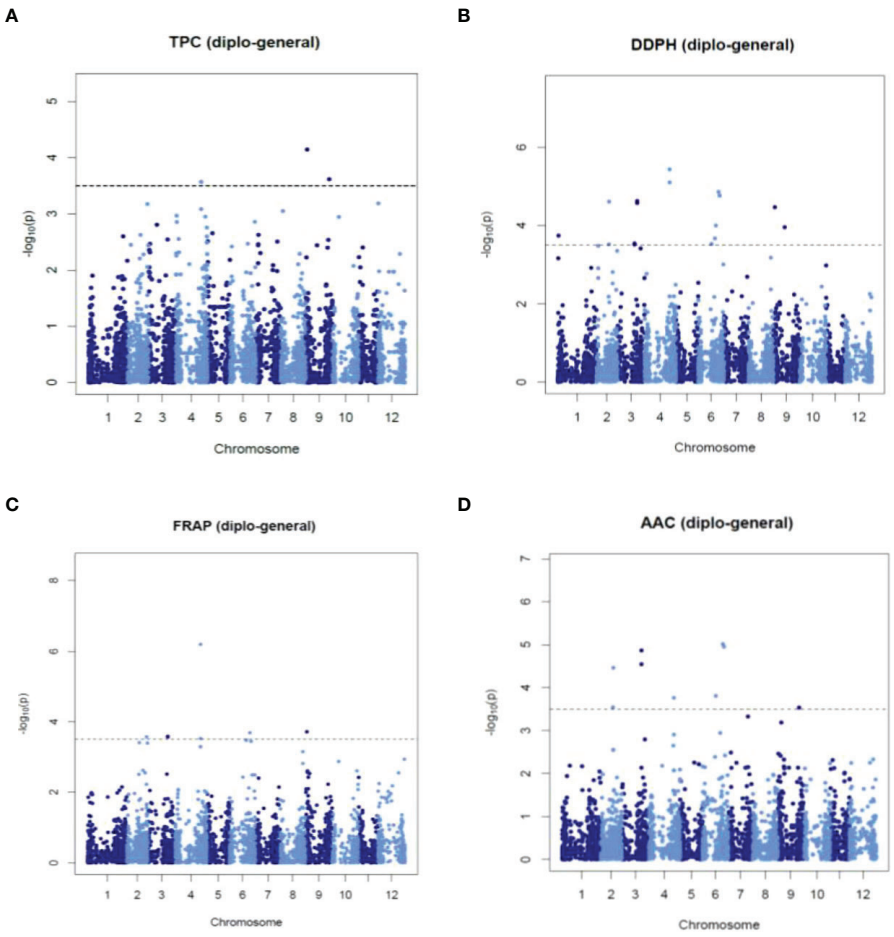


FIGURE 5
Manhattan plots showing significant STAs for TPC **(A)**, AA_DPPH **(B)**, AA_FRAP **(C)**, and AAC **(D)** obtained using single-locus GWAS methods in the mixed population.

AAC were shown. In the tetraploid population, the analysis exhibited four STAs in chromosomes 2, 4, and 9 with TPC (1.3-6.6%), nine STAs in chromosomes 2, 4, 6, 8, 9, and 11 with AA_DPPH (1.7-15.9%), 12 STAs in chromosomes 1, 4, 6, 8, 9, and 11 with AA_FRAP (0.8-10.2%), and one STA in the chromosome 4 with AAC (5.4%) (Supplementary Table S6 and Supplementary Figure S6).

Quantitative-trait nucleotides using multi-locus GWAS analyses

Using the same three potato populations, we ran six models of multi-locus GWAS analyses. A total of 28 QTNs had significant associations with TPC, AA_DPPH, AA_FRAP, and AAC in at least two models implemented in *mrMLM*. This analysis revealed the highest number of QTNs in the tetraploid population (15 QTNs). In mixed and diploid populations, the analyses displayed 11 QTNs and two QTNs (Supplementary Table S7). Three QTNs were pleiotropic, and four QTNs were present in more than two populations. Most QTNs were revealed in only one population: two QTNs in diploid, nine in mixed, and 13 in tetraploid populations (Supplementary Table S7). We present the significant QTNs in Manhattan and q-q plots (Figure 6 and Supplementary Figures S7-S9). In the mixed population, four QTNs in chromosomes 4 and 10 with a PVE-r^2 between 0.8-22.4% were associated with TPC, five QTNs in chromosomes 2, 3, 4, and 9 were associated with AA_DPPH (2.4-28.6%), two QTNs in the chromosomes 1 (5.1-32.5%) and 4 (10.4-11.0%) were associated with AA_FRAP. Finally, one QTN in chromosome 9 was associated with AAC (13.6-30.8%) (Supplementary Table S7 and Figure S6). In the diploid population, only two QTNs had an association with TPC (25.76%) in chromosome 6 and AA in chromosome 6 with DPPH (63.8%) and FRAP assays (25.3%) (Supplementary Table S7 and Supplementary Figure S8). In the tetraploid population, six QTNs associated with TPC were revealed in chromosomes 2, 4, 6, and 11, explaining between 4.6-26.8% of PVE. Three QTNs in chromosomes 1, 6, and 4 were associated with AA_DDPH (6.6-36.4%). Five QTNs were associated with AA_FRAP in chromosomes 1, 2, 4, and 12 (2.4-31.2%). Finally, only two QTNs in chromosomes one and 11 were associated with AAC (10.1-14.2%) (Supplementary Table S7 and Supplementary Figure S9).

Potential candidate genes revealed by single and multi-locus GWAS

We compared the single and multi-locus GWAS results to identify the more reliable genes associated with TPC, AA, and ACC. A total of eight STAs/QTNs were identified that flanked seven genes in chromosomes 1, 2, 3, 4, and 9 of the potato genome (Table 2). Four candidate genes were pleiotropic. The *PGSC0003DMG400024824* gene (Chr 4) of the CAZy family with a Glycosyltransferase function was detected by two STAs/QTNs (c1_12945 and c2_43998) associated with TPC (phenotype variation estimated PVE-r^2 3.3-22.4%), AA_FRAP (2.1-36.4%), AA_DPPH (0.7-36.4%) and AAC (0.2-5.1%). The *PSC0003DMG400002675* gene (Chr 9) that has a fructose-bisphosphate aldolase function was associated with TPC (0.2-1.3%), AA_DPPH (4.8-24.7%) and AA_FRAP (4.2-4.9%). The *Soltu.DM.03G022550.1* gene (Chr 4) with a pectin methyl esterase function was associated with AA_DPPH (0.1-63.8%), AA_FRAP (22.7-25.3%) and AAC (0.1%) and the *PGSC0003DMG400001451* gen (Chr 2) with folate carrier protein were associated with TPC (13.9-22.6%) and AA_FRAP (7.8%). Additionally, three genes were associated with only one trait. The *Soltu.DM.01G033780.1* gen (Chr 1) with a cysteine synthase C1 function was associated with AA_FRAP (6.6-11%), the *Soltu.DM.03G015260.1* gene (Chr 3) with a BCL-2-associated athanogene function was associated with AA_DPPH (0.1-9.5%), and the *Soltu.DM.09G021110.1* gene (Chr 9) annotated as a basic helix-loop-helix (bHLH) DNA-binding superfamily protein was associated with AAC (0.6-30.8%) (Table 2).

Discussion

The benefits in human health of consuming fruits and vegetables with high antioxidant properties are well known. The antioxidant compounds can protect against oxidative damage by scavenging ROS and reducing oxidation reactions. The phenolic compounds and ascorbic acid have antioxidant and anti-inflammatory properties, inhibiting, preventing, and treating various human diseases, including cardiovascular disease, neurodegenerative disorders, obesity, and hypertension (Arruda et al., 2020; Rahman et al., 2022). The antioxidant components extracted from potato tubers have prevented stomach, prostate, colon, and liver cancer and diseases such as type 2 diabetes (Burgos et al., 2019; Hellmann et al., 2021). In this study, we evaluated the total phenolic compounds (TPC) and ascorbic acid (AAC) contents and their antioxidant activities (AA) in a diverse panel of diploid and tetraploid potatoes genotypes to select promising genotypes to use in future breeding programs and to identify STAs and QTNs flanked in genes with a possible genetic association with the synthesis of TPC, and AAC, and their AA in potatoes.

Diploid and tetraploid potato genotypes had differential values of TPC, AAC, and AA

In this study, we analyzed the influence of ploidy levels on the production of antioxidant compounds in potatoes. The

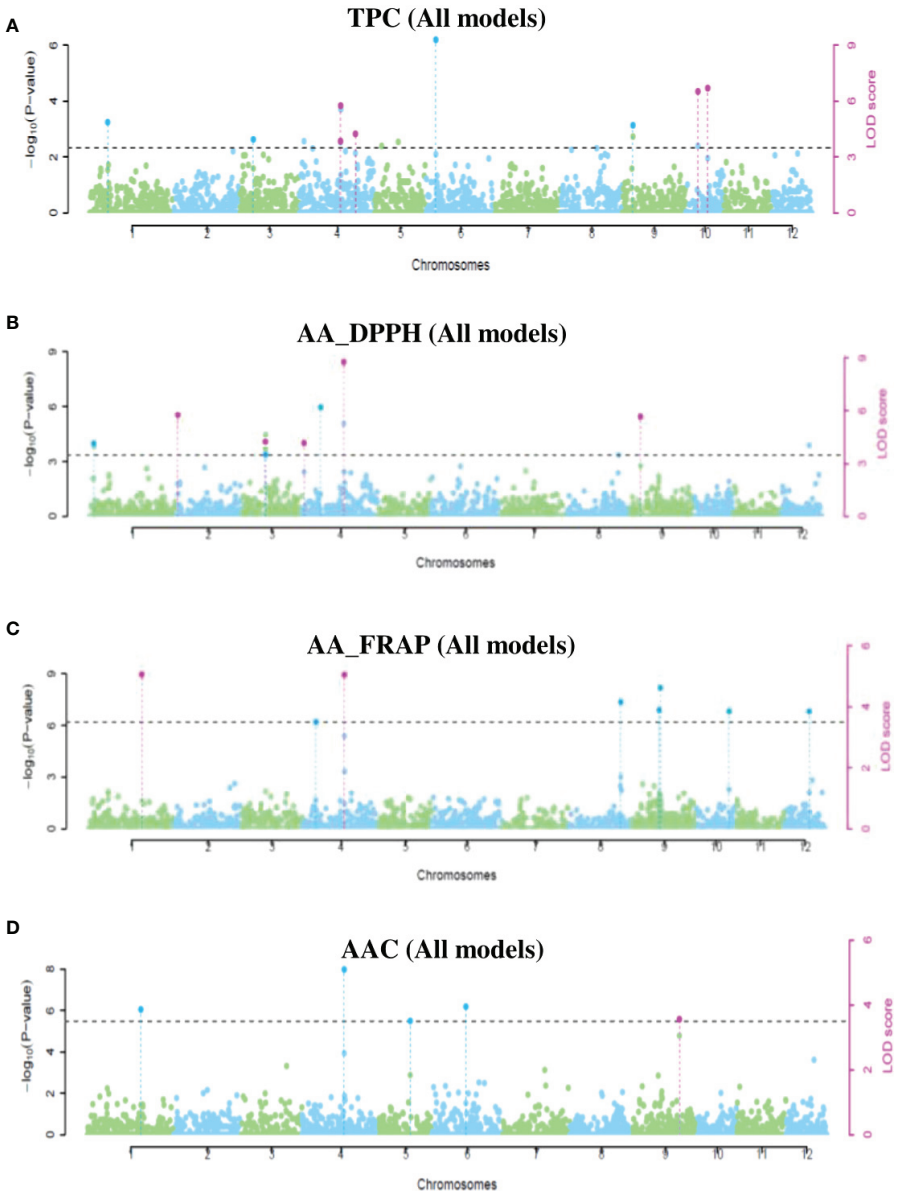


FIGURE 6
Manhattan plots showing significant QTNs for TPC **(A)**, AA_DPPH **(B)**, AA_FRAP **(C)**, and AAC **(D)** obtained using multi-locus GWAS methods in the mixed population.

analyzed accessions from AGROSAVIA's collection showed a high variation in TPC, AAC, and AA values. This variation was associated with potatoes' ploidy level and tuber colors; this collection conserves diploid and tetraploid native potatoes with a wide tuber phenotypic diversity of forms and colors (Berdugo-Cely et al., 2017). Here, the results showed that purple-blackish tetraploid and pink-red diploid potatoes had elevated levels of TPC and AAC with good antioxidant activity. In agreement with our results, previous studies reported a correlation between high levels of phenolic compounds, ascorbic acid, and antioxidant activity in potatoes and genotypes with skin and flesh dark colors tubers (Hu et al., 2012; Perla et al., 2012; Albishi et al., 2013). Skin and flesh potato colors can be considered phenotypic traits to identify genotype accessions with high nutritional values.

The ploidy level can affect TPC, AAC, and AA values. Here, diploid potatoes had a higher concentration of ascorbic acid and antioxidant activity than tetraploid genotypes. Tetraploid genotypes tended to have higher concentrations of phenolic

TABLE 2 List of candidate genes obtained from STAs and QTNs detected in common in single and multi-locus GWAS methods with significant association with the content of phenolic compounds (TPC), ascorbic acid (AAC), and Antioxidant Activity (AA) in potatoes populations.

STA/ QTN	SNP	Chrom	Position (bp)	Alleles	Number gene	Gene	Molecular Function	GWAS Model	Trait	LOD Scores	R^2 (%)
STA_QTN_1	c1_9566	1	72320482	A/G	Gen_1	Soltu.DM.01G033780.1	cysteine synthase C1	Single-locus: 3,4,6	AA_FRAP	3,5	6,6
								Multi-locus: 4,6	AA_FRAP	5,0579	10,4 - 11,0
STA_QTN_2	c2_15040	2	45649739	A/G	Gen_2	PGSC0003DMG400001451	folate carrier protein	Single-locus: 3,4,5	AA_FRAP	3,5	7,8
								Multi-locus: 4,5	TPC	4,8 - 7,7	13,9 - 22,6
STA_QTN_3	c2_48381	3	39198462	T/C	Gen_3	Soltu.DM.03G015260.1	BCL-2-associated athanogene	Single-locus: 3,6	AA_DPPH	3,5 - 3,7	0,1
								Multi-locus: 3,5	AA_DPPH	3,9 - 4,5	2,6 - 9,5
STA_QTN_4	c2_55279	3	45386107	A/C	Gen_4	Soltu.DM.03G022550.1	pectin methylesterase PCR fragment F	Single-locus: 2,4	AA_DPPH	3,6	31,1
								Single-locus: 1,3	AA_DPPH	3,9 - 4,5	0,1
								Single-locus: 1,3	AAC	3,5 - 4,5	0,1
								Multi-locus: 1,4,6	AA_DPPH	6,4205	49,7 - 63,8
								Multi-locus: 4,6	AA_FRAP	3,0091	22,7 - 25,3
								Single-locus: 1,3,4,5	AA_DPPH	4,6 - 6	0,7
								Single-locus: 1,3,4,5	AA_FRAP	5,4 - 6,8	2,1
								Single-locus: 3,4,5	AAC	3,5 - 4,4	0,2
								Single-locus: 1,5	AA_FRAP	3,5	2,4
								Single-locus: 1,3,4,5	AA_DPPH	3,5 - 4,7	4,1
								Single-locus: 1,3,4,5	AA_FRAP	4,8 - 4,7	5,1
								Single-locus: 4,5	AAC	3,9	5,1
STA_QTN_5	c1_12945	4	57899405	T/C				Multi-locus: 2, 3, 5, 6	TPC	3,5 - 4,8	3,3 - 10
					Gen_5	PGSC0003DMG400024824	Glycosyltransferase, CAZy family GT8	Multi-locus: 1,2,3,4,5,6	AA_DPPH	4,2 - 15,8	5,1 - 28,6
								Multi-locus: 1,2,3,4,5,6	AA_FRAP	3,6 - 15,4	5,1 - 32,5
								Multi-locus: 1,2,4,5,6	AA_DPPH	5,5 - 12,3	17,3 - 36,4
								Multi-locus: 1,2,3,4,5,6	AA_FRAP	3,8 - 11,3	5,2 - 31,2
								Single-locus: 4,6	AA_FRAP	3,7 - 4,0	0,2
STA_QTN_6	c2_43998	4	57899561	A/C				Single-locus: 2,3,4,6	TPC	3,5 - 4,2	4,4
								Multi-locus: 1,4	TPC	5,4 - 6,0	

(Continued)

10.3389/fpls.2022.1007104

10.3389/fpls.2022.1007104

TABLE 2 Continued

STA/ QTN	SNP	Chrom	Position (bp)	Alleles	Number gene	Gene	Molecular Function	GWAS Model	Trait	LOD Scores	R^2 (%)
											9,3 - 22,4
								Single-locus: 1,3,5	AA_DPPH	3,5 - 5,1	5,2
								Single-locus: 3,4,5	AA_FRAP	3,5 - 4,4	4,9
								Single-locus: 1,3,4,5	TPC	3,6 - 4,8	0,2
STA_QTN_7	c2_4030	9	3619115	T/C	Gen_6	PGSC0003DMG400002675	fructose-bisphosphate aldolase	Single-locus: 1,3,5	AA_DPPH	3,6 -5,2	4,8
								Single-locus: 1,3,4,5	AA_FRAP	3,6 - 5,4	4,2
								Single-locus: 3,5	TPC	3,7 - 4,4	1,3
								Multi-locus: 1,2,3	AA_DPPH	5,2 - 6,4	8,4 - 24,7
								Single-locus: 3,4	AAC	3,5 - 4,1	0,6
STA_QTN_8	c2_51155	9	49754441	C/T	Gen_7	Soltu.DM.09G021110.1	basic helix-loop-helix (bHLH) DNA-binding superfamily protein	Multi-locus: 1,2,4,5,6	AAC	3,5 - 4,7	13,6 - 30,8

Single-locus models: 1: General; 2: Additive; 3: Diplo-General; 4: Diplo-Additive; 5: 1-dom-alt; 6: 1-dom-ref; 7: 2-dom-alt; 8: 2-dom-ref. Multi-locus models: 1: mrMLM; 2: FASTmrMLM; 3: FASTmrEMMA; 4: pLARmEB; 5: pKWmEB; 6: ISIS EM-BLASSO.

contents, but this difference was not significant. Potato diploid genotypes can present up to 13 times more carotenoid content than tetraploid genotypes, suggesting that tetraploid potatoes have lower nutritional value than diploid genotypes (Lu et al., 2001). Similar results in native South American potato cultivars were reported by Brown (2008). Some reports indicate that polyploid genotypes of *Glycine max*, *Solidago canadensis*, and *Fagopyrum esculentum* generally produce a more diverse and higher secondary metabolite concentration than diploids, explained by the presence of multiple gene copies generated by chromosome and gene duplication (Zagoskina et al., 2018; Gaynor et al., 2020; Yang et al., 2021). Environmental and genotypic effects can affect the concentrations of secondary metabolites in plants (Yuan et al., 2013), and they could be species-specific. For instance, diploid genotypes of *Anchusa officinalis* and *Camellia sinensis* presented higher concentrations of secondary metabolites than polyploid genotypes (Zagoskina et al., 2018). However, other plant species did not show differences in secondary metabolite concentration associated with ploidy levels (Gaynor et al., 2020). Our results show the importance of including diploid genotypes in a potato breeding program to improve antioxidant traits for the generation of biofortified cultivars with high nutritional values.

Phenolic compounds, reducing agents with redox properties, are one of the main responsible for the antioxidant activity in plants (Roman et al., 2013). In our study, we found that TPC and AA are correlated in agreement with earlier potato studies (Hu et al., 2012; Pinhero et al., 2016; Kim et al., 2019). TPC is one of the main contributors to antioxidant activity. This result would support the use of TPC as an indicator to assess the AA in fruits and vegetables (Hesam et al., 2012). In contrast, TPC and AAC were not correlated in tetraploids and had a low correlation in diploids. Although ascorbic acid is a well-known antioxidant, we found a weak correlation between AAC and AA in tetraploids. These results coincide with studies in pineapple, papaya, plum, and tamarind, in which no correlations between AAC and AA were found (Almeida et al., 2011). In a study analyzing juices from different fruit species, the authors found that the antioxidant capacity needs the synergetic action of ascorbic acid and polyphenols; the AAC depended more on the polyphenol's concentration than on ascorbic acid content (Nowak et al., 2018). In agreement with Nowak´s study, our results suggest that the antioxidant activity in potatoes is more governed by the phenolic compounds than by the ascorbic acid; results in berries and flaxseed were similar (Silva et al., 2013; Manganaris et al., 2014).

The selection indexes (SI) allowed the choice of 20 promising genotypes from AGROSAVIA's collection (13 diploids and seven tetraploids) with mainly dark skin tuber color. Some of the selected materials can be attractive for a breeding program (And_186, And_209, And_210, Cha_49, and Cha_16) because they presented good levels of both TPC and AA. The heritability values varied between medium and high (h^2 =0.466-0.695), in agreement with previous reports

(h^2 =0.413-0.657) (Tierno and Ruiz de Galarreta, 2018). The h^2 values support using selected materials as elite parentals to initiate crossing processes in a breeding program for TPC, AA, and AAC (Tierno and Ruiz de Galarreta, 2018). In this study, we obtained an absolute value of the global concentration of phenolic compounds. In a next step, we are interested to identify the individual compounds and will analyze the concentration of each specific phenolic compound. A complete assessment of the nutritional value of this germplasm should include the analysis of other antioxidant components such as chlorogenic, vanillin, p-coumaric acid, and caffeic acid, among others.

Single and multi-locus GWAS revealed STAs and QTNs associated with TPC, AAC, and AA

In the present work, GWAS could dissect the genetic basis of the natural variation of phenolic compounds, ascorbic acid content, and antioxidant activity in potato tubers. So far, few studies have evaluated the genetic control of secondary metabolites in potatoes (Parra-Galindo et al., 2019; Parra-Galindo et al., 2021), and few reports studied the genetics of quality traits such as fry color and starch content (Schönhals et al., 2017; Byrne et al., 2020; Naeem et al., 2021). Then, until this moment, this could be the first GWAS that identified STAs and QTNs that flanked genes with a possible genetic association with TPC, AA, and AAC in potatoes using diploid and tetraploid genotypes. In other cultivated plant species, GWAS revealed genes associated with TPC in rice (Xu et al., 2016), apple (McClure et al., 2019), sorghum (Habyarimana et al., 2019; Kimani et al., 2020), barley (Han et al., 2018), and tomatoes (Ruggieri et al., 2014) and associated with AAC in tomatoes (Ruggieri et al., 2014; Sauvage et al., 2014; Ye et al., 2019).

Here, we implemented single and multi-locus GWAS analyses of the complete dataset that included diploid and tetraploid potatoes and also ran the analyses separating the potatoes by their ploidy level. The CCC has a strong population structure based on the ploidy level; this collection is structured into diploid and tetraploid potatoes (Berdugo-Cely et al., 2017). Ploidy levels have been reported as one of the main factors to explain the population structure in other potato collections (Stich et al., 2013; Hardigan et al., 2015; Berdugo-Cely et al., 2021). Identifying the genetic structure of panels used in association mapping analyses is essential to eliminate false marker-traits associations influenced by genetic relationships and population structure tendencies. These population characteristics must be included in the GWAS models to correct and exclude these false associations (Flint-Garcia et al., 2005; Zhu et al., 2008). To implement GWAS analysis, it is also necessary to identify the LD degree among SNPs in the population because these statistics would establish if the population is suitable for association mapping studies and the degree of the mapping resolution (Flint-Garcia et al., 2005; Yousaf et al., 2021). The LD mean value and the genetic diversity (Ho) of diploids contrasted with the values of the mixed and tetraploid populations. The same pattern of high LD and low Ho in diploid potatoes and low LD and high Ho in tetraploids was previously reported (Berdugo-Cely et al., 2017; Berdugo-Cely et al., 2021). As reported in other studies (Gebhardt et al., 2004; Simko et al., 2006), the LD decayed slowly in all populations, and it is explained by the clonal propagation of potatoes that limits the number of meiotic generations and, in consequence, the recombination events (Gebhardt et al., 2004; Simko et al., 2006). The LD pattern in potatoes makes possible the use of a modest number of SNPs, as included in the 8K potato SNP array, to find market-trait associations (Gebhardt et al., 2004; Simko et al., 2006). In this study, we confirmed that the number of SNPs was enough to find STAs and QTNs for the analyzed traits. However, the number of STAs and QTNs in the diploid population was minimal.

In the single and multi-locus GWAS analyses, the number of significant associations found for TPC, AAC, and AA (AA_DPPH and AA_FRAP) traits depended on the dataset/populations (mixed, diploid, and tetraploid) and the GWAS models (single and multi-locus) used. At the level of the dataset, the number of detected STAs and QTNs was higher in the mixed (in single locus GWAS) and tetraploid (in multi-locus GWAS) datasets than in the diploid dataset (in both GWAS models: single and multi-locus), probably due to the higher sample number in the mixed and tetraploid populations. The size population affects the obtention of reproducible and robust results in GWAS analyses (Uffelmann et al., 2021); for this, it would be necessary to increase the number of individuals in the diploid population to ensure that the number of genetic associations detected was not affected by this difference in population size. In sugarcane, a species with a complex population structure based on ploidy levels, GWAS identified the highest number of genetic associations in the complete dataset and the largest subpopulations. Associations detected in the subpopulations were not detected in the complete dataset, probably because of the population size and structure population (Yang et al., 2020). Similar results were found in our study; the highest number of the STAs and QTNs detected in the diploid and tetraploid populations were not detected in the complete dataset. This result suggests that GWAS cannot detect the overall genetic associations even with a correction by population structure in a mixed population. Then, GWAS in potatoes should be conducted using complete datasets and data separated by ploidy level to identify the highest number of genetic associations.

The single-locus GWAS detected more genetic associations than the multi-locus model. In contrast, the LOD scores and the phenotype variation estimated (PVE-r^2) were higher in multi-locus than in the single-locus GWAS. The conventional single-

locus GWAS, such as the general linear model (GLM) and mixed linear model (MLM), is considered less robust because it neglects the overall effects of multiple loci and needs corrections for critical values. QTLs with minor effects are sometimes not detected, and the effect estimations (r^2) are less accurate with single-locus models. GWAS multi-locus methods would have lower false-positive error and higher statistical power in detecting a higher number of genetic associations (Wang et al., 2016; Cui et al., 2018; Hu et al., 2018; Zhong et al., 2021), as reported in studies in rice and cotton (Li et al., 2018; Zhong et al., 2021). Our results differ because we used the *GWASpoly* package, which implements models with different types of polyploid gene action, including additive, simplex dominant, and duplex dominant useful in the analysis of an autopolyploid species such as potato (Rosyara et al., 2016). We also used an FDR correction instead of a Bonferroni correction because Bonferroni is a very stringent correction method that identifies a minor number of genetic associations (Merrick et al., 2022). Another difference is that only QTNs detected by at least two models in the multi-locus GWAS were kept. Variations in GWAS analyses used in this study increase the probability of detecting a significant number of genetic associations with single-locus models. The single and multi-locus GWAS detected eight QTNs/STAs in common that flanked seven genes statistically associated with TPC, AAC, and AA. Implementing single and multi-locus GWAS methods allows to improve the power and robustness of association analyses (Li et al., 2018).

The phenolic compounds and the ascorbic acids are antioxidants and co-factors of many processes in plants. Genes implicated in their synthesis involve several biotic and abiotic stress responses in plants (Martínez-Lüscher et al., 2014) in congruence with our results, where four pleiotropic genes were associated with TPC and AAC with antioxidant activity responses. In the present study, the gene *PGSC0003DMG400024824* (Chr4) with a *Glycosyltransferase* function was associated with the three analyzed traits, TPC, AAC, and AA (AA_DPPH and AA_FRAP). This gene belongs to the *UDP-glycosyltransferase (UGTs) superfamily* that codifies proteins involved in the biosynthesis of phenolic compounds such as anthocyanins and flavonoids (Bowles et al., 2005; Dong et al., 2020). In rice, glycosyltransferases were associated with heat, drought, and salt stress responses (Obaid et al., 2016; Li et al., 2020; Liu et al., 2021). The *PGSC0003DMG400002675* (Chr 9) gene with a *fructose-1, 6-bisphosphate aldolase (FBA)* function was associated with TPC and AA (AA_DPPH and AA_FRAP). Recently, a possible association of the FBA with the phenylalanine biosynthesis of a phenolic compound in *Cannabis sativa* was reported (González Camargo et al., 2022). Enzymes with these functions also are mainly reported in wheat as a response to salt, drought, heat, and low-temperature stresses (Lv et al., 2017). The *Soltu.DM.03G022550.1* (Chr3) gene with a *pectin methylesterase (PMEs)* function was associated with AAC and AA (AA_DPPH and AA_FRAP). This enzyme has been associated with the production of ascorbic acid through the D-galacturonate pathway in tomato and tobacco (Rigano et al., 2018) and is involved in pectin remodeling, disassembly of the cell wall, and heat stress responses in Arabidopsis (Huang et al., 2017; Wu et al., 2018). The *PGSC0003DMG400001451* gene (Chr 2) with *folate carrier proteins* function was associated with TPC and AA_FRAP; these proteins are involved in the downregulation of folate biosynthesis as a response to cold in rice (Neilson et al., 2011; Gorelova et al., 2017).

Another three genes detected in common among single and multi-locus GWAS were associated with only one trait. The *Soltu.DM.09G021110.1* (Chr 9) gene that transcribed for *basic helix-loop-helix (bHLH) transcription factor (TF)* was associated with AAC. The bHLH TF acts as a co-regulator in the biosynthesis of phenolic compounds such as anthocyanins (Liu et al., 2016). But, in maize and tomato, bHLH TF regulates the expression of genes involved in ascorbic acid biosynthesis through the GDP-mannose/L-galactose pathway (Ye et al., 2019; Yu et al., 2021) and causes the accumulation of ascorbic acid under salt stress in maize (Yu et al., 2021). The bHLH TF has also been associated with water, salt, and drought stress responses in tobacco, maize, and peanut, respectively (Zhao et al., 2020; Li et al., 2021; Yu et al., 2021). The *Soltu.DM.01G033780.1* (Chr 1) gene with a *cysteine synthase C1 (CSase)* function was associated with AA_FRAP. This enzyme responds to stress caused by high-salt conditions, heavy metals, and pathogen responses in alfalfa and brassica (Xie et al., 2013; Yuan et al., 2022). Finally, the *Soltu.DM.03G015260.1* (Chr 3) gene with *Bcl-2-associated athanogene (BAG) family protein* functions are involved in heat stress responses in rice (Rashid et al., 2011).

Many detected STAs/QTNs presented minor effects, suggesting that multiple genes probably control the phenotypic traits. It is recognized that phenolic compounds and ascorbic acid are quantitative traits (Habyarimana et al., 2019; Kimani et al., 2020). However, some associations explained until 63.8% of the phenotypic variation of the evaluated traits. Similarly, in blueberries, genetic factors (>40%) controlling secondary metabolite concentrations (e.g., anthocyanins, flavanols, and phenolic acids) were reported (Mengist et al., 2020). The analyzed phenotypic traits had a high heritability enabling the implementation of MAS for these attributes in potatoes that could start using the four identified pleiotropic genetic associations. According to the genotype scoring data, these markers would allow the identification of potato individuals with elevated levels of these compounds and good antioxidant activity. It is then required to design and validate these molecular markers *via* fine mapping and expression analyses to determine their efficiency in selecting promising genotypes and their role in the genetic control of the traits analyzed here. In plant breeding, validated molecular markers are used for initial screening or selecting individuals from crosses between elite parents. Marker-assisted selection can

be performed using easy and fast strategies based on PCR, such as KASP (Kompetitive Allele-Specific PCR) markers (Chen et al., 2021; Fu et al., 2021). In the future, it is necessary to select new associated genomic regions using trait evaluation data conducted in multiple locations for several years, as realized in wheat (Vikas et al., 2022) and peach (da Silva Linge et al., 2021). It is also necessary to explore other strategies, such as Genomic Selection (GS), to improve these features in potatoes.

Conclusions

Genotypic and phenotypic results revealed that the potato germplasm conserved in AGROSAVIA holds valuable and promising genotypes with high levels of TPC and AAC with good antioxidant properties, especially the diploid genotypes. We identified higher values for the studied traits than those found in other studies, probably as an effect of the reduced number of potato accessions or breeding lines evaluated in those studies (Hesam et al., 2012; Tierno et al., 2015; Pinhero et al., 2016; Silva-Beltrán et al., 2017; Kim et al., 2019; Soare et al., 2020). We were able to select 20 genotypes as candidate materials for breeding programs looking to improve these traits in potatoes. From literature reports, we confirmed the nutritional importance of potatoes over other plant species, such as pineapple, apples, carrots, onions, tamarind, and tomatoes, as a source of TPC and AAC (Leja et al., 2013; Kschonsek et al., 2018; Sagar et al., 2020). However, the genotype-environment interaction affects the production of these compounds, making sometimes invalid the comparison between studies conducted in different countries and environments (Reddivari et al., 2007; André et al., 2009). Therefore, we recommend that future research projects evaluate AGROSAVIA's collection in multiple producing regions of Colombia for several years to establish if TPC, AAC, and AA are affected by the environment and other variables. For instance, in blueberries, the secondary metabolite concentrations, despite having high heritability (>40%), were affected by the environment (Mengist et al., 2020). Evaluating morpho-agronomic and biochemical compounds in a high number of potato genotypes in the field in multiple locations for several years could be very expensive. Consequently, we recommend evaluating the CCC Core collection that includes 10% (128 genotypes) of the complete genetic diversity of the CCC (Manrique-Carpintero et al., 2022. Submitted). In this study, we also found that STAs and QTNs associated with TPC, AAC, and AA differed between mixed, diploid, and tetraploid populations and that analyses are affected by the population size and structure and the GWAS model used. GWAS in potatoes should be done in mixed ploidy populations and in populations separated by ploidy, to identify a major number of genetic associations. GWAS allowed the identification of seven candidate genes associated with phenolic compounds, ascorbic acid content, and antioxidant properties. Many of these associations were mapped in genes with molecular functions involved in abiotic stress responses. However, further functional validation is needed to confirm their biological role. In a breeding program, these associated genes would allow the implementation of MAS of potato materials with good antioxidant properties and tolerance to abiotic stresses (Burgos et al., 2019; Arruda et al., 2020; Hellmann et al., 2021; Rahman et al., 2022).

Data availability statement

The SNP data used in this manuscript were provided in the supplementary data (Supplementary Material Table S2). The data was obtained from the potato SNP array. Further queries should be directed to the corresponding author.

Author contributions

JB-C carried out phenotype and genotype statistical analyses, made graphics, and wrote the manuscript. MC-L achieved the funding for the biochemical process and contributed to revising the manuscript. RY contributed to interpreting data, editing the graphics, and writing and revising the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study received funding from the Ministerio de Agricultura de Colombia using funds from TV12 and TV19-20 transfers. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

Acknowledgments

We want to thank Raúl Iván Valbuena, curator of the potato germplasm bank of AGROSAVIA until 2021, for providing qualitative phenotype information about the potato color traits of the genotypes used in this study. We want to thank Benjamín Rojano from the National University for facilitating the biochemical analyses. We would also like to thank Andrés Cortés, Jaime Andrés Osorio, and Gina Garzón for kindly reviewing the manuscript. Finally, we want to thank the reviewers and the editor of this document, since all their comments allowed us to generate a more robust document.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1007104/full#supplementary-material

References

Agius, F., González-Lamothe, R., Caballero, J. L., Muñoz-Blanco, J., Botella, M. A., and Valpuesta, V. (2003). Engineering increased vitamin c levels in plants by overexpression of a d-galacturonic acid reductase. *Nat. Biotechnol.* 21, 177–181. doi: 10.1038/nbt777

Ah-Hen, K., Fuenzalida, C., Hess, S., Contreras, A., Vega-Gálvez, A., and Lemus-Mondaca, R. (2012). Antioxidant capacity and total phenolic compounds of twelve selected potato landrace clones grown in southern Chile. *Chil J. Agric. Res.* 72, 3–9. doi: 10.4067/s0718-58392012000100001

Albishi, T., John, J. A., Al-Khalifa, A. S., and Shahidi, F. (2013). Phenolic content and antioxidant activities of selected potato varieties and their processing by-products. *J. Funct. Foods* 5, 590–600. doi: 10.1016/j.jff.2012.11.019

Almeida, M. M. B., de Sousa, P. H. M., Arriaga, Â.M.C., do Prado, G. M., Magalhães, C. E., de, C., et al. (2011). Bioactive compounds and antioxidant activity of fresh exotic fruits from northeastern Brazil. *Food Res. Int.* 44, 2155–2159. doi: 10.1016/j.foodres.2011.03.051

André, C., Oufir, M., Hoffmann, L., Hausman, J.-F., Rogez, H., Larondelle, Y., et al. (2009). Influence of environment and genotype on polyphenol compounds and *in vitro* antioxidant capacity of native Andean potatoes (Solanum tuberosum l.). *J. Food Composition Anal.* 22, 517–524. doi: 10.1016/j.jfca.2008.11.010

Arruda, H. S., Neri-Numa, I. A., Kido, L. A., Maróstica Júnior, M. R., and Pastore, G. M. (2020). Recent advances and possibilities for the use of plant phenolic compounds to manage ageing-related diseases. *J. Funct. Foods* 75, 104203. doi: 10.1016/J.JFF.2020.104203

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Society. Ser. B (Methodological)* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Benzie, I. F. F., and Strain, J. J. (1996). The ferric reducing ability of plasma (FRAP) as a measure of "Antioxidant power": The FRAP assay. *Anal. Biochem.* 239, 70–76. doi: 10.1006/abio.1996.0292

Berdugo-Cely, J. A., Martínez-Moncayo, C., and Lagos-Burbano, T. C. (2021). Genetic analysis of a potato (Solanum tuberosum l.) breeding collection for southern Colombia using single nucleotide polymorphism (SNP) markers. *PloS One* 16, e0248787. doi: 10.1371/journal.pone.0248787

Berdugo-Cely, J., Valbuena, R. I., Sánchez-Betancourt, E., Barrero, L. S., and Yockteng, R. (2017). Genetic diversity and association mapping in the Colombian central collection of solanum tuberosum l. andigenum group using SNPs markers. *PloS One* 12, e0173039. doi: 10.1371/journal.pone.0173039

Bowles, D., Isayenkova, J., Lim, E. K., and Poppenberger, B. (2005). Glycosyltransferases: managers of small molecules. *Curr. Opin. Plant Biol.* 8, 254–263. doi: 10.1016/J.PBI.2005.03.007

Bozonet, S. M., and Carr, A. C. (2019). The role of physiological vitamin c concentrations healthy individuals. *Nutrients* 11, 1363. doi: 10.3390/NU11061363

Brand-Williams, W., Cuvelier, M. E., and Berset, C. (1995). Use of a free radical method to evaluate antioxidant activity. *LWT - Food Sci. Technol.* 28, 25–30. doi: 10.1016/S0023-6438(95)80008-5

Brown, C. R. (2008). Breeding for phytonutrient enhancement of potato. *Am. J. Potato Res.* 85, 298–307. doi: 10.1007/S12230-008-9028-0/FIGURES/8

Burgos, G., zum Felde, T., Andre, C., and Kubow, S. (2019). The potato and its contribution to the human diet and health. *Potato Crop: Its Agricultural Nutr. Soc. Contribution to Humankind* 37–74. doi: 10.1007/978-3-030-28683-5_2/FIGURES/5

Byrne, S., Meade, F., Mesiti, F., Griffin, D., Kennedy, C., and Milbourne, D. (2020). Genome-wide association and genomic prediction for fry color in potato. *Agronomy* 10, 90. doi: 10.3390/AGRONOMY10010090

Caruso, P., Russo, M. P., Caruso, M., di Guardo, M., Russo, G., Fabroni, S., et al. (2021). A transcriptional analysis of the genes involved in the ascorbic acid pathways based on a comparison of the juice and leaves of navel and anthocyanin-rich sweet orange varieties. *Plants* 10 (7). doi: 10.3390/PLANTS10071291/S1

Chambial, S., Dwivedi, S., Shukla, K. K., John, P. J., and Sharma, P. (2013). Vitamin c in disease prevention and cure: An overview. *Indian J. Clin. Biochem.* 28, 314–328. doi: 10.1007/s12291-013-0375-3

Chen, Z., Tang, D., Ni, J., Li, P., Wang, L., Zhou, J., et al. (2021). Development of genic KASP SNP markers from RNA-seq data for map-based cloning and marker-assisted selection in maize. *BMC Plant Biol.* 21, 1–11. doi: 10.1186/S12870-021-02932-8/FIGURES/4

Cory, H., Passarelli, S., Szeto, J., Tamez, M., and Mattei, J. (2018). The role of polyphenols in human health and food systems: A mini-review. *Front. Nutr.* 5. doi: 10.3389/fnut.2018.00087

Cui, Y., Zhang, F., and Zhou, Y. (2018). The application of multi-locus GWAS for the detection of salt-tolerance loci in rice. *Front. Plant Sci.* 9. doi: 10.3389/FPLS.2018.01464/BIBTEX

da Silva Linge, C., Cai, L., Fu, W., Clark, J., Worthington, M., Rawandoozi, Z., et al. (2021). Multi-locus genome-wide association studies reveal fruit quality hotspots in peach genome. *Front. Plant Sci.* 12. doi: 10.3389/FPLS.2021.644799

De Haan, S., and Rodriguez, F. (2016). Potato origin and production. In: J. Singh and L. Kaur (eds). *Advances in potato chemistry and technology* (Elsevier Inc., London, GB) doi: 10.1016/B978-0-12-800002-1.00001-7

Dong, N. Q., Sun, Y., Guo, T., Shi, C. L., Zhang, Y. M., Kan, Y., et al. (2020). UDP-Glucosyltransferase regulates grain size and abiotic stress tolerance associated with metabolic flux redirection in rice. *Nat. Commun.* 11, 1–16. doi: 10.1038/s41467-020-16403-5

Earl, D. A., and vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the evanno method. *Conserv. Genet. Resour* 4, 359–361. doi: 10.1007/s12686-011-9548-7

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x

FAOSTAT (2022) *FAOSTAT*. Available at: https://www.fao.org/faostat/en/#data/QCL/visualize (Accessed June 30, 2022).

Felcher, K. J., Coombs, J. J., Massa, A. N., Hansey, C. N., Hamilton, J. P., Veilleux, R. E., et al. (2012). Integration of two diploid potato linkage maps with the potato genome sequence. *PloS One* 7, e36347. doi: 10.1371/journal.pone.0036347

Flint-Garcia, S. A., Thuillet, A. C., Yu, J., Pressoir, G., Romero, S. M., Mitchell, S. E., et al. (2005). Maize association population: a high-resolution platform for quantitative trait locus dissection. *Plant J.* 44, 1054–1064. doi: 10.1111/J.1365-313X.2005.02591.X

Fraga, C. G., Croft, K. D., Kennedy, D. O., and Tomás-Barberán, F. A. (2019). The effects of polyphenols and other bioactives on human health. *Food Funct.* 10, 514–528. doi: 10.1039/C8FO01997E

Fu, Y., Mason, A. S., Zhang, Y., and Yu, H. (2021). Identification and development of KASP markers for novel mutant BnFAD2 alleles associated with elevated oleic acid in brassica napus. *Front. Plant Sci.* 12. doi: 10.3389/FPLS.2021.715633/BIBTEX

Gaynor, M. L., Lim-Hing, S., and Mason, C. M. (2020). Impact of genome duplication on secondary metabolite composition in non-cultivated species: a systematic meta-analysis. *Ann. Bot.* 126, 363–376. doi: 10.1093/AOB/MCAA107

Gebhardt, C., Ballvora, A., Walkemeier, B., Oberhagemann, P., and Schüler, K. (2004). Assessing genetic potential in germplasm collections of crop plants by marker-trait association: a case study for potatoes with quantitative variation of resistance to late blight and maturity type. *Mol. Breed.* 13, 93–102. doi: 10.1023/B:MOLB.0000012878.89855.DF

GenomeStudio Software (2022). Available at: https://www.illumina.com/techniques/microarrays/array-data-analysis-experimental-design/genomestudio.html (Accessed May 3, 2022).

Gómez, R. (2014) *Guía para las caracterizaciones morfológicas básicas en colecciones de papas nativas*. Available at: https://www.researchgate.net/publication/267991711.

González Camargo, F. D., Santamaría, M., Cala, M., Guevara Suárez, M. I., Restrepo Restrepo, S., Corujo, M., et al. (2022) *Genome-scale metabolic reconstruction of cannabis sativa and validation with non-targeted LC-MS based leaf metabolomics data*. Available at: http://hdl.handle.net/1992/59262.

Gorelova, V., Ambach, L., Rébeillé, F., Stove, C., and van der Straeten, D. (2017). Folates in plants: Research advances and progress in crop biofortification. *Front. Chem.* 5. doi: 10.3389/FCHEM.2017.00021

Goyani, Z. (2021). Package "selection.index". *Title Anal. Selection Index Plant Breed.* doi: 10.1111/j.1469

Grun, P. (1990). The evolution of cultivated potatoes. *Economic Bot.* 44, 39–55. doi: 10.1007/BF02860474

Habyarimana, E., Dall'Agata, M., de Franceschi, P., and Baloch, F. S. (2019). Genome-wide association mapping of total antioxidant capacity, phenols, tannins, and flavonoids in a panel of sorghum bicolor and s. bicolor × s. halepense populations using multi-locus models. *PloS One* 14, e0225979. doi: 10.1371/journal.pone.0225979

Hamilton, J. P., Hansey, C. N., Whitty, B. R., Stoffel, K., Massa, A. N., Van Deynze, A., et al. (2011). Single nucleotide polymorphism discovery in elite north american potato germplasm. *BMC Genomics* 12, 302. doi: 10.1186/1471-2164-12-302

Han, Z., Zhang, J., Cai, S., Chen, X., Quan, X., and Zhang, G. (2018). Association mapping for total polyphenol content, total flavonoid content and antioxidant activity in barley. *BMC Genomics* 19, 81. doi: 10.1186/s12864-018-4483-6

Hardigan, M. A., Bamberg, J., Buell, C. R., and Douches, D. S. (2015). Taxonomy and genetic differentiation among wild and cultivated germplasm of solanum sect. petota. *Plant Genome* 8, plantgenome2014.06.0025. doi: 10.3835/plantgenome2014.06.0025

Hawkes, J.G. (1990). *The potato: evolution, biodiversity and genetic resources* (London: Belhaven Press a division of Pinter Publishers).

Hejtmánková, K., Pivec, V., Trnková, E., Hamouz, K., and Lachman, J. (2009). Quality of coloured varieties of potatoes. *Czech J. Food Sci.* 27, S310–S313. doi: 10.17221/605-cjfs

Hellmann, H., Goyer, A., and Navarre, D. A. (2021). Antioxidants in potatoes: A functional view on one of the major food crops worldwide. *Molecules* 26, 1–29. doi: 10.3390/MOLECULES26092446

Hesam, F., Balali, G. R., and Tehrani, R. T. (2012). Evaluation of antioxidant activity of three common potato (Solanum tuberosum) cultivars in Iran. *Avicenna J. Phytomed* 2, 79–85.

Huang, Y. C., Wu, H. C., Wang, Y., Liu, C. H., Lin, C. C., Luo, D. L., et al. (2017). PECTIN METHYLESTERASE34 contributes to heat tolerance through its role in promoting stomatal movement. *Plant Physiol.* 174, 748–763. doi: 10.1104/PP.17.00335

Hu, C., Tsao, R., Liu, R., Alan Sullivan, J., and McDonald, M. R. (2012). Influence of cultivar and year on phytochemical and antioxidant activity of potato (Solanum tuberosum l.) in Ontario. *Can. J. Plant Sci.* 92, 485–493. doi: 10.4141/CJPS2011-212/ASSET/IMAGES/CJPS2011-212TAB5.GIF

Hu, X., Zuo, J., Wang, J., Liu, L., Sun, G., Li, C., et al. (2018). Multi-locus genome-wide association studies for 14 main agronomic traits in barley. *Front. Plant Sci.* 871. doi: 10.3389/FPLS.2018.01683/BIBTEX

Jombart, T., and Ahmed, I. (2011). Adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* 27, 3070–3071. doi: 10.1093/bioinformatics/btr521

Kamvar, Z. N., Tabima, J. F., and Grünwald, N. J. (2014). Poppr: an r package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2, e281. doi: 10.7717/peerj.281

Kassambara, A., and Mundt, F. (2020) *Factoextra - extract and visualize the results of multivariate data analyses*. Available at: https://CRAN.R-project.org/package=factoextra (Accessed December 27, 2020).

Kelebek, H., Selli, S., Canbas, A., Cabaroglu, T., Singleton, V. L., Rossi, J. A., et al. (2009). HPLC determination of organic acids, sugars, phenolic compositions and antioxidant capacity of orange juice and orange wine made from a Turkish cv. kozan. *LWT - Food Sci. Technol.* 239, 144 LP–158. doi: 10.1016/S0023-6438(95)80008-5

Kimani, W., Zhang, L.-M., Wu, X.-Y., Hao, H.-Q., and Jing, H.-C. (2020). Genome-wide association study reveals that different pathways contribute to grain quality variation in sorghum (Sorghum bicolor). *BMC Genomics* 21, 112. doi: 10.1186/s12864-020-6538-8

Kim, J., Soh, S. Y., Bae, H., and Nam, S.-Y. (2019). Antioxidant and phenolic contents in potatoes (Solanum tuberosum l.) and micropropagated potatoes. *Appl. Biol. Chem.* 62, 17. doi: 10.1186/s13765-019-0422-8

Kowalczewski, P.Ł., Olejnik, A., Białas, W., Rybicka, I., Zielińska-Dawidziak, M., Siger, A., et al. (2019). The nutritional value and biological activity of concentrated protein fraction of potato juice. *Nutrients* 11, 1523. doi: 10.3390/nu11071523

Kruijer, W. (2019). Package "heritability" title marker-based estimation of heritability using individual plant or plot data. *Genetics* 199, 379–398. doi: 10.1534/GENETICS.114.167916/-/DC1

Kschonsek, J., Wolfram, T., Stöckl, A., and Böhm, V. (2018). Polyphenolic compounds analysis of old and new apple cultivars and contribution of polyphenolic profile to the *In vitro* antioxidant capacity. *Antioxidants (Basel)* 7, 20. doi: 10.3390/antiox7010020

Leja, M., Kamińska, I., Kramer, M., Maksylewicz-Kaul, A., Kammerer, D., Carle, R., et al. (2013). The content of phenolic compounds and radical scavenging activity varies with carrot origin and root color. *Plant Foods Hum. Nutr.* 68, 163–170. doi: 10.1007/s11130-013-0351-3

Li, C., Fu, Y., Sun, R., Wang, Y., and Wang, Q. (2018). Single-locus and multi-locus genome-wide association studies in the genetic dissection of fiber quality traits in upland cotton (gossypium hirsutum l.). *Front. Plant Sci.* 9. doi: 10.3389/FPLS.2018.01083/FULL

Li, Y., Liu, F., Li, P., Wang, T., Zheng, C., and Hou, B. (2020). An arabidopsis cytokinin-modifying glycosyltransferase UGT76C2 improves drought and salt tolerance in rice. *Front. Plant Sci.* 11. doi: 10.3389/FPLS.2020.560696/BIBTEX

Liu, Q., Dong, G. R., Ma, Y. Q., Zhao, S. M., Liu, X., Li, X. K., et al. (2021). Rice glycosyltransferase gene UGT85E1 is involved in drought stress tolerance through enhancing abscisic acid response. *Front. Plant Sci.* 12. doi: 10.3389/FPLS.2021.790195/BIBTEX

Liu, Y., Lin-Wang, K., Deng, C., Warran, B., Wang, L., Yu, B., et al. (2015). Comparative transcriptome analysis of white and purple potato to identify genes involved in anthocyanin biosynthesis. *PloS One* 10, e0129148. doi: 10.1371/journal.pone.0129148

Liu, Y., Lin-Wang, K., Espley, R., Wang, L., Yang, H., Yu, B., et al. (2016). Functional diversification of the potato R2R3 MYB anthocyanin activators AN1, MYBA1, and MYB113 and their interaction with basic helix-loop-helix cofactors. *J. Exp. Bot.* 67, 2159–2176. doi: 10.1093/JXB/ERW014

Liu, Y., Liu, C., and Li, J. (2020). Comparison of vitamin c and its derivative antioxidant activity: Evaluated by using density functional theory. *ACS Omega* 5, 25467–25475. doi: 10.1021/acsomega.0c04318

Liu, F., Yang, Y., Gao, J., Ma, C., and Bi, Y. (2018). A comparative transcriptome analysis of a wild purple potato and its red mutant provides insight into the mechanism of anthocyanin transformation. *PloS One* 13, e0191406. doi: 10.1371/journal.pone.0191406

Li, C., Yan, C., Sun, Q., Wang, J., Yuan, C., Mou, Y., et al. (2021). The bHLH transcription factor AhbHLH112 improves the drought tolerance of peanut. *BMC Plant Biol.* 21, 1–12. doi: 10.1186/S12870-021-03318-6/FIGURES/7

Lorence, A., Chevone, B. I., Mendes, P., and Nessler, C. L. (2004). Myo-inositol oxygenase offers a possible entry point into plant ascorbate biosynthesis. *Plant Physiol.* 134, 1200–1205. doi: 10.1104/pp.103.033936

Lourenço, S. C., Moldão-Martins, M., and Alves, V. D. (2019). Antioxidants of natural plant origins: From sources to food industry applications. *Molecules* 24, 1–25. doi: 10.3390/molecules24224132

Lu, W., Haynes, K., Wiley, E., and Clevidence, B. (2001). Carotenoid content and color in diploid potatoes. *J. Am. Soc. Hortic. Sci. jashs* 126, 722–726. doi: 10.21273/JASHS.126.6.722

Lv, G. Y., Guo, X. G., Xie, L. P., Xie, C. G., Zhang, X. H., Yang, Y., et al. (2017). Molecular characterization, gene evolution, and expression analysis of the fructose-1, 6-bisphosphate aldolase (FBA) gene family in wheat (Triticum aestivum l.). *Front. Plant Sci.* 8. doi: 10.3389/FPLS.2017.01030/FULL

Manganaris, G. A., Goulas, V., Vicente, A. R., and Terry, L. A. (2014). Berry antioxidants: small fruits providing large benefits. *J. Sci. Food Agric.* 94, 825–833. doi: 10.1002/jsfa.6432

Manrique-Carpintero, N. C., Berdugo-Cely, J. A., Cerón-Souza, I., Lasso-Paredes, Z., Reyes-Herrera, P., and Yockteng, R. (2022). Defining a diverse core collection of the Colombian Central Collection of Potato: A tool for advancing research. *Front Plant Sci.*.

Martínez-Lüscher, J., Sánchez-Díaz, M., Delrot, S., Aguirreolea, J., Pascual, I., and Gomès, E. (2014). Ultraviolet-b radiation and water deficit interact to alter flavonol and anthocyanin profiles in grapevine berries through transcriptomic regulation. *Plant Cell Physiol.* 55, 1925–1936. doi: 10.1093/pcp/pcu121

McClure, K. A., Gong, Y., Song, J., Vinqvist-Tymchuk, M., Campbell Palmer, L., Fan, L., et al. (2019). Genome-wide association studies in apple reveal loci of large effect controlling apple polyphenols. *Hortic. Res.* 6, 107. doi: 10.1038/s41438-019-0190-y

Mengist, M. F., Grace, M. H., Xiong, J., Kay, C. D., Bassil, N., Hummer, K., et al. (2020). Diversity in metabolites and fruit quality traits in blueberry enables ploidy and species differentiation and establishes a strategy for future genetic studies. *Front. Plant Sci.* 11. doi: 10.3389/FPLS.2020.00370/BIBTEX

Merrick, L. F., Burke, A. B., Zhang, Z., and Carter, A. H. (2022). Comparison of single-trait and multi-trait genome-wide association models and inclusion of correlated traits in the dissection of the genetic architecture of a complex trait in a breeding program. *Front. Plant Sci.* 12. doi: 10.3389/FPLS.2021.772907/BIBTEX

Mishra, T., Raigond, P., Thakur, N., Dutt, S., and Singh, B. (2020). Recent updates on healthy phytoconstituents in potato: a nutritional depository. *Potato Res.* 63, 323–343. doi: 10.1007/s11540-019-09442-z

Murniece, I., Tomsone, L., Skrabule, I., and Vaivode, A. (2014). Carotenoids and total phenolic content in potatoes with different flesh colour. *Foodbalt* 206–211.

Naeem, M., Demirel, U., Yousaf, M. F., Caliskan, S., Caliskan, M. E., and Wehling, P. (2021). Overview on domestication, breeding, genetic gain and improvement of tuber quality traits of potato using fast forwarding technique (GWAS): A review. *Plant Breed.* 140, 519–542. doi: 10.1111/PBR.12927

Nagel, M., Dulloo, M., Bissessur, P., Gavrilenko, T., Bamberg, J., Ellis, D., et al. (2022). Global strategy for the conservation of potato. *Global Crop Diversity Trust*. Bonn, Germany. doi: 10.5447/IPK/2022/29

Navarre, D. A., Goyer, A., and Shakya, R. (2009). Nutritional value of potatoes: Vitamin, phytonutrient, and mineral content. *Adv. Potato Chem. Technol.*, 395–424. doi: 10.1016/B978-0-12-374349-7.00014-3

Neilson, K. A., Mariani, M., and Haynes, P. A. (2011). Quantitative proteomic analysis of cold-responsive proteins in rice. *Proteomics* 11, 1696–1706. doi: 10.1002/PMIC.201000727

Nowak, D., Gośliński, M., Wojtowicz, E., and Przygoński, K. (2018). Antioxidant properties and phenolic compounds of vitamin c-rich juices. *J. Food Sci.* 83, 2237–2246. doi: 10.1111/1750-3841.14284

Obaid, A. Y., Sabir, J. S. M., Atef, A., Liu, X., Edris, S., El-Domyati, F. M., et al. (2016). Analysis of transcriptional response to heat stress in rhazya stricta. *BMC Plant Biol.* 16, 1–18. doi: 10.1186/S12870-016-0938-6

Parra-Galindo, M.-A., Piñeros-Niño, C., Soto-Sedano, J. C., and Mosquera-Vasquez, T. (2019). Chromosomes I and X harbor consistent genetic factors associated with the anthocyanin variation in potato. *Agronomy* 9, 366. doi: 10.3390/agronomy9070366

Parra-Galindo, M. A., Soto-Sedano, J. C., Mosquera-Vásquez, T., and Roda, F. (2021). Pathway-based analysis of anthocyanin diversity in diploid potato. *PloS One* 16, e0250861. doi: 10.1371/journal.pone.0250861

Patil, V. M., and Masand, N. (2018). Anticancer potential of flavonoids: Chemistry, biological activities, and future perspectives. *Stud. Natural Products Chem.* 59, 401–430. doi: 10.1016/B978-0-444-64179-3.00012-8

Pawlowska, E., Szczepanska, J., and Blasiak, J. (2019). Pro- and antioxidant effects of vitamin c in cancer in correspondence to its dietary and pharmacological concentrations. *Oxid. Med. Cell Longev* 2019, 7286737. doi: 10.1155/2019/7286737

Perla, V., Holm, D. G., and Jayanty, S. S. (2012). Effects of cooking methods on polyphenols, pigments and antioxidant activity in potato tubers. *LWT - Food Sci. Technol.* 45, 161–171. doi: 10.1016/j.lwt.2011.08.005

PGSC Data Download. Available at: http://solanaceae.plantbiology.msu.edu/pgsc_download.shtml (Accessed March 16, 2021).

Phan, H. T. T., Rybak, K., Bertazzoni, S., Furuki, E., Dinglasan, E., Hickey, L. T., et al. (2018). Novel sources of resistance to septoria nodorum blotch in the vavilov wheat collection identified by genome-wide association studies. *Theor. Appl. Genet.* 131, 1223–1238. doi: 10.1007/S00122-018-3073-Y/FIGURES/6

Pinhero, R. G., Tsao, R., Liu, Q., Sullivan, J. A., Bizimungu, B., Y. Yada, R., et al. (2016). Protein and phenolic contents and antioxidant activities of 14 early maturing potatoes as affected by processing. *Am. J. Plant Sci.* 07, 69–81. doi: 10.4236/ajps.2016.71008

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945 LP–959. doi: 10.1093/genetics/155.2.945

R (2022) *R: The r project for statistical computing*. Available at: https://www.r-project.org/ (Accessed February 7, 2021).

Rahman, M. M., Rahaman, M. S., Islam, M. R., Rahman, F., Mithi, F. M., Alqahtani, T., et al. (2022). Role of phenolic compounds in human disease: Current knowledge and future prospects. *Molecules* 27, 233. doi: 10.3390/MOLECULES27010233

Rashid, M. R., Shinan, D., Zulfiqar, A., Azeem, I. K., and Hong, S. (2011). Identification and characterization of the bcl-2-associated athanogene (BAG) protein family in rice. *Afr J. Biotechnol.* 11, 88–98. doi: 10.5897/ajb11.3474

Reddivari, L., Hale, A. L., and Miller, J. C. (2007). Genotype, location, and year influence antioxidant activity, carotenoid content, phenolic content, and composition in specialty potatoes. *J. Agric. Food Chem.* 55, 8073–8079. doi: 10.1021/jf071543w

Riaz, A., Athiyannan, N., Periyannan, S. K., Afanasenko, O., Mitrofanova, O. P., Platz, G. J., et al. (2018). Unlocking new alleles for leaf rust resistance in the vavilov wheat collection. *Theor. Appl. Genet.* 131, 127–144. doi: 10.1007/S00122-017-2990-5/FIGURES/2

Rigano, M. M., Lionetti, V., Raiola, A., Bellincampi, D., and Barone, A. (2018). Pectic enzymes as potential enhancers of ascorbic acid production through the d-galacturonate pathway in solanaceae. *Plant Sci.* 266, 55–63. doi: 10.1016/J.PLANTSCI.2017.10.013

Roman, I., Stănilă, A., and Stănilă, S. (2013). Bioactive compounds and antioxidant activity of Rosa canina l. biotypes from spontaneous flora of Transylvania. *Chem. Cent J.* 7, 1–10. doi: 10.1186/1752-153X-7-73/FIGURES/9

Rosyara, U. R., de Jong, W. S., Douches, D. S., and Endelman, J. B. (2016). Software for genome-wide association studies in autopolyploids and its application to potato. *Plant Genome* 9, plantgenome2015.08.0073. doi: 10.3835/plantgenome2015.08.0073

R Stats (2022). Available at: https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html (Accessed November 30, 2022).

Ruggieri, V., Francese, G., Sacco, A., D'Alessandro, A., Rigano, M. M., Parisi, M., et al. (2014). An association mapping approach to identify favourable alleles for tomato fruit quality breeding. *BMC Plant Biol.* 14, 337. doi: 10.1186/s12870-014-0337-9

Sagar, N. A., Pareek, S., and Gonzalez-Aguilar, G. A. (2020). Quantification of flavonoids, total phenols and antioxidant properties of onion skin: a comparative study of fifteen Indian cultivars. *J. Food Sci. Technol.* 57, 2423–2432. doi: 10.1007/s13197-020-04277-w

Sauvage, C., Segura, V., Bauchet, G., Stevens, R., Do, P. T., Nikoloski, Z., et al. (2014). Genome-wide association in tomato reveals 44 candidate loci for fruit metabolic traits. *Plant Physiol.* 165, 1120 LP–1132. doi: 10.1104/pp.114.241521

Schönhals, E. M., Ding, J., Ritter, E., Paulo, M. J., Cara, N., Tacke, E., et al. (2017). Physical mapping of QTL for tuber yield, starch content and starch yield in tetraploid potato (Solanum tuberosum l.) by means of genome wide genotyping by sequencing and the 8.3 K SolCAP SNP array. *BMC Genomics* 18 (642), 1–20. doi: 10.1186/S12864-017-3979-9

Sharma, S. K., Bolser, D., de Boer, J., Sønderkær, M., Amoros, W., Carboni, M. F., et al. (2013). Construction of reference chromosome-scale pseudomolecules for potato: Integrating the potato genome with genetic and physical maps. *G3: Genes|Genomes|Genetics* 3, 2031 LP–2047. doi: 10.1534/g3.113.007153

Silva-Beltrán, N. P., Chaidez-Quiroz, C., López-Cuevas, O., Ruiz-Cruz, S., López-Mata, M. A., Del-Toro-Sánchez, C., et al. (2017). Phenolic compounds of potato peel extracts: Their antioxidant activity and protection against human enteric viruses. *J. Microbiol. Biotechnol.* 27, 234–241. doi: 10.4014/jmb.1606.06007

Silva, F. G. D., O'Callagahan, Y., O'Brien, N. M., and Netto, F. M. (2013). Antioxidant capacity of flaxseed products: The effect of *In vitro* digestion. *Plant Foods Hum. Nutr.* 68, 24–30. doi: 10.1007/s11130-012-0329-6

Simko, I., Haynes, K. G., and Jones, R. W. (2006). Assessment of linkage disequilibrium in potato genome with single nucleotide polymorphism markers. *Genetics* 173, 2237–2245. doi: 10.1534/GENETICS.106.060905

Singh, J., and Kaur, L. (2009). "Introduction," in *Advances in Potato Chemistry and Technology* Eds. J. Singh and T. Kaur (San Diego: Academic Press), ix–xii. L. B. T.-A. @ in P. C. doi: 10.1016/B978-0-12-374349-7.00019-2

Singleton, V. L., and Rossi, J. A. (1965). Colorimetry of total phenolics with phosphomolybdic-phosphotungstic acid reagents. *Am. J. Enol Vitic* 16, 144 LP–158.

Soare, R., Dinu, M., Babeanu, C., and Soare, M. (2020). Evaluation and comparison of antioxidant activity and biochemical compounds in some coloured potato cultivars. *Plant Soil Environ.* 66, 281–286. doi: 10.17221/202/2020-PSE

Spooner, D. M., Ghislain, M., Simon, R., Jansky, S. H., and Gavrilenko, T. (2014). Systematics, diversity, genetics, and evolution of wild and cultivated potatoes. *Botanical Rev.* 80, 283–383. doi: 10.1007/S12229-014-9146-Y

Stich, B., Urbany, C., Hoffmann, P., and Gebhardt, C. (2013). Population structure and linkage disequilibrium in diploid and tetraploid potato revealed by genome-wide high-density genotyping using the SolCAP SNP array. *Plant Breed.* 132, 718–724. doi: 10.1111/pbr.12102

Stushnoff, C., Ducreux, L. J. M., Hancock, R. D., Hedley, P. E., Holm, D. G., McDougall, G. J., et al. (2010). Flavonoid profiling and transcriptome analysis reveals new gene–metabolite correlations in tubers of solanum tuberosum l. *J. Exp. Bot.* 61, 1225–1238. doi: 10.1093/jxb/erp394

Tierno, R., Hornero-Méndez, D., Gallardo-Guerrero, L., López-Pardo, R., and de Galarreta, J. I. R. (2015). Effect of boiling on the total phenolic, anthocyanin and carotenoid concentrations of potato tubers from selected cultivars and introgressed breeding lines from native potato species. *J. Food Composition Anal.* 41, 58–65. doi: 10.1016/j.jfca.2015.01.013

Tierno, R., and Ruiz de Galarreta, J. I. (2018). Genetic analyses for bioactive potential in tetraploid potato (Solanum tuberosum l.). *Plant Breed.* 137, 210–219. doi: 10.1111/PBR.12570

Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., et al. (2021). Genome-wide association studies. *Nat. Rev. Methods Primers* 1, 1–21. doi: 10.1038/s43586-021-00056-9

USDA (2022) *FoodData central*. Available at: https://fdc.nal.usda.gov/fdc-app.html#/food-details/2346402/nutrients (Accessed November 30, 2022).

Valcarcel, J., Reilly, K., Gaffney, M., and O'Brien, N. M. (2016). Levels of potential bioactive compounds including carotenoids, vitamin C and phenolic compounds, and expression of their cognate biosynthetic genes vary significantly in different varieties of potato (Solanum tuberosum L.) grown under uniform cultural con. *J. Sci. Food Agric.* 96, 1018–1026. doi: 10.1002/jsfa.7189

Valiñas, M. A., Lanteri, M. L., ten Have, A., and Andreu, A. B. (2015). Chlorogenic acid biosynthesis appears linked with suberin production in potato tuber (Solanum tuberosum). *J. Agric. Food Chem.* 63, 4902–4913. doi: 10.1021/jf505777p

Vikas, V. K., Pradhan, A. K., Budhlakoti, N., Mishra, D. C., Chandra, T., Bhardwaj, S. C., et al. (2022). Multi-locus genome-wide association studies (ML-GWAS) reveal novel genomic regions associated with seedling and adult plant stage leaf rust resistance in bread wheat (Triticum aestivum l.). *Heredity 2022 128:6* 128, 434–449. doi: 10.1038/s41437-022-00525-1

Vogt, T. (2010). Phenylpropanoid biosynthesis. *Mol. Plant* 3, 2–20. doi: 10.1093/mp/ssp106

Voorrips, R. E., Gort, G., and Vosman, B. (2011). Genotype calling in tetraploid species from bi-allelic marker data using mixture models. *BMC Bioinf.* 12, 172. doi: 10.1186/1471-2105-12-172

Vos, P. G., Paulo, M. J., Voorrips, R. E., Visser, R. G. F., van Eck, H. J., and van Eeuwijk, F. A. (2017). Evaluation of LD decay and various LD-decay estimators in simulated and SNP-array data of tetraploid potato. *Theor. Appl. Genet.* 130, 123–135. doi: 10.1007/S00122-016-2798-8/FIGURES/4

Vos, P. G., Uitdewilligen, J. G. A. M. L., Voorrips, R. E., Visser, R. G. F., and van Eck, H. J. (2015). Development and analysis of a 20K SNP array for potato (Solanum tuberosum): an insight into the breeding history. *Theor. Appl. Genet.* 128, 2387–2401. doi: 10.1007/s00122-015-2593-y

Wang, S. B., Feng, J. Y., Ren, W. L., Huang, B., Zhou, L., Wen, Y. J., et al. (2016). Improving power and accuracy of genome-wide association studies *via* a multi-locus mixed linear model methodology. *Sci. Rep. 2016 6:1* 6, 1–10. doi: 10.1038/srep19444

Wheeler, G. L., Jones, M. A., and Smirnoff, N. (1998). The biosynthetic pathway of vitamin c in higher plants. *Nature* 393, 365–369. doi: 10.1038/30728

Wheeler, G. L., Jones, M. A., Smirnoff, N., Wolucka, B. A., Van Montagu, M., Agius, F., et al. (2003). Engineering increased vitamin c levels in plants by overexpression of a d-galacturonic acid reductase. *J. Biol. Chem.* 393, 365–369. doi: 10.1074/jbc.M309135200

Wolucka, B. A., and Van Montagu, M. (2003). GDP-Mannose 3′,5′-epimerase forms GDP-l-gulose, a putative intermediate for the de Novo biosynthesis of vitamin c in plants*. *J. Biol. Chem.* 278, 47483–47490. doi: 10.1074/jbc.M309135200

Wu, H. C., Bulgakov, V. P., and Jinn, T. L. (2018). Pectin methylesterases: Cell wall remodeling proteins are required for plant response to heat stress. *Front. Plant Sci.* 871, 1612 doi: 10.3389/FPLS.2018.01612/BIBTEX

Xie, Y., Lai, D., Mao, Y., Zhang, W., Shen, W., and Guan, R. (2013). Molecular cloning, characterization, and expression analysis of a novel gene encoding l-cysteine desulfhydrase from brassica napus. *Mol. Biotechnol.* 54, 737–746. doi: 10.1007/S12033-012-9621-9/FIGURES/6

Xu, F., Bao, J., Kim, T.-S., and Park, Y.-J. (2016). Genome-wide association mapping of polyphenol contents and antioxidant capacity in whole-grain rice. *J. Agric. Food Chem.* 64, 4695–4703. doi: 10.1021/acs.jafc.6b01289

Xu, X., Pan, S., Cheng, S., Zhang, B., Mu, D., Ni, P., et al. (2011). Genome sequence and analysis of the tuber crop potato. *Nature* 475, 189–195. doi: 10.1038/nature10158

Yang, X., Cheng, J., Yao, B., Lu, H., Zhang, Y., Xu, J., et al. (2021). Polyploidy-promoted phenolic metabolism confers the increased competitive ability of solidago canadensis. *Oikos* 130, 1014–1025. doi: 10.1111/OIK.08280

Yang, X., Luo, Z., Todd, J., Sood, S., and Wang, J. (2020). Genome-wide association study of multiple yield traits in a diversity panel of polyploid sugarcane (Saccharum spp.). *Plant Genome* 13, e20006. doi: 10.1002/TPG2.20006

Ye, J., Li, W., Ai, G., Li, C., Liu, G., Chen, W., et al. (2019). Genome-wide association analysis identifies a natural variation in basic helix-loop-helix transcription factor regulating ascorbate biosynthesis *via* d-mannose/L-galactose pathway in tomato. *PloS Genet.* 15, e1008149. doi: 10.1371/journal.pgen.1008149

Yousaf, M. F., Demirel, U., Naeem, M., and Çalışkan, M. E. (2021). Association mapping reveals novel genomic regions controlling some root and stolon traits in tetraploid potato (Solanum tuberosum l.). *3 Biotech.* 11, 1–16. doi: 10.1007/S13205-021-02727-6/FIGURES/8

Yuan, Y., Song, T., Yu, J., Zhang, W., Hou, X., Kong Ling, Z., et al. (2022). Genome-wide investigation of the cysteine synthase gene family shows that overexpression of CSase confers alkali tolerance to alfalfa (Medicago sativa l.). *Front. Plant Sci.* 12. doi: 10.3389/FPLS.2021.792862/FULL

Yuan, Y., Wang, B., Zhang, S., Tang, J., Tu, C., Hu, S., et al. (2013). Enhanced allelopathy and competitive ability of invasive plant solidago canadensis in its introduced range. *J. Plant Ecol.* 6, 253–263. doi: 10.1093/JPE/RTS033

Yu, C., Yan, M., Dong, H., Luo, J., Ke, Y., Guo, A., et al. (2021). Maize bHLH55 functions positively in salt tolerance through modulation of AsA biosynthesis by directly regulating GDP-mannose pathway genes. *Plant Sci.* 302, 110676. doi: 10.1016/J.PLANTSCI.2020.110676

Zagoskina, N. v., Kazantseva, V. v., Fesenko, A. N., and Shirokova, A. v. (2018). Accumulation of phenolic compounds at the initial steps of ontogenesis of fagopyrum esculentum plants that differ in their ploidy levels. *Biol. Bull.* 45, 171–178. doi: 10.1134/S1062359018020140

Zhang, Y., Jung, C. S., and De Jong, W. S. (2009). Genetic analysis of pigmented tuber flesh in potato. *Theor. Appl. Genet.* 119, 143–150. doi: 10.1007/s00122-009-1024-3

Zhang, Y. W., Tamba, C. L., Wen, Y. J., Li, P., Ren, W. L., Ni, Y. L., et al. (2020). mrMLM v4.0.2: An r platform for multi-locus genome-wide association studies. *Genomics Proteomics Bioinf.* 18, 481–487. doi: 10.1016/J.GPB.2020.06.006

Zhao, Q., Fan, Z., Qiu, L., Che, Q., Wang, T., Li, Y., et al. (2020). MdbHLH130, an apple bHLH transcription factor, confers water stress resistance by regulating stomatal closure and ROS homeostasis in transgenic tobacco. *Front. Plant Sci.* 11. doi: 10.3389/FPLS.2020.543696/BIBTEX

Zheng, X., Gong, M., Zhang, Q., Tan, H., Li, L., Tang, Y., et al. (2022). Metabolism and regulation of ascorbic acid in fruits. *Plants* 11, 1602. doi: 10.3390/PLANTS11121602

Zhong, H., Liu, S., Sun, T., Kong, W., Deng, X., Peng, Z., et al. (2021). Multi-locus genome-wide association studies for five yield-related traits in rice. *BMC Plant Biol.* 21, 1–12. doi: 10.1186/S12870-021-03146-8/FIGURES/4

Zhu, C., Gore, M., Buckler, E. S., and Yu, J. (2008). Status and prospects of association mapping in plants. *Plant Genome* 1 (1), 5–20. doi: 10.3835/PLANTGENOME2008.02.0089

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 30 January 2023
DOI 10.3389/fpls.2023.1044718

Check for updates

OPEN ACCESS

EDITED BY
Shawn Carlisle Kefauver,
University of Barcelona, Spain

REVIEWED BY
Cengiz Toker,
Akdeniz University, Türkiye
Mauricio Parra Quijano,
National University of Colombia, Colombia

*CORRESPONDENCE
Diego A. Sotomayor
dsotomayor@lamolina.edu.pe

SPECIALTY SECTION
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

RECEIVED 14 September 2022
ACCEPTED 12 January 2023
PUBLISHED 30 January 2023

CITATION
Sotomayor DA, Ellis D, Salas A, Gomez R, Sanchez RA, Carrillo F, Giron C, Quispe V, Manrique-Carpintero NC, Anglin NL and Zorrilla C (2023) Collecting wild potato species (*Solanum* sect. Petota) in Peru to enhance genetic representation and fill gaps in *ex situ* collections. *Front. Plant Sci.* 14:1044718. doi: 10.3389/fpls.2023.1044718

COPYRIGHT
© 2023 Sotomayor, Ellis, Salas, Gomez, Sanchez, Carrillo, Giron, Quispe, Manrique-Carpintero, Anglin and Zorrilla. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Collecting wild potato species (*Solanum* sect. Petota) in Peru to enhance genetic representation and fill gaps in *ex situ* collections

Diego A. Sotomayor[1,2]*, David Ellis[3], Alberto Salas[3], Rene Gomez[3], Rosa A. Sanchez[1,2], Fredesvinda Carrillo[1], Carolina Giron[1], Violeta Quispe[3], Norma C. Manrique-Carpintero[3], Noelle L. Anglin[3,4] and Cinthya Zorrilla[1,5]

[1]Direccion de Recursos Geneticos y Biotecnologia, Instituto Nacional de Innovacion Agraria (INIA), Lima, Peru, [2]Facultad de Ciencias, Universidad Nacional Agraria La Molina (UNALM), Lima, Peru, [3]Centro Internacional de la Papa (CIP), Lima, Peru, [4]USDA ARS Small Grains and Potato Germplasm Unit, Aberdeen, ID, United States, [5]International Atomic Energy Agency, Plant Breeding and Genetics Section, Joint FAO/IAEA Center of Nuclear Techniques in Food and Agriculture, Vienna, Austria

Crop wild relatives (CWRs) are important sources of novel genes, due to their high variability of response to biotic and abiotic stresses, which can be invaluable for crop genetic improvement programs. Recent studies have shown that CWRs are threatened by several factors, including changes in land-use and climate change. A large proportion of CWRs are underrepresented in genebanks, making it necessary to take action to ensure their long-term *ex situ* conservation. With this aim, 18 targeted collecting trips were conducted during 2017/2018 in the center of origin of potato (*Solanum tuberosum* L.), targeting 17 diverse ecological regions of Peru. This was the first comprehensive wild potato collection in Peru in at least 20 years and encompassed most of the unique habitats of potato CWRs in the country. A total of 322 wild potato accessions were collected as seed, tubers, and whole plants for *ex situ* storage and conservation. They belonged to 36 wild potato species including one accession of *S. ayacuchense* that was not conserved previously in any genebank. Most accessions required regeneration in the greenhouse prior to long-term conservation as seed. The collected accessions help reduce genetic gaps in *ex situ* conserved germplasm and will allow further research questions on potato genetic improvement and conservation strategies to be addressed. These potato CWRs are available by request for research, training, and breeding purposes under the terms of the International Treaty for Plant Genetic Resources for Food and Agriculture (ITPGRFA) from the Instituto Nacional de Innovacion Agraria (INIA) and the International Potato Center (CIP) in Lima-Peru.

KEYWORDS

potato, crop wild relatives (CWRs), *ex situ* conservation, genetic resources, *Solanum* species, germplasm collecting

Introduction

Crop wild relatives (CWRs) are non-domesticated "wild" plant species that share a common ancestry with cultivated plants. They typically possess a wider range of genetic diversity in comparison with their cultivated counterparts due to their continued interaction with the environment and lack of genetic manipulation or selection by humans (Smýkal et al., 2018). CWRs offer a critical, often untapped resource, to address food security needs by providing genetic diversity of important agronomic traits for crop improvement, leading to increased plasticity and productivity of farming systems (Jansky et al., 2013). Genes found in CWRs can be introgressed into the crop by breeding programs (Harlan and de Wet, 1971; Singh, 2001; Castañeda-Álvarez et al., 2016) to provide traits such as pest and disease resistance, tolerance to abiotic stresses, increased yield, male fertility and quality, increasing the value and sustainability in crops (Dempewolf et al., 2017).

In the past 20 years, there has been a steady increase in the rate of cultivar releases containing genes from CWRs, and their contribution should only increase as the development of molecular technologies makes identification and utilization of diverse germplasm more efficient (Prescott-Allen and Prescott-Allen, 1986; Tanksley and McCouch, 1997; Singh, 2001; Hajjar & Hodgkin, 2007; Dempewolf et al., 2017). Plant breeders frequently obtain CWRs from genebanks. Having a representative diverse sample is of critical importance, especially at centers of origin for the crop, where diversity is the highest. However, major gaps in the genetic diversity of important crop gene pools such as potato (*Solanum tuberosum* L.), brinjal eggplant (*S. melongena* L.), and tomato (*S. lycopersicum* L.) remain to be filled in *ex situ* germplasm collections (Castañeda-Álvarez et al., 2015; Syfert et al., 2016; Vilchez et al., 2019). Moreover, the survival of some of these wild plant species are threatened by conversion of their natural habitats to agriculture, urbanization, invasive species, mining, climate change and/or pollution, and land-use change (Wilkes, 2007; Jarvis et al., 2008; Ford-Lloyd et al., 2011; Brummitt et al., 2015; Wettberg et al., 2022).

Potato (*S. tuberosum*) is the most important tuber crop worldwide, and one of the top staple foods in the world. According to FAOSTAT (2020) more than 20 million hectares globally are cultivated with potatoes, which produce more than 400 million tons. It continues to gain significance in temperate and tropical regions as a source of carbohydrates, vitamins, and minerals, as well as for industrial purposes (Navarre et al., 2009; Aversano et al., 2017). Potato wild relatives constitute a morphologically and genetically diverse group of plants and are geographically distributed from central Chile and Argentina to the southwestern United States covering 16 countries with high levels of endemism. Mexico, Bolivia, Argentina, and especially Peru, are considered to possess the greatest total diversity of potato wild relatives (Spooner and Salas, 2006) where they occupy a variety of habitats including deserts, forests, and mountainous regions (Hijmans et al., 2002). Potato is also one of the ten crops with the most breeding uses of CWRs documented (Dempewolf et al., 2017).

Currently, the taxonomic classification of tuber-bearing *Solanum* species (section Petota) recognizes 107 species of wild tuber-bearing potato species and four cultivated species (Spooner et al., 2014). From them, 53 species are distributed in Peru including 39 endemic species (Sarkinen et al., 2015). The taxonomy of potato is contentious as some conservation institutions still use older classification systems. For instance, the International Potato Center (CIP from its Spanish acronym) Genebank uses a classification based on the descriptions of Hawkes (1990) and Ochoa (1999) and has not yet adopted Spooner taxonomy (Spooner et al., 2014). Harmonization of the taxonomy used in global potato genebanks is needed for the extensive genetic comparison of potato collections to identify unique and redundant material between genebanks (Ellis et al., 2020).

The variety of ecosystems that are present in Peru, from north to south and from sea level to the rainforest, including seasonally humid habitats in the arid coast, locally known as "lomas", have contributed to the evolution and survival of a vast variety of potato species. Wild potato species in Peru are diverse, not only in their geographical distribution, but also in their relative regional abundance and in the extensive gene flow that occurs between populations. Potato CWRs also vary widely in their life cycle duration, vegetative, and sexual reproduction (allogamous and autogamous species), time of flowering, pollination systems, as well as fruit and seed ripening (Salas et al., 2008; Ellis et al., 2020). Simulating these highly varied environmental conditions in *ex situ* conservation systems can be extremely challenging as many species may need specialized conditions (often unknown) to produce ample flowers, fruits, and seeds for genebank regeneration, conservation, and subsequent distributions of seed samples. Further, many species may not produce many flowers, fruits, and seeds even in their natural environments. Jansky et al. (2013) concluded that *ex situ* and *in situ* preservation are essential for a comprehensive conservation plan for potato CWRs and that collecting gaps detected in genebanks was a top priority. Castañeda-Álvarez et al. (2015), in an analysis of the state of *ex situ* conservation of potato CWRs, highlighted a high-priority need for the collection (*ex situ* gaps) of over 40% of the species of wild potatoes. This same study classified 26 species of wild potatoes native to Peru as high to medium priority for collecting. Thus, it is imperative that collection of these species is made before more habitats, populations, or species are lost.

To our knowledge there has not been any collections of wild potatoes (*Solanum* section Petota) for the past two decades or more in Peru, certainly not one that has placed the material into the Multilateral System (MLS) of the International Treaty for Plant Genetic Resources for Food and Agriculture (ITPGRFA) with materials readily available for the research community. Hence, there was and still is a dire need for plant collecting, as habitats are changing drastically throughout Peru due to urbanization, mining, and extremely rapid changes in the environment due to climate change, particularly affecting the Andes. This article presents the results of a series of wild potato collecting expeditions in Peru conducted under a global program led by the Global Crop Diversity Trust (GCDT) [for more specific details on the collecting in multiple crops and countries see (Eastwood et al., 2022)]. Here, we detail only the collection trips for potato in Peru, the biological material collected, the localities visited, the filling of *ex situ* conservation gaps, and discuss other specifics related to the collection and multiplication of the potato CWRs for long-term conservation in *ex situ* and routine distribution for research, training, and breeding.

Materials and methods

Study sites

Potato CWRs are widely distributed in the Andean region of Peru (Figure 1); however, a few species occur in the arid coast, as well as in the Amazon basin. To determine the most appropriate timing and collecting route for each species, we based our decision on expert's opinion and reported observation sites. This was complemented with the available data on the target species distribution as reported in global databases, such as Global Biodiversity Information Facility-GBIF (https://www.gbif.org/), Tropicos (https://www.tropicos.org/) and Genesys (https://www.genesys-pgr.org/), the CWR collecting guide (RBG Kew 2016), and the CIP Genebank database (https://genebank.cipotato.org/gringlobal/search.aspx). The former curator of wild potato species at CIP, Alberto Salas with ca. 50 years of experience collecting potato CWRs, led the planning of the collecting trips and provided detailed expertise on where potato species should be found based on previous collection trips in Peru. One goal was to re-visit sites where the species of interest had been collected decades prior to make new collections of the species, since genetic drift could be reshaping the allelic composition of many CWR species. The collection schedule was coordinated to target the best timing for as many populations as possible for most of the species and/or sites. Some of the collecting trips were performed in parallel with two crews going to separate geographic locations to maximize the potential of having a team be present during the growing season for collection. The crews of each trip were comprised by at least three researchers to help find the potato CWR plants and to record the data associated with the accessions collected.

Germplasm collection

Prioritization of potato species for collecting included those species with low representation (from 0 to 10 accessions) in germplasm banks or herbariums globally. The main efforts for collecting were focused on three species, *S. ayacuchense*, *S. olmosense*, and *S. salasianum*. These species are endemic to Peru and do not exist in any genebank as reported by Castañeda-Álvarez et al. (2015), making them critical for collecting and *ex situ* conservation. In addition, species that were not represented or represented only by a single accession in the CIP Genebank were also priority targets for the collection trips. These include *S. arahuayum* (0 accessions), *S. jaenense* (0 accessions) and *S. ortegae* (1 accession). The taxonomic classifications developed by both Spooner et al. (2014) and Hawkes (1990) were used as an aid for targeting sites to visit. To determine synonymous species, the Solanaceae Source database (http://solanaceaesource.org/) was used in addition to the list of *Solanum* species in Peru (Sarkinen et al., 2015). Collecting for this project was conducted following Peru's legal requirements, and hence, the Servicio Nacional Forestal y de Fauna

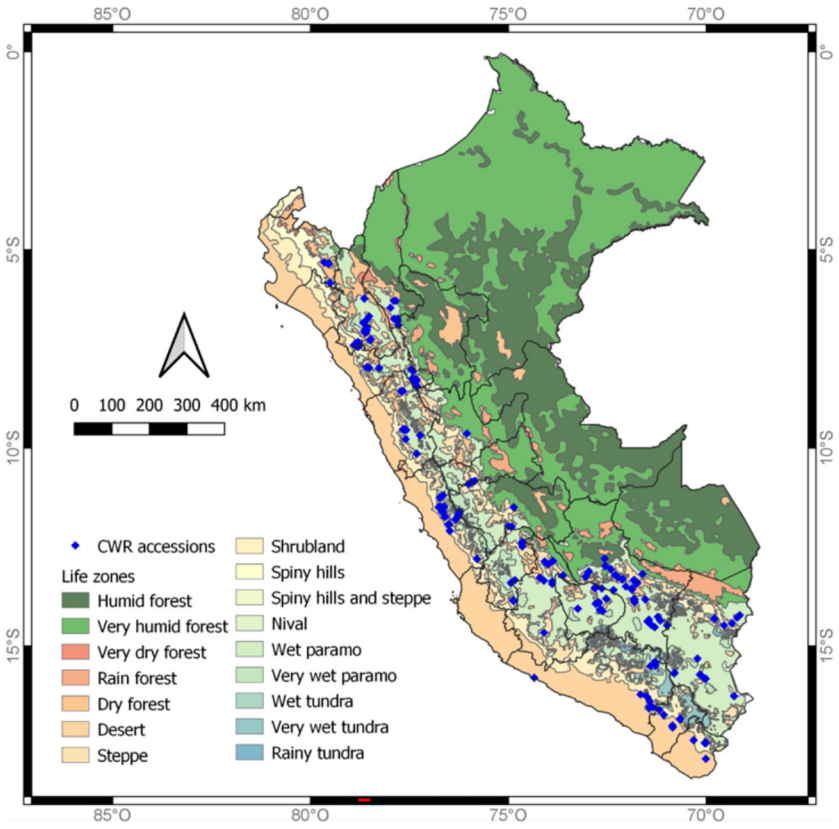


FIGURE 1
Map of potato CWRs accessions collected 2017/2018 in Peru with life zones (or habitats) depicted. The blue diamonds shown on the map are the sites in which the CWRs in this study were collected.

Silvestre (SERFOR) authorized the collecting expeditions under permit number AUT-IFL-2016-038.

The unit of germplasm collected, or the accession, was a population defined as a set of plants growing near each other that belong to the same species and were at the same collecting site. A standard was set to have collection sites at least 2 km in distance from an adjacent one. Collections were made from areas walking distance of roads or paths, due to security and accessibility reasons.

The targeted collecting sample was mature berries containing mature seed obtained from at least five to ten plants. However, when berries were not present, tubers and/or whole plants from at least five to ten plants were collected for later greenhouse regeneration of seed. In addition, herbarium samples were taken to conserve and document a voucher of each population collected using a portable plant press. Following the instructions of the collecting permit granted by Peru´s national authority (SERFOR), "Free and Prior Informed Consent (PIC)" was obtained prior to collecting from any farmer´s community or private properties. Farmers were frequently consulted to help identify areas of interest for collecting.

A minimum set of passport descriptors that included collecting site and date, collectors, GPS coordinates, a brief description of the site (including biophysical characteristics and surrounding flora), a quick visual assessment of potential diseases and pests, and other features of interest, were recorded. Further, photographs of the material being collected, both in their natural habitat, as well as close-up of collected materials were taken and recorded in the database/collecting notes. The collecting notebook included the following data: collecting ID, collector names, genus, species, native name, state of the sample (wild, weedy), type of sample taken (voucher, fruits, tubers, seedlings, etc.), reported uses, pathogens, associated flora, topography, vegetation, and soil type.

Sample processing

Mature berries were collected in paper bags and seed was extracted, washed, dried, and stored at CIP at -20°C according to the standard procedures established at CIP which is aligned with the international standards for seed conservation (FAO, 2014). In cases where the number of seed collected was below the minimum needed for conservation (6000 seeds per accession), seed was regenerated in greenhouses (CWRs are never multiplicated in a field setting due to the concern about invasive escapes). In cases where berries were collected but not fully mature, these were allowed to ripen prior to seed extraction by leaving them in paper bags at greenhouse temperature (20°C – 25°C). After seeds were extracted, they were split between the two collaborating institutions (INIA and CIP) for long-term conservation at -20°C in heat-sealed aluminium pouches.

When berries were not available, either tubers or whole plants were collected. Tubers were collected in paper bags, washed, and disinfected after arrival to the laboratory, and stored at 4°CC until needed for regeneration. Whole plants with roots and soil attached were collected, wrapped in moistened newspaper, and placed in plastic bags during transport back to CIP in Lima, Peru where they were immediately planted. Tubers and plants, were planted in 20 cm diameter pots, using Promix 8 as the substrate, grown in a screenhouse, and fertilized with Jack's Professional fertilizer (3g/l before and 5g/l after flowering) and allowed to grow until mature plants with berries developed.

Herbarium samples consisted of whole plants pressed in a portable plant press on site using newspaper sheets between samples for transport back to the laboratory. Ideally, when possible, three samples were taken, the first sample for INIA, the second for CIP, and the third to be deposited in the Herbarium of the Natural History Museum in Lima, Peru. In the laboratory, the procedure for processing the herbarium samples started by carefully pressing, rapidly drying at 50°C for about 48 hours and then permanently mounting whole plant samples on herbarium sheets along with barcoded labels to identify each specimen and collecting information in the database. The herbarium sheets were then stored in polyethylene bags in controlled temperature (19°C-21°C) herbarium cabinets with controlled relative humidity (45-50%), following CIP standards for herbarium specimen conservation (https://cipotato.org/genebankcip/process/herbarium/, Vargas et al, 2016) prior to placement in the permanent herbarium collection.

Seed regeneration

Seed regeneration was carried out for all accessions collected as tubers and/or whole plants, as well as for accessions with a low number of seed (less than 6000 seeds). Six thousand seeds were selected to ensure both *ex situ* institutions (INIA and CIP) maintaining this material, would have ample material for distribution requests and conserving the material into perpetuity. These accessions were grown under greenhouse conditions in Lima (coastal site), Huancayo, and Cusco (higher elevation sites) depending on the origin of the accessions and understanding of the best environmental conditions for regeneration.

In the case of seed collections needing a seed increase, a minimum of 100 seeds were germinated. A set of 25-30 seedlings were transplanted in Jiffys-7 pots for 30 days, and then transferred to 20 cm pots using a Promix 8 substrate (Salas et al., 2008). Due to the ISO 17025:2017 accreditation at CIP which regulates workflows to ensure that only virus free germplasm is moved around the globe, each seedling was tested for the following viruses using standard laboratory testing (ELISA, PCR, etc.) and complemented with biological indexing on indicator plants: Arracacha Virus B Oca strain (AVB-O), Alfalfa Mosaic Virus (AMV), Andean Potato Latent Virus (APLV), Potato Yellowing Virus (PYV), Tobacco Mosaic Virus (TMV), Potato Virus T (PVT), and the quarantine viroid Potato Spindle Tuber Viroid (PSTVd), after which only virus-free plants were used for seed regeneration. Any plants that were positive for these viruses were destroyed to prevent dissemination of viruses infecting potato.

When plants were flowering, depending on their mode of reproduction, autogamous or allogamous, open pollination or a combination of sib-crosses and bulk crosses were made, respectively, to obtain seed for conservation. The regeneration of seed was performed between 2017 and 2021. The resulting seed was divided and shared between CIP and INIA for long term conservation.

Results and discussion

Collecting expeditions

Prioritization of species and collecting sites was done *via* expert judgment and through a review of past collection sites, herbarium specimens, genebank passport data, and gap analysis results (Royal Botanical Garden Kew, 2016). An important aspect for collections such as these, is the identification of priority sites for *in situ* conservation of CWRs. It was observed that the habitat of potato CWRs are vanishing at an alarming rate, primarily due to changes in land use in Peru. Indeed, we found populations that were lost due to urban/agricultural expansion including *S. limbaniense* Ochoa in Cullucachi, Sandia-Puno, *S. saxatilis* Ochoa in Kkakkapata, Sandia-Puno, *S. medians* Bitter in San Juan de Lurigancho-Lima, *S. wittmackii* Bitter in Chorrillos-Lima, *S. mochiquense* Ochoa in Cerro Chipuitur and Cerro Cabras-Trujillo, and *S. olmosense* Ochoa in Porcuya, Huancabamba-Piura. In addition, populations are also lost due to intensified food production such as the development of chicken farms which has specifically impacted *S. chancayense* Ochoa and *S. inmite* Dunal in Chancay, Huaral-Lima. The loss of these populations emphasizes the need for continued collections and *ex situ* conservation of these valuable genetic resources.

In this sense, systematic conservation planning (Margules and Pressey, 2000) could be used to identify sites at a local scale and implement specific conservation actions to complement efforts at the global scale (Vincent et al., 2019). Comprehensive efforts, combining *ex situ* and *in situ* strategies employing multiple institutions working together are needed to conserve the genetic diversity of potato CWRs, especially in Peru, where large species diversity and ecogeographical conditions exist, especially because there are significantly increasing rates of land use change occurring in Peru.

A total of 18 collecting trips were conducted across a large portion of the varying habitats where potato CWRs have been documented or thought to be present in Peru. Where possible, trips included local INIA staff or farmers that guided and actively participated in the collection. The collecting trips are presented chronologically (Table 1) and locations are shown on a map (Figure 1). The collecting trips were performed at different times of the year between May-October 2017 and January-October 2018 to sample different life cycle stages of potato CWRs. The optimum time to collect a population is when they are in the later stages of flowering (with flowers desired for species identification and the collection of proper herbarium specimens) and ripe fruit present and still attached to the plant. Unfortunately, increased climatic uncertainty and consequent changes in patterns of development of plants made this condition harder to find than expected and, in most cases, either tubers or whole plants were collected, or a second visit was needed to collect the samples in the right stage of development. Effects of climate change disturbing crops in the Andes has been reviewed by Perez et al. (2010) - these factors also influence the development of their crop wild relatives.

Unforeseen events limited the capacity to collect accessions over the course of 2017/2018. The main challenge was limited access to the north and central parts of the country in 2017 due to the occurrence of a coastal El Niño event bringing strong rains, landslides, and extensive road closures which were often unpaved and not well maintained (Rodríguez-Morata et al., 2019). As well, a severe drought after the coastal El Niño event in 2017, reduced the amount of rain in the coastal regions (also known as "lomas") limiting plant development. To prevent situations that might affect the safety of the collectors, most collecting trips were rescheduled to 2018. The 2018 trips accounted for 105 collecting days, which in addition to the 36 days of collections in 2017, gave a total of 141 collecting days for the entire study. Other challenges that occurred included a landslide occurring in front of collectors on a road near the buffer zone of the Manu National Park in Cusco and an armed robbery of a collecting crew on the first day of the 2018 collecting campaign in Apurimac (central Peru), which resulted in some re-scheduling and delays of collecting activities along with loss of supplies and equipment needed for collecting. These unforeseen events, although not unique to the collection of potato CWRs or other crops for that matter, demonstrate the multitude of challenges that typically occur when conducting field research and the need for great flexibility in planning and travel. Atypical climatic events highlight the need for flexibility and future collecting trips to be planned to allow visits to sites in multiple years to ensure success and the preservation of target species before they vanish.

TABLE 1 Collecting trips of potato CWRs performed in Peru during 2017/2018.

Trips	Regions	Collecting days	Accessions collected
1	Tacna, Moquegua, Arequipa, Apurimac, Ayacucho, Cusco	12	33
2	Puno, Tacna, Arequipa, Moquegua	12	4
3	Apurimac, Cusco	12	34
4	Trips suspended due to robbery, targeted Apurimac	0	0
5	Puno	10	19
6	Ayacucho, Huancavelica, Junín, Pasco	10	24
7	Pasco, Lima	11	64
8	Piura, Amazonas, Cajamarca	11	13
9	La Libertad, Ancash	11	38
10	Cajamarca, Amazonas, Piura	11	36
11	Huánuco, Ancash	7	2
12	Lima	3	1
13	Huancavelica	2	2
14	San Martín	5	0
15	Arequipa, Tacna, Moquegua	6	1
16	La Libertad, Lima	6	0
17	Lima	1	0
18	Cusco, Apurimac	11	51
Total	**17**	**141**	**322**

The collecting permit granted by SERFOR and the funding available for the project only allowed potato CWRs to be sampled over a two-year period so extension to another year was not possible. This was an additional challenge in collecting potatoes (time limitations). The collecting trips did not always unearth the prioritized species, but often other species were found in their place. This could be because the target species were either relatively rare or that the weather conditions were just not suitable for their emergence in the landscape during these years which is a typical challenge faced by plant collectors. For example, after the El Niño event of 1998, the population of *S. augustii* Ochoa in Ancash (central Peru) greatly increased making it possible to collect (Alberto Salas, pers. Comm.). Another challenge when visiting sites explored 20-40 years ago, was the effect of new human settlements and expansion of economic activities such as mining which have drastically changed the natural landscape in Peru and beyond. This resulted in some species becoming locally extinct. Thus, expanding the geographical area and time allotted for collecting particular potato species was greatly needed instead of making collections only near roads for efficiency and safety reasons. It is important to note that for this study, collecting sites could only be visited once or at maximum twice, and therefore, it was impossible to confirm if species still existed at a site or not, due to the time limitations imposed at each stop. Clearly, future collecting trips and their associated permits need to be carefully planned, optimally for more than a two-year time span to ensure the collection of species that may only germinate under certain climatic conditions. This flexibility should be recognized and supported by funding agencies.

Copies of the successfully collected material are deposited at both, INIA and CIP genebanks and are available for breeding, research, and training under the terms of the Standard Material Transfer Agreement (SMTA) of the ITPGRFA. In this context, the material collected is now available for use under the MLS and is safely conserved long-term in two genebanks in Peru. The collected material can be requested by users from CIP at (https://genebank.cipotato.org/gringlobal/search.aspx) or from the Agricultural Innovation Division of INIA (https://www.inia.gob.pe/requisitos-acceso-rrgg/), establishing a new chapter in the study of potato CWR.

Seed regeneration and conservation

The amount of material collected that needed regeneration was greater than anticipated at the beginning of the study, in terms of (1) the number of accessions that required extensive seed regeneration for long-term storage; (2) the geographic locations where it had to be regenerated based on its origin, (3) the necessity to identify or validate the taxonomy of the accessions, and (4) the highly regimented virus testing procedures for accessions in order to only store, maintain, and distribute disease-free material. A considerable amount of care is required to meet these objectives in the regeneration effort, to maximize conservation of the representative genetic variation in the collected populations especially when small numbers of seed/genetic material were available. Another important point of consideration for planning future collection trips is building in sufficient time and labor for the effort required for extensive regenerations.

A total of 245 accessions, representing 76% of the 322 accessions originally collected, required regeneration. Regeneration of the material was initiated in September 2017 and finished in June 2021 at INIA and CIP greenhouses in Lima (coastal-central Peru), Huancayo (Andean-central Peru) and Cusco (Andean-southern Peru), with most of the regenerations performed in Huancayo. The decision to regenerate at a certain location was based on the daylength and elevation of the species natural habitat.

Geographical representation of the collection

Our collecting trips acquired accessions from 17 out of the 25 political regions of Peru, regions not included were those located in the rainforest and the coastal desert, which are not known habitats of potato CWRs. The Peruvian Meteorological Service following Holdridge life zones (Holdridge, 1987) has divided Peru into 16 life zones: humid forest, very humid forest, very dry forest, rain forest, dry forest, desert, steppe, shrubland, spiny hills, spiny hills and steppe, nival, wet paramo, very wet paramo, wet tundra, very wet tundra, and rainy tundra. Collections were made in 12 (humid and very humid forests, dry forests, desert or arid environments, steppe, shrublands, spiny hills and steppe, very wet and paramo, and very wet and rainy tundra) of these specified zones (Figures 1, 2). Wet paramo and steppe were the life zones from which more accessions were collected.

Castañeda-Álvarez et al. (2015) presented the regions and localities where high priority species were distributed, and all were included in the collecting campaign. However, as previously suggested, a more precise method for priorization of collecting sites and the identification of candidates for *in situ* conservation areas should be developed in the future to enhance the conservation efforts currently made by genebanks. Gap analysis (Ramirez-Villegas et al., 2010; Castañeda-Álvarez et al., 2015 at the global scale; Vilchez et al., 2019 at the national scale) and systematic conservation planning (Margules and Pressey, 2000; Vincent et al., 2019) that combine both herbarium/germplasm bank distribution data with species distribution modeling could be used for the priorization of sites for collecting and for *in situ* conservation purposes.

Species representation in the collection

Taxonomic classification is a very important step for adequate conservation and utilization by researchers and breeders. A total of 322 accessions (populations) were collected between 2017 and 2018 of which 284 accessions were taxonomically classified belonging to 36 species according to Spooner et al. (2014), the respective species equivalency in Hawkes (1990) is also presented (Table 2). Additionally, 38 accessions have no species designation to date as certainty of the taxonomic identity of the species at the collection site or the greenhouse has not been possible and were thus labeled as *Solanum* spp. In this case, we propose that the complication with species identity could be due to the presence of natural hybrids. This highlights the need to further delineate species boundaries.

Collected species with reported traits of interest for breeding in cultivated potatoes include *S. acaule* Bitter, *S. boliviense* Dunal, *S.*

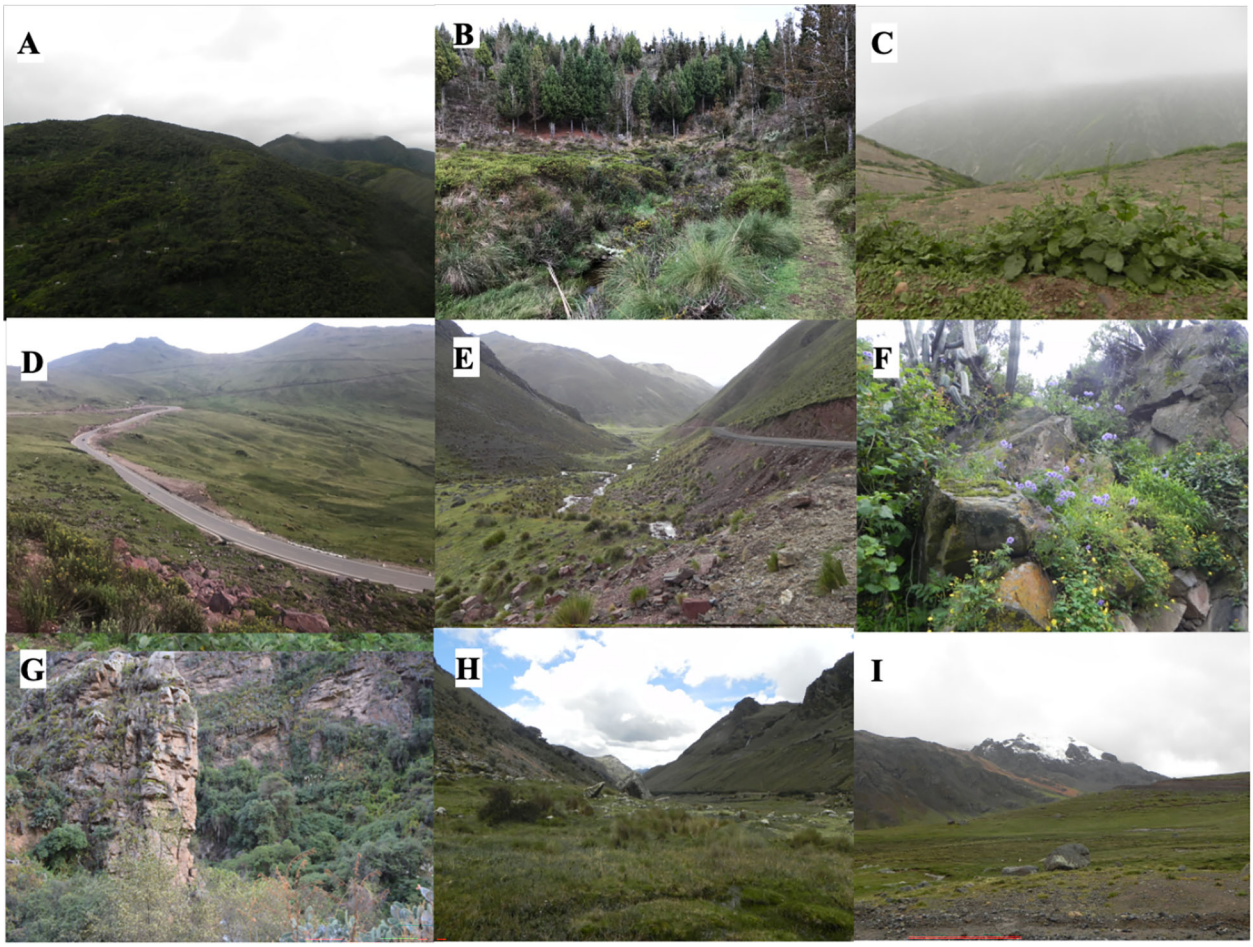


FIGURE 2
Habitats in Peru where potato CWRs were collected. **(A)** Humid and very humid forest habitats in Puno; **(B)** Dry forest habitat in Piura; **(C)** Desert habitats, lomas vegetation; Steppe habitats in **(D)** Ayacucho, and **(E)** Lima; **(F)** Shrubland habitats in Lima; **(G)** Spiny hills and steppe habitats in Apurimac; **(H)** Wet and very wet paramo habitats in Ancash; and **(I)** Tundra habitat in Pasco.

brevicaule Bitter, *S. candolleanum* Berthault, *S. chacoense* Bitter, *S. chomatophilum* Bitter, *S. hypacrarthrum* Bitter, *S. medians*, *S. piurae* Bitter, *S. raphanifolium* Cárdenas & Hawkes, *S. sogarandinum* Ochoa and *S. violaceimarmoratum* Bitter. Traits include tolerance to below freezing temperatures, tolerance to semi-desert environments; resistance to viruses, pests and diseases such as late blight and others, and/or traits amenable to processing (Machida-Hirano, 2015; Kumari et al., 2018; Karki et al., 2021). The availability of a larger number of accessions for these species is an advantage for screening of the reported traits, as well as for the discovery of new traits of interest that may be mined for breeding purposes.

Future collecting efforts are still needed to fill genetic gaps and have a well-represented *ex situ* collection of potato CWRs conserved in genebanks. Although very successful, with this initial two-year collection effort, it was not possible to collect all high priority potato CWRs species. Examples of targeted species which were not found include *S olmosense* and *S. salasianum* Ochoa which are not represented in any genebank (Castañeda-Álvarez et al., 2015) and could be at risk due to their growth in specific habitats with a narrow geographical distribution. *S. olmosense* is distributed in the north of Peru (Lambayeque, Piura) and Ecuador (Loja) in the rain forests and *S. salasianum* is endemic to central Peru (Huanuco) in moist and humid habitats (Solanaceae Source, http://solanaceaesource.org/). Similarly, *S. arahuayum* Ochoa, *S. jaenense* Ochoa, and *S. ortegae* Ochoa (respectively considered synonyms of *S. medians*, *S. colombianum* Dunal and *S. candolleanum* by Spooner et al. (2014) were not found and are either not represented, or represented by a single accession, in the CIP Genebank.

It is important to note that the inability to find a species in this limited expedition cannot be inferred as evidence of species extinction as far more geographical surveillance is needed for such an assessment. Humphreys et al. (2019) reported two potato species as extinct, *S. cajamarquense* Ochoa and *S. hygrothermicum* Ochoa, both of which are native to Peru and this assessment had to be later corrected after evidence of their existence was demonstrated. *S. cajamarquense* is not extinct in the wild as it was collected during this collecting campaign, and in fact seed is available from multiple genebanks around the world (USDA, IPK, and CIP; see https://www.grin-global.org/, Genesys). In the case of *S. hygrothermicum*, Spooner's revision of potato taxonomy did not recognize this

10.3389/fpls.2023.1044718

TABLE 2 The number of accessions collected and conserved, on a species basis, during eighteen collecting trips in Peru (2017/2018).

Spooner et al. (2014)	Hawkes (1990)	priority	collected	conserved
S. acaule Bitter	*S. acaule* Bitter	NFCR	35	33
S. acroscopicum Ochoa	*S. acroscopicum* Ochoa	HPS	10	9
	S. lopez-camarenae Ochoa	HPS	2	0
S. albicans Ochoa	*S. albicans* Ochoa	LPS	1	1
S. amayanum Ochoa	*S. amayanum* Ochoa	HPS	1	1
S. ayacuchense Ochoa	*S. ayacuchense* Ochoa	HPS	1	1
S. boliviense Ochoa	*S. megistacrolobum* Bitter	MPS	7	7
S. brevicaule Bitter	*S. sparsipilum* (Bitter) Juz. and Bukasov	MPS	8	7
S. burkartii Ochoa	*S. burkartii* Ochoa	HPS	5	2
S. cajamarquense Ochoa	*S. cajamarquense* Ochoa	HPS	7	7
S. candolleanum Berthault	*S. aymaraesense* Ochoa	HPS	2	2
	S. bill-hookeri Ochoa	HPS	1	1
	S. bukasovii	LPS	45	43
	S. coelestispetalum Vargas	MPS	3	3
	S. longiusculus Ochoa	HPS	2	1
	S. marinasense Vargas	MPS	7	4
	S. orophilum Correll	LPS	1	1
	S. ortegae Ochoa	HPS	1	1
	S. pampasense Hawkes	HPS	1	1
	S. tarapatanum Ochoa	HPS	1	1
	S. velardei Ochoa	HPS	2	0
S. cantense Ochoa	*S. cantense* Ochoa	HPS	2	2
S. chacoense Bitter	*S. yungasense* Hawkes	HPS	3	3
S. chiquidenum Ochoa	*S. chiquidenum* Ochoa	MPS	9	6
S. chomatophilum Bitter	*S. chomatophilum* Bitter	MPS	13	8
	S. jalcae Ochoa	HPS	4	2
S. contumazaense Ochoa	*S. contumazaense* Ochoa	HPS	1	1
S. dolichocremastrum Bitter	*S. dolichocremastrum* Bitter	MPS	3	3
S. gracilifrons Bitter	*S. gracilifrons* Bitter	HPS	1	1
S. hastiforme Correll	*S. hastiforme* Correll	HPS	1	0
S. huancabambense Ochoa	*S. huancabambense* Ochoa	MPS	1	1
S. hypacrarthrum Bitter	*S. guzmanguense* Whalen & Sagást.	HPS	1	1
S. immite Dunal	*S. immite* Dunal	HPS	3	1
S. laxissimum Bitter	*S. laxissimum* Bitter	HPS	5	3
	S. santolallae Vargas	HPS	1	0
S. lignicaule Vargas	*S. lignicaule* Vargas	HPS	3	3
S. limbaniense Ochoa	*S. limbaniense* Ochoa	HPS	2	1
S. medians Bitter	*S. medians* Bitter	MPS	17	15
	S. sandemanii Hawkes	HPS	7	6
	S. tacnaense Ochoa	HPS	12	8

(Continued)

TABLE 2 Continued

Spooner et al. (2014)	Hawkes (1990)	priority	collected	conserved
S. mochiquense Ochoa	*S. mochiquense* Ochoa	HPS	2	2
S. multiinterruptum Bitter	*S. multiinterruptum* Bitter	LPS	14	14
S. nubicola Ochoa	*S. nubicola* Ochoa	HPS	2	2
S. paucissectum Ochoa	*S. paucissectum* Ochoa	LPS	4	0
S. piurae Bitter	*S. piurae* Bitter	HPS	3	3
S. raphanifolium Cárdenas & Hawkes	*S. raphanifolium* Cárdenas & Hawkes	LPS	12	12
S. rhomboideilanceolatum Ochoa	*S. rhomboideilanceolatum* Ochoa	HPS	2	2
S. simplicissimum Ochoa	*S. simplicissimum* Ochoa	HPS	1	1
S. sogarandinum Ochoa	*S. sogarandinum* Ochoa	MPS	1	1
S. violaceimarmoratum Bitter	*S. urubambae* Juz.	HPS	2	1
S. wittmackii Bitter	*S. wittmackii* Bitter	HPS	8	8
Solanum spp.	*Solanum* spp.	NA	39	34
36 species	**51 species**		**322**	**238**

Taxonomic identification as of August 2019. Collecting priorities (priority) according to Castañeda-Álvarez et al, (2015): high priority species (HPS), medium priority species (MPS), low priority species (LPS), no further collecting required (NFCR), not available (NA).

species as valid, lumping it into *Solanum tuberosum* Andigena group; however, a specimen of this taxa was collected in 2006 and is available at the CIP genebank as an herbarium specimen. We agree, however with Humphreys et al. (2019), that extinction can be under or overestimated and that these caveats can be alleviated through increased study of poorly known biodiverse areas such as the Peruvian highlands.

Few exploratory plant collection trips have taken place in the relevant potato diversity areas since the entering into force of the Convention of Biological Diversity (CBD) and the bottleneck created by navigating the complex legal requirements to obtain the permission from a country through the required permits for collecting. Targeted searches of prioritized species need to be made in the future for multi-year collecting efforts to determine current distribution and conservation status, population dynamics and phenology, as well as for conservation of genetic resources in long-term *ex situ* storage.

Filling the *ex situ* conservation gaps for potato CWR

Accessions collected in this study will improve the representation of potato CWR conserved *ex situ* in Peru and globally, along with ensuring that they are publicly available under the terms of the SMTA from two genebanks, INIA and CIP; in addition, these accessions will be deposited at the Svalbard Global Seed Vault ensuring secure long-term conservation. However, there are several species endemic to other countries from the United States of America to Chile and Argentina that still need to be collected and conserved. Among the 36 species collected in this campaign for *ex situ* conservation, 24 species had high priority, seven had moderately high priority, four species with low priority, and one species (*S. acaule*) was not a priority for further collecting (Table 2). The *S. acaule* collections described here were mainly opportunistic (populations found while looking for another species). Clearly in this campaign, most accessions collected were high priority species indicating that the collecting efforts were successful in filling gaps of *ex situ* conservation (Figure 3). Of particular note is the gap filled by the collection of one *S. ayacuchense* Ochoa accession, a species with high conservation priority that had not been previously conserved in any genebank (Figure 4). Another important accession collected is a population of *S. candolleanum*, synonym of *S. ortegae* under Hawkes (1990) and Ochoa (1999) taxonomical treatment, also with high conservation priority and not available in any genebank until now (Figure 4).

Recommendations for plant collection trips

Four major aspects that should be considered in future collecting campaigns include:

(1) Strong phenologic changes exist from year to year affecting the duration of growth cycles, the number of flowers formed, pollination efficiency and the number of berries formed. These changes would have a greater impact on CWRs than on their cultivated counterparts which could be further impacted by climate change and consequently will increase their unpredictability.

(2) Regeneration in greenhouses or field (if applicable) is necessary to conserve sufficient quantities of virus-free seed. Approximately 76% of the collected samples required further multiplication to ensure sufficient seed for conservation and distribution. Future collecting campaigns need to include

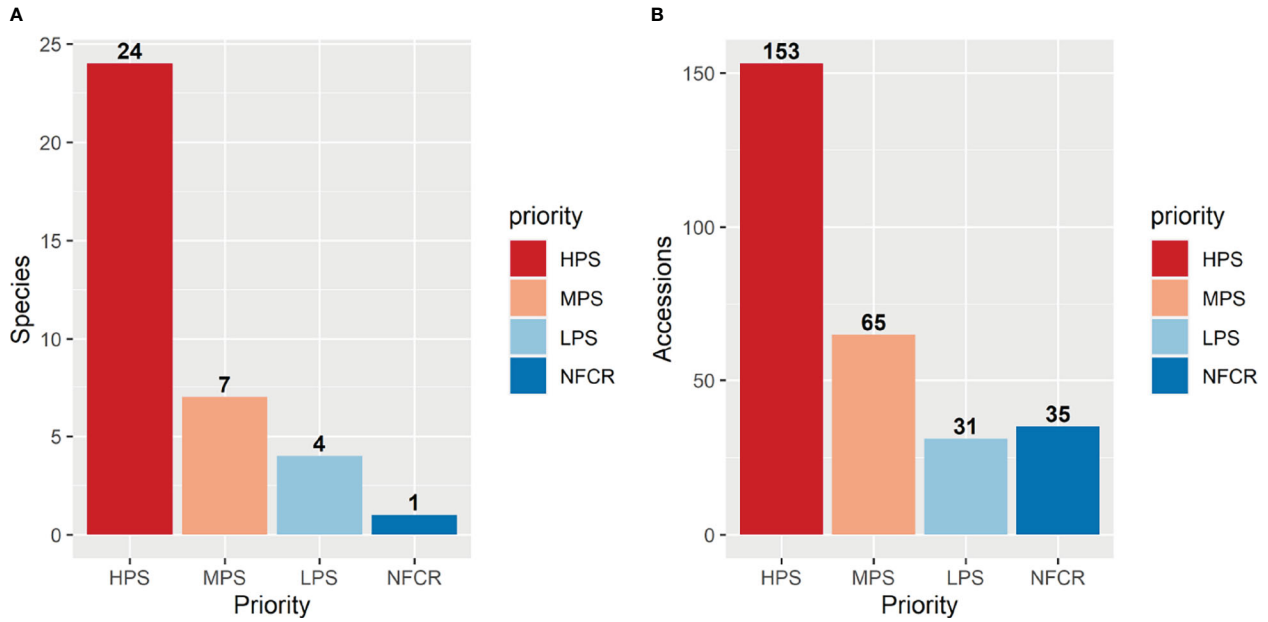


FIGURE 3
Number of species **(A)** and accessions **(B)** collected and conserved. Identification and prioritization of material for *ex situ* conservation was done using the Spooner et al. (2014) taxonomic classification. Collecting priorities according to Castañeda-Álvarez et al. (2015): high priority species (HPS), medium priority species (MPS), low priority species (LPS), and no-further collecting required (NFCR). Accessions with undetermined taxonomic classification (38 accessions) were not included in this figure.

several extra years after collection for multiplication and virus testing before material can be made available to researchers. Proper timing of collections to ensure the ability to collect ample samples at the time of collection and/or adequate regeneration success in the first regeneration cycle can avoid subsequent regenerations of small populations that could cause variation in the population structure and loss of key alleles.

(3) Local collaborators (farmers) are essential for successful collecting trips. Local inhabitants have unique knowledge of the geography and species habitats. Collaborations with local communities and/or farmers during collection improves the depth of searches, strengthens local capacities, and can significantly decrease the time required to organize collecting trips. This could also lead to an increased sense of the need of conservation of habitats locally. Further, local farmers tend to be invested in safeguarding plant material and understand the importance of conservation efforts for food security.

(4) Allowing sufficient time to navigate and obtain host country collecting permits and ensuring previous and informed consent (PIC) are obtained during collection process is of paramount importance to comply with obligations set forth from the Convention in Biological Diversity. The process of obtaining the permits can take quite a lot of time to navigate and may require building relationships and trust with key partners to ensure success of getting the necessary permits. The system for obtaining collecting permits can change country by country as it is not standardized, hence having host country partners can help point research to the necessary institutions and requirements for obtaining a permit. Funding agencies should also consider this aspect by giving flexibility to researchers.

Conclusions

Between 2018 and 2019 we conducted 18 collecting trips accounting for 141 collecting days across the main distribution of potato's CWR in Peru. This is the first comprehensive expedition in the last 20 years in a country that is center of origin of one of the most important crops in the world. We collected 322 CWR accessions belonging to 36 species of potato CWR, filling the gap for ex situ conservation (Castañeda-Álvarez et al., 2015) with 24 species designated as a high priority for collection, including *S. ayacuchense* that had not been previously conserved in any genebank until now. During the collection we found that the habitats of potato's CWR are quickly disappearing, which underlines the importance of strengthening both *in situ* and ex situ conservation efforts.

The potato CWRs collected in this study have contributed to the reduction of genetic gaps in *ex situ* collections worldwide. Nevertheless, future collecting efforts are still needed to have a well-represented potato CWR *ex situ* collection conserved in global genebanks, preventing loss of valuable genetic materials due to changes in natural habitats. Further, more intensive collecting efforts could help document habitat loss and the possibility of associated extinctions of key species, highlighting the need to continually document changes and collect species before they are lost in their natural environments. Strategic conservation initiatives that combine *in situ* and *ex situ* conservation are necessary to ensure the

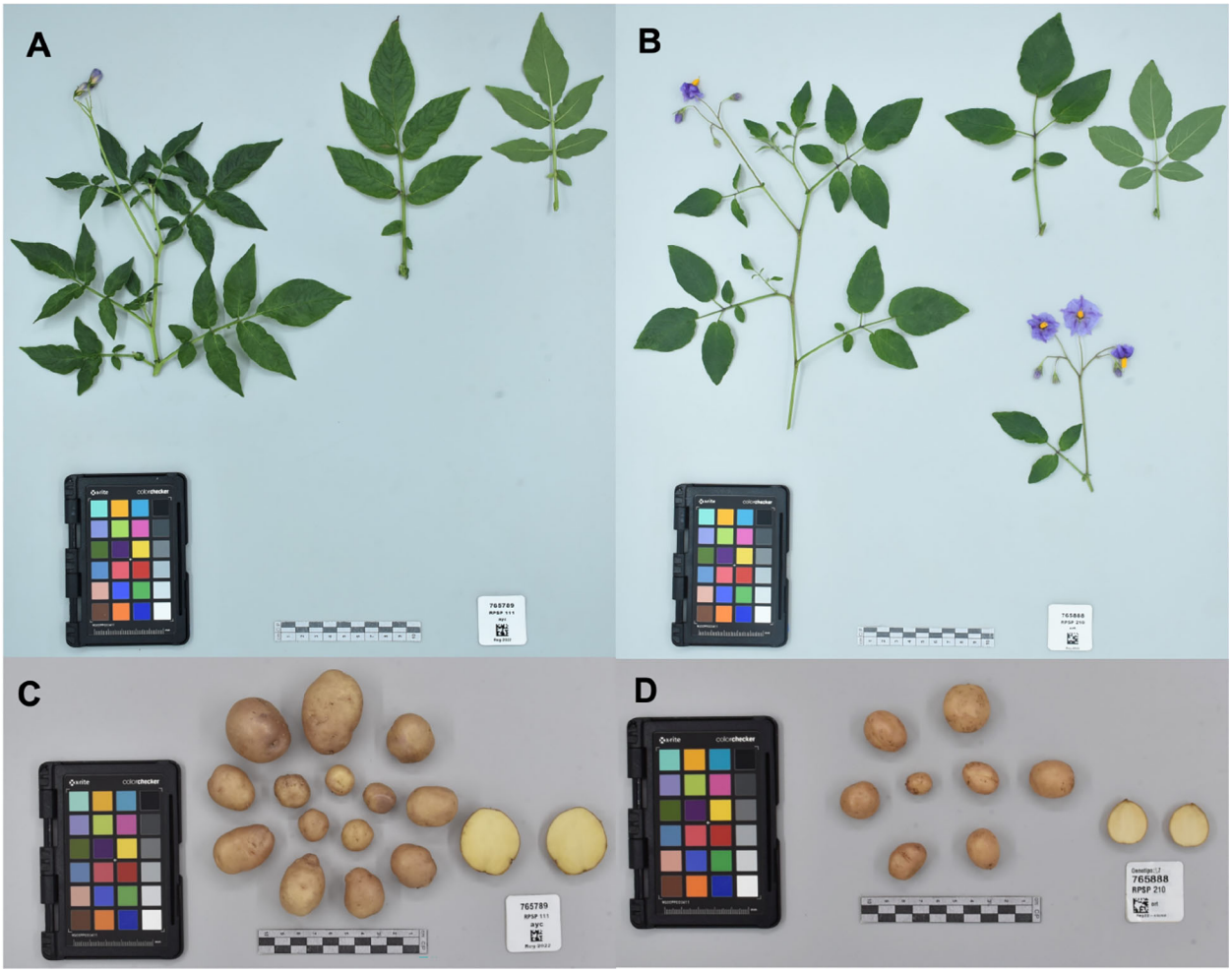


FIGURE 4
Herbarium specimen and tubers for: **(A, C)** *Solanum ayacuchense* Ochoa collected in Ayaerapampa, La Mar Province at the Department of Ayacucho. **(B, D)** *S. candolleanum* Berthault (*S. ortegae* Ochoa) collected in Paccaypata, Abancay Province at the Department of Huancavelica. No other accessions of these species are conserved in genebanks as far as the authors are aware. Photos taken by CIP's Potato Genebank.

sustainable conservation of potato CWRs into the future. This is further accentuated by the accelerated changes in the habitats of potato CWRs, as well as the effect of climate change, that could drive local extinction of some populations.

Data availability statement

The datasets presented in this study can be found in online repositories. Data is available on GENESYS: https://www.genesys-pgr.org/subsets/a55a79df-703c-46a7-9af8-c9e128d2a3bc

Author contributions

DE and CZ conceived the project. All authors conducted the collection. DS wrote the manuscript with support from DE, NA, and CZ. All authors contributed to the article and approved the submitted version.

Funding

This project was funded by the Global Crop Diversity Trust, through the Crop Wild Relatives Project and the CGIAR Genebank Platform.

Acknowledgments

We gratefully thank the Servicio Nacional Forestal y de Fauna Silvestre (SERFOR) of Peru for authorizing the collecting expeditions under permit number AUT-IFL-2016-038. We also thank Pedro Vergara, Wilder Rodriguez, Milton Gamarra, Edmundo Vilca, Marino Garay, Juan Tineo, Armando Linares, Jesús Arcos, Victor Nina, Severin Polreich, Edith Poma, Alina Gonzales, Vidal Mamani, Emma Manco, Julio Chávez and Ramiro Ortega for their committed participation in the collecting trips and multiplication activities. We also would like to acknowledge all the support we obtained from farmers and farming communities around Peru who gave us valuable information to navigate the roads and find the species we hoped to

collect. Finally, we thank the Global Crop Diversity Trust for their support of this collection campaign.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Aversano, R., Contaldi, F., Adelfi, M. G., D'Amelia, V., Diretto, G., De Tommasi, N., et al. (2017). Comparative metabolite and genome analysis of tuber-bearing potato species. *Phytochemistry* 137, 42–51. doi: 10.1016/j.phytochem.2017.02.011

Brummitt, N. A., Bachman, S. P., Griffiths-Lee, J., Lutz, M., Moat, J. F., et al. (2015). Green plants in the red: a baseline global assessment for the IUCN sampled red list index for plants. *PloS One* 10(8), e0135152. doi: 10.1371/journal.pone.0135152

Castañeda-Álvarez, N. P., de Haan, S., Juarez, H., Khoury, C. K., Achicanoy, H. A., Sosa, C. C., et al. (2015). Ex situ conservation priorities for the wild relatives of potato (Solanum l. section petota). *PloS One* 10 (4), e0122599. doi: 10.1371/journal.pone.0122599

Castañeda-Álvarez, N. P., Khoury, C. K., Achicanoy, H. A., Bernau, V., Dempewolf, H., Eastwood, R. J., et al. (2016). Global conservation priorities for crop wild relatives. *Nat. Plants* 2 (4), 16022. doi: 10.1038/nplants.2016.22

Dempewolf, H., Baute, G., Anderson, J., Kilian, B., Smith, C., and Guarino, L. (2017). Past and future use of wild relatives in crop breeding. *Crop Sci.* 57 (3), 1070–1082. doi: 10.2135/cropsci2016.10.0885

Eastwood, R. J., Tambam, B. B., Aboagye, L. M., Akparov, Z. I., Aladele, S. E., Allen, R., et al. (2022). Adapting agriculture to climate change: A synopsis of coordinated national crop wild relative seed collecting programs across five continents. *Plants* 11, 1840. doi: 10.3390/plants11141840

Ellis, D., Salas, A., Chavez, O., Gomez, R., and Anglin, N. (2020). "Ex situ conservation of potato [Solanum section petota (Solanaceae)] genetic resources in genebanks," in *The potato crop*(Springer, Cham), (pp. 109–138).

FAO (2014). *Genebank standards for plant genetic resources for food and agriculture. Rev. ed* (Rome).

Ford-Lloyd, B. V., Schmidt, M., Armstrong, S. J., Barazani, O., Engels, J., Hadas, R., et al. (2011). Crop wild relatives–undervalued, underutilized and under threat? *Bioscience* 61 (7), 559–565. doi: 10.1525/bio.2011.61.7.10

Hajjar, R., and Hodgkin, T. (2007). The use of wild relatives in crop improvement: A survey of developments over the last 20 years. *Euphytica* 156, 10–13. doi: 10.1007/s10681-007-9363-0

Harlan, J. R., and de Wet, J. M. (1971). Toward a rational classification of cultivated plants. *Taxon* 20 (4), 509–517. doi: 10.2307/1218252

Hawkes, J. G. (1990). *The potato: evolution, biodiversity and genetic resources* (Belhaven Press).

Hijmans, R., Spooner, D. M., Salas, A. R., Guarino, L., and Cruz, J. (2002). Atlas of wild potatoes. systematic and ecogeographic studies on crop genepools. *Int. Plant Genet. Resour. Institute (IPGRI)* 10), 130.

Holdridge, L. R. (1987). *Ecología basada en zonas de vida (No. 83)* (Agroamérica).

Humphreys, A. M., Govaerts, R., Ficinski, S. Z., Lughadha, E. N., and Vorontsova, M. S. (2019). Global dataset shows geography and life form predict modern plant extinction and rediscovery. *Nat. Ecol. Evol.* 3, 1043–1047. doi: 10.1038/s41559-019-0906-2

Jansky, S. H., Dempewolf, H., Camadro, E. L., Simon, R., Zimnoch-Guzowska, E., Bisognin, D. A., et al. (2013). A case for crop wild relative preservation and use in potato. *Crop Sci.* 53 (3), 746–754. doi: 10.2135/cropsci2012.11.0627

Jarvis, A., Lane, A., and Hijmans, R. J. (2008). The effect of climate change on crop wild relatives. *Agric. Ecosyst. Environ.* 126, 13–23. doi: 10.1016/j.agee.2008.01.013

Karki, H. S., Jansky, S. H., and Halterman, D. S. (2021). Screening of wild potatoes identifies new sources of late blight resistance. *Plant Dis.* 105, 368–376. doi: 10.1094/PDIS-06-20-1367-RE

Kumari, M., Kumar, M., and Solankey, S. S. (2018). "Breeding potato for quality improvement," in *Potato: From incas to all over the world*, 37.

Machida-Hirano, R. (2015). Diversity of potato genetic resources. *Breed. Sci.* 65 (1), 26–40. doi: 10.1270/jsbbs.65.26

Margules, C. R., and Pressey, R. L. (2000). Systematic conservation planning. *Nature* 405, 243–253. doi: 10.1038/35012251

Navarre, D. A., Goyer, A., and Shakya, R. (2009). "Nutritional value of potatoes: vitamin, phytonutrient, and mineral content," in *Advances in potato chemistry and technology* (Academic Press), (pp. 395–424).

Ochoa, C. M. (1999). *Las papas de sudamerica: Peru* (International Potato Center).

Perez, C., Nicklin, C., Dangles, O., Vanek, S., Sherwood, S. G., Halloy, S., et al. (2010). Climate change in the high Andes: Implications and adaptation strategies for small-scale farmers. *Int. J. Environmental Cultural Economic Soc. Sustainability* 6, 71–88.

Prescott-Allen, C., and Prescott-Allen, R. (1986). *The first resource: Wild species in the north american economy* (New Haven: Yale University).

Ramírez-Villegas, J. , Khoury, C., Jarvis, A., Debouck, D. G., and Guarino, L. (2010). A Gap Analysis Methodology for Collecting Crop Genepools: A Case Study with Phaseolus Beans , *PLOS ONE.* 5(10) e13497. doi: 10.1371/journal.pone.0013497

Rodríguez-Morata, C., Diaz, H. F., Ballesteros-Canovas, J. A., Rohrer, M., and Stoffel, M. (2019). The anomalous 2017 coastal El niño event in Peru. *Climate Dynamics* 52 (9), 5605–5622. doi: 10.1007/s00382-018-4466-y

Royal Botanic Gardens Kew (2016). *Peru Seed collecting guide.*

Salas, A., Gaspar, O., Rodríguez, W., Vargas, M., Centeno, R., and Tay, D. (2008). "Regeneration guidelines: wild potato. crop specific regeneration guidelines [CD-ROM]," in *CGIAR system-wide genetic resource programme*(Rome, Italy).

Sarkinen, T., Baden, M., Gonzales, P., Cueva, M., Giacomin, L. L., Spooner, D. M., et al. (2015). *Annotated checklist of solanum l. (Solanaceae) for Peru* (Revista Peruana de Biología).

Singh, S. P. (2001). Broadening the genetic base of common bean cultivars: A review. *Crop Sci.* 41, 1659–1675. doi: 10.2135/cropsci2001.1659

Smýkal, P., Nelson, M. N., Berger, J. D., and Von Wettberg, E. J. (2018). The impact of genetic changes during crop domestication. *Agronomy* 8 (7), 119.

Spooner, D. M., Ghislain, M., Simon, R., Jansky, S. H., and Gavrilenko, T. (2014). Systematics, diversity, genetics, and evolution of wild and cultivated potatoes. *Botanical Rev.* 80, 283–383. doi: 10.1007/s12229-014-9146-y

Spooner, D. M., and Salas, A. (2006). "Structure, biosystematics, and germplasm resources," in *Handbook of potato production, improvement, and post-harvest management*. Ed. J. Gopal (Binghampton, NY: The Haworth Press), 1–39.

Syfert, M. M., Castañeda-Álvarez, N. P., Khoury, C. K., Särkinen, T., Sosa, C. C., Achicanoy, H. A., et al. (2016). Crop wild relatives of the brinjal eggplant (Solanum melongena): Poorly represented in genebanks and many species at risk of extinction. *Am. J. Bot.* 103 (4), 635–651. doi: 10.3732/ajb.1500539

Tanksley, S. D., and McCouch, S. R. (1997). Seed banks and molecular maps: Unlocking genetic potential from the wild. *Science* 277, 1063–1066. doi: 10.1126/science.277.5329.1063

Vargas, F., Liau, S., Grandez, C., Anglin Barkley, N., and Ellis, D. (2016). . *collection and preservation of herbarium specimens for the CIP genebank. CIP-OP252. v.1.* 1–36 pp.

Vilchez, D., Sotomayor, D. A., and Zorrilla, C. (2019). Ex situ conservation priorities for the Peruvian wild tomato species (Solanum l. SECT. lycopersicum (MILL.) WETTST.). *Ecologia Aplicada* 18, 2.

Vincent, H., Amri, A., Castañeda-Álvarez, N. P., Dempewolf, H., Dulloo, E., Guarino, L., et al. (2019). Modeling of crop wild relative species identifies areas globally for *in situ* conservation. *Commun. Biol.* 2, 136. doi: 10.1038/s42003-019-0372-z

Wettberg, E. V., Toker, C., Ozkan, H., and & Smykal, P. (2022). "Endangered wild crop relatives of the fertile crescent," in *Imperiled: The encyclopedia of conservation*. Eds. D. A. DellaSala and M. I. Goldstein (Elsevier), 673–682.

Wilkes, G. (2007). Urgent notice to all maize researchers: disappearance and extinction of the last wild teosinte population is more than half completed. a modest proposal for teosinte evolution and conservation *in situ*: the balsas, Guerrero, Mexico. *Maydica* 52, 49–58.

frontiers | Frontiers in Plant Science

TYPE Review
PUBLISHED 20 February 2023
DOI 10.3389/fpls.2022.1046702

Check for updates

OPEN ACCESS

EDITED BY
Wellington Ronildo Clarindo,
Universidade Federal de Viçosa, Brazil

REVIEWED BY
Leonardo Lopes Bhering,
Federal University of Viçosa, Brazil
Leandro Giacomin,
Federal University of Paraíba, Brazil

*CORRESPONDENCE
Rodrigo Nicolao
rodnicolao@gmail.com

SPECIALTY SECTION
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

RECEIVED 16 September 2022
ACCEPTED 28 December 2022
PUBLISHED 20 February 2023

CITATION
Nicolao R, Gaiero P, Castro CM and Heiden G (2023) *Solanum malmeanum*, a promising wild relative for potato breeding. *Front. Plant Sci.* 13:1046702. doi: 10.3389/fpls.2022.1046702

COPYRIGHT
© 2023 Nicolao, Gaiero, Castro and Heiden. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Solanum malmeanum, a promising wild relative for potato breeding

Rodrigo Nicolao[1]*, Paola Gaiero[2], Caroline M. Castro[3] and Gustavo Heiden[3]

[1]Programa de Pós-Graduação em Agronomia/Fitomelhoramento - Universidade Federal de Pelotas (UFPel), Pelotas, RS, Brazil, [2]Departamento de Biología Vegetal, Facultad de Agronomía, Universidad de la República, Montevideo, Uruguay, [3]Laboratório de Recursos Genéticos, Embrapa Clima Temperado, Pelotas, RS, Brazil

Crop wild relatives are gaining increasing attention. Their use in plant breeding is essential to broaden the genetic basis of crops as well as to meet industrial demands, for global food security and sustainable production. *Solanum malmeanum* (*Solanum* sect. *Petota*, Solanaceae) is a wild relative of potatoes (*S. tuberosum*) from Southern South America, occurring in Argentina, Brazil, Paraguay and Uruguay. This wild potato has been largely mistaken for or historically considered as conspecific with *S. commersonii*. Recently, it was reinstated at the species level. Retrieving information on its traits and applied uses is challenging, because the species name has not always been applied correctly and also because species circumscriptions and morphological criteria applied to recognize it have not been consistent. To overcome these difficulties, we performed a thorough literature reference survey, herbaria specimens' identification revision and genebank database queries to review and update the information available on this potato wild relative, contributing to an increase in research on it to fully understand and explore its potential for potato breeding. Scarce studies have been carried out concerning its reproductive biology, resistance against pests and diseases as well as tolerance to abiotic stresses and evaluation of quality traits. The scattered information available makes it less represented in genebanks and genetic studies are missing. We compile, update and present available information for *S. malmeanum* on taxonomy, geographical distribution, ecology, reproductive biology, relationship with its closest relatives, biotic and abiotic stresses resistance and quality traits and discuss ways to overcome sexual barriers of hybridization and future perspectives for its use in potato breeding. As a final remark, we highlight that this species' potential uses have been neglected and must be unlocked. Thus, further studies on morphological and genetic variability with molecular tools are fundamental for an efficient conservation and applied use of this promising genetic resource.

KEYWORDS

conservation, crop wild relatives, food security, genetic resources, genebank, germplasm, *Petota*, Solanaceae

1 Introduction

Potato (*Solanum tuberosum* L.) is the third most important food crop in terms of global food security, after rice and wheat and the most important non-grain food crop (Birch et al., 2012). It is a prominent source of calories, fibers, proteins, mineral micronutrients, vitamins B6 and C, potassium, and antioxidant compounds, and can pack all those nutrients while using less area (m^2) than any other crop (Birch et al., 2012; Devaux et al., 2014). Conventional potato cultivars are bred at the tetraploid level, have high heterozygosity, reproduce vegetatively and have less genetic variation than their wild relatives (Fu et al., 2009; Hardigan et al., 2015; Kolech et al., 2016; Hardigan et al., 2017; Duan et al., 2019; Drapal et al., 2020). Climate change and the dispersion of pests and new races of pathogens are currently a matter of concern among potato breeders and farmers. The outbreaks caused by new isolates of the late blight oomycete *Phytophthora infestans* (Mont.) de Bary in Europe and the Americas (Santana et al., 2013; Göre et al., 2019; Maurice et al., 2019), or by the new sequevars of bacterial wilt *Ralstonia solanacearum* found in Brazil, Peru and Iran (Gutarra et al., 2017; Santiago et al., 2017; Sedighian et al., 2020) have the potential to destroy potato fields. Lack of resistance of the current potato cultivars to the main diseases that affect the crop makes it one of the most dependent crops on agrochemical applications (Vleeshouwers et al., 2011). This scenario requires genetic variability for the continuous process of crop improvement to face biotic stressors and climate change, while increasing productivity to ensure global food security.

Crop Wild Relatives (CWR) are a reservoir of valuable genetic diversity for crop improvement, especially those genes that confer resistance to pests and diseases, tolerance to abiotic stressors and climate change scenarios, as well as providing nutritional compounds (Maxted et al., 2006; Jansky et al., 2013; Zhang et al., 2017; Pradheep et al., 2019; Perrino and Perrino, 2020). CWR contain pharmaceutical and industrial properties and can potentially contribute for a sustainable and smarter agroecosystem way (Jansky et al., 2013; Dempewolf et al., 2014; Warschefsky et al., 2014).

The cultivated potato has more than 100 wild relative species (*Solanum* sect. *Petota*, Solanaceae) (Douches et al., 1989; Hawkes, 1990; Pavek and Corsini, 2001; Spooner et al., 2014; Hardigan et al., 2017). Potato wild relatives (PWR) are widely distributed across the Americas, from the geographical and ecological points of view, ranging from Southwestern United States (38°N), through the dry areas of Mexico, until the adjacent mainland areas of Chile (41°S) in the Southern Cone of South America, at elevation ranging from sea level up to 4,500 m (Hijmans and Spooner, 2001; Spooner et al., 2004; Spooner et al., 2014; Spooner et al., 2016; Spooner et al., 2019). About 70% of these species are diploids while autopolyploids and allopolyploids go up to 6x chromosome numbers (x=12). PWR possess a common (basic) A genome, which, modified to different degrees, originated four additional genomes: B, C, D and E (Matsubayashi, 1991). Furthermore, the differentiation in the basic A genome is assumed to play a minor role as an isolation mechanism (Camadro et al., 2004).

To guide potato breeders in the efficient use of PWR according to their gene pool level, Spooner and collaborators (2014) proposed five crossability groups, based on the Endosperm Balance Number (EBN) along the ploidy, and self-compatible/self-incompatible systems. Diploid potato species (1EBN and 2EBN) are mostly self-incompatible (Cipar et al., 1964; Spooner et al., 2014). All 4EBN potato species are self-compatible due to a lack of pre-zygotic gametophytic barriers (Spooner et al., 2014). A self-compatible plant of *S. chacoense* was found (Hanneman, 1985) and the responsible *S-loci-inhibitor* (*Sli*) gene was identified and mapped to chromosome 12 (Hosaka and Hanneman, 1998a; b). The EBN post-zygotic barrier of hybridization acts at the embryo-endosperm development level, leading to aberrations in chromosome pairing and is observed through embryonic abortion, male sterility and breakdown of F1 or F2 hybrid offspring (Camadro et al., 2004; Camadro et al., 2012). According to Johnston et al. (1980), this post-zygotic barrier is explained under the hypothesis of the Endosperm Balance Number (EBN), in which *Solanum* sect. *Petota* species have a true ploidy number (determined by their actual number of chromosomes) and an "effective" ploidy generated by the action of hypothetical genetic factors (called EBN) (Johnston et al., 1980). The EBN system requires a 2:1 maternal to paternal EBN dosage for a normal endosperm development, which could predict the success of hybridization of some specific crosses. EBN is under genetic control (Ehlenfeldt and Hanneman, 1988; Camadro and Mansuelli, 1995; Johnston and Hanneman, 1996) and its values have been determined empirically for most PWR species by crossing each one with a standard species of known EBN (Hanneman, 1994; Mansuelli and Camadro, 1997). Due to the strong post-zygotic barrier of EBN differences, 2x (1EBN) species are sexually isolated from 2x (2EBN) and both 4x (4EBN) (Camadro and Peloquin, 1981; Hanneman, 1994).

Solanum malmeanum Bitter (Solanaceae) is a 1EBN PWR species, belonging to *Solanum* sect. *Petota* Dumort. ser. *Commersoniana* Buk., classified into the tertiary gene pool of potato (*S. tuberosum* L.) (Hanneman, 1994; Spooner et al., 2014). *Solanum malmeanum* is native to the Southern Cone of South America, occurring in Argentina, Brazil, Paraguay and Uruguay. This PWR was collected for the first time in 1893 by the Swedish botanist Gustaf Oskar Andersson Malme along the Ijuí river, at Rio Grande do Sul state in Southern Brazil (53° 54' 53" W, 28° 23' 16" S) and the new species was named after him when formally described later by the German botanist Friedrich August Georg Bitter in 1913 (Bitter, 1913). Taxonomic treatments of PWR by Bitter (1913); Correll (1962); Hawkes and Hjerting (1969); Mentz and Oliveira (2004); Matesevach and Barboza (2005), and Spooner et al. (2016) based on morphological traits, considered *S. malmeanum* as *S. malmeanum*, *S. commersonii* f. *malmeanum*, *S. commersonii* subsp. *malmeanum*, *S. commersonii* f. *malmeanum*, *S. commersonii* subsp. *malmeanum*, and *S. malmeanum*, respectively. *Solanum malmeanum* is a long day adapted PWR, highly flowering and a good pollen producer, having a geographical distribution that partially overlaps with *S. chacoense* Bitter and *S. commersonii* Dunal (Correll, 1962; Hawkes and Hjerting, 1969; Spooner et al., 2016). After a long history of being considered conspecific with *S. commersonii*, recently Spooner et al. (2014); Spooner et al. (2016) have reinstated the taxon at species level.

Solanum malmeanum populations have co-evolved with many diverse environmental conditions, including humid tropical and subtropical areas, semidesert conditions (e.g., dry, cold, frost, heat,

salinity) and biotic stressors (Hawkes, 1958; Hawkes and Hjerting, 1969). Therefore, *S. malmeanum* stands out as a promising important source of allelic diversity and valuable agronomic traits for potato breeding, including resistance to pests and diseases, abiotic stress tolerance and industrial traits (Douches et al., 1989; Hanneman, 1989; Flanders and Radcliffe, 1992; Spooner and Bamberg, 1994; Jansky, 2000; Siri et al., 2009; Jansky et al., 2013; Bachmann-Pfabe and Dehmer, 2020).

Due to a changing taxonomic history, several morphological shared traits and geographical distribution patterns that greatly overlap, *S. malmeanum* has been largely confused or historically considered as conspecific with *S. commersonii*. Nevertheless, it was reinstated at the species level (Spooner et al., 2016). Due to the instability of the cohesive application of this species name and discordant species circumscriptions and morphological borders applied to recognize it along the taxonomic history, retrieving information on its traits and applied uses is tricky. To overcome these difficulties, we performed a thorough literature review, genebank database queries and a revision of herbaria voucher specimens cited in published studies to update the information available on this wild potato relative, aiming to broadening the scientific knowledge on the species to fully understand and explore its potential for potato breeding. Thus, our study aims to compile, synthetize and update all the scattered information about reproductive biology and cytogenetics, taxonomic treatments and geographical data, biotic and abiotic resistance of *S. malmeanum*, providing it in an elucidative and comprehensive way.

2 Materials and methods

2.1 Species circumscription

Our circumscription of *S. malmeanum* is based on the species original publication by Bitter (1913), the revised taxonomic treatment of Spooner et al. (2016); Nicolao (2021) and complementary observations from characterized genebank accessions available at database AleloVegetal (https://av.cenargen.embrapa.br/avconsulta/Home/index.do). When the original data discussed in this review was formerly published as *S. commersonii*, but the cited accession or specimen corresponds to *S. malmeanum*, we added the remark "as *S. commersonii*" right after.

2.2 Data review

To perform this review, we searched for "*Solanum commersonii*", "*Solanum malmeanum*", "*Solanum commersonii* subsp. *malmeanum*", "*Solanum commersonii* f. *malmeanum*" and their abbreviated forms in literature, herbaria and genebank databases as follows.

2.3 Literature consulted

We reviewed the monographs for taxonomic treatments of tuber-bearing *Solanum* species conducted by Bitter (1913); Correll (1962); Hawkes and Hjerting (1969), and Spooner et al. (2016), as well as regional taxonomic studies for Argentina by Matesevach and Barbosa (2005), South of Brazil by Mentz and Oliveira (2004) and Paraguay by De Egea et al. (2018). The scientific reports published until October 31st 2022 were searched in the databases Base de Dados da Pesquisa Agropecuária (BDPA 2022), Biblioteca Digital Brasileira de Teses e Dissertações (BDBTD 2022), CAPES Portal de Periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (2022), Google Scholar (2022), JSTOR (2022), Scientific Electronic Library Online (SciELO 2022), ScienceDirect system (2022), and printed publications available at the Embrapa Clima Temperado library.

2.4 Taxonomic and geographical distribution data

To update the taxonomic and geographic data for *S. malmeanum*, we reviewed, either through specimens loans or online, the label data of specimens from the herbaria B, BAL, BHCB, BM, CEN, ECT, F, ICN, K, MVFA, MVJB, MVM, MO, NY, P, R, RB, S, SPF, U, and US (acronyms according to Thiers 2022). Georeferenced records were extracted from the online platform databases Global Biodiversity Information Facility, speciesLink. Duplicate records were removed. Manual checking and correction of wrong or doubtful coordinates were carried out in GEOlocate (http://www.geo-locate.org/), Google Earth and Google Maps. The occurrence data were plotted on a map rendered in QGIS (2022).

2.5 Genebank databases search

To assess the representativeness of *S. malmeanum* in the germplasm banks we explored the online databases AleloVegetal (2022) by the Brazilian Agricultural Research Corporation (EMBRAPA) and Genesys (2022), a Global Portal on Plant Genetic Resources for Food and Agriculture (PGRFA). We also accessed offline data from genebank curators from Instituto Nacional de Tecnología Agropecuária (INTA) of Argentina and Instituto Nacional de Investigación Agropecuaria (INIA) of Uruguay. Thus, passport information of *Solanum malmeanum* from INIA and INTA were obtained personally because they host critical collections of the focused species, although their data are not available at Genesys.

2.6 Agronomical traits

To compile the agronomical traits (i.e., biotic resistance, abiotic tolerance levels and quality traits), we explored the online databases Genesys (2022) and literature cited accordingly.

2.7 Criteria for updating taxonomic information

We checked and updated the taxonomic determinations of retrieved records of *S. commersonii* and *S. malmeanum* for taxonomic accuracy through the revision of voucher specimens

deposited in herbaria or other verifiable sources as genebank pictures of the plants under cultivation. The confirmation, updating or corrections of identifications were possible when the studied specimens matched the current circumscription of *S. malmeanum* based on the observation of diagnostic morphological features present in the physical specimens or in the high-quality digital images retrieved. Accessions cited in previously published works as "*S. commersonii*" and "*S. commersonii* subsp. *malmeanum*" were checked by its code accession at the Germplasm Resources Information Network (GRIN Taxonomy, 2022), to confirm their correct taxonomic identification.

3 Results

3.1 Taxonomic history, habitat and geographical distribution

Bitter (1913), when describing *S. malmeanum* as a new species to science, emphasized that it can be confused with *S. commersonii* due to their similar habit. Later, Correll (1962) downgrades *S. malmeanum* to a form of *S. commersonii* because he considers it could represent a result of hybridization between *S. commersonii* and *S. chacoense* stating that "*Solanum* form *malmeanum*" differs from the typical form of *S. commersonii* because the former contains 4-5 lateral leaflets pairs and two or more lateral interstitial segments while the typical form of *S. commersonii* contains 2-5 lateral leaflet segments pairs, without or rarely bearing lateral interstitial segments pairs. Another difference on flower characteristic reported by Correll (1962) was that *S. malmeanum* has its corolla narrowly and deeply lobed, often to near the base than *S. commersonii*.

Hawkes and Hjerting (1969) states that Correll was mistaken in classifying *S. malmeanum* as a form of *S. commersonii* because they consider it consists of an assemblage of biotypes with a distinct geographical distribution which are adapted to a habitat range different from that of the typical *S. commersonii*. According to them, *S. malmeanum* has lateral leaflet segments generally decreasing gradually to the leaf base and are narrowly decurrent, being normally petiolulate, while the inflorescence peduncle is unbranched or has the branches not markedly contracted with pedicels articulation median to higher than in *S. commersonii*, and corollas are always white and the lobes about as long as broad. Conversely, *S. commersonii* shows lateral leaflet segments decreasing rapidly in size towards the base of the leaf, being often markedly decurrent and normally sessile, while the inflorescence peduncle branches and is somewhat contracted with pedicels articulation lower to median, and corollas are generally tinged with purple and have the lobes about 1 and a ½ time as long as broad or even longer than in *S. malmeanum*. The two authors reinforced that the two taxa could almost be considered to differ from each other at the species level.

The last comprehensive taxonomic treatment including *S. malmeanum* was by Spooner and collaborators (2016). According to them, *S. malmeanum* can be differentiated from *S. commersonii* by its generally subequal uppermost lateral leaflets that do not decrease rapidly in size towards the leaf base, by its generally petiolulate and larger lateral interstitial segments and by its white corollas. The corollas are almost white, although in a very few cases they have observed a small brushstroke on the abaxial midribs of the corolla lobes very different from the characteristic dark stripes along the abaxial midribs of the corolla lobes in flowers of *S. commersonii*. Maybe, these specimens could represent hybrids between *S. malmeanum* and *S. commersonii*.

Solanum malmeanum is native to the Southern Cone of South America, in Argentina, Brazil, Paraguay and Uruguay. It is found in diverse environments such as shady forests, grasslands, damp pastures, roadsides, stream sides, as well as invasive at crop cultivations, including potato fields (Hawkes and Hjerting, 1969). In Argentina it is found in the Eastern Subtropical Biogeographic Province at grasslands or savannas and the Mixed Forest Biogeographic Districts, Chaco Biogeographic Province by Eastern Chaco District (more humid eastern parts) and along the Espiñal Biogeographic Province by Santa Fé and Entre Ríos (Hawkes and Hjerting, 1969). In Brazil, it occurs in the *Araucaria angustifolia* subtropical Mixed Forests of the Paranense Biogeographic Province, with high annual rainfall, as well as in the Pampean Biogeographic Province with a dominant vegetation of grasses and bushes (Hawkes and Hjerting, 1969). In Paraguay, it spreads through the Mixed Forest Biogeographic District into the wetter parts of the Chaco Biogeographic Province and the Chaqueño Biogeographic Domain that are characterized by xerophytic vegetation (Hawkes and Hjerting, 1969). In Uruguay it is found mainly in the so-called 'Galería Uruguayense', that comprises the forests along the Uruguay river basin, belonging to the Mixed Forest Biogeographic District of the Eastern Subtropical Biogeographic Province and to the Pampean Biogeographic Province, with a dominant vegetation of grasses and bushes (Hawkes and Hjerting, 1969; Morrone et al.,2014; Spooner et al., 2016). Records found in the surveyed herbaria and genebanks range from sea level up to 765 m of elevation (AleloVegetal, 2022; Global Biodiversity Information Facility, 2022).

3.2 Herbarium representativeness

A total of 281 records for *S. malmeanum* specimens are preserved in many herbaria, of which 197 with geographical coordinates were plotted (Figure 1).

3.3 Genebank representativeness

Currently, *S. malmeanum* is represented in genebanks by a total of 89 accessions worldwide (Table 1). The *in vitro* collection at the Instituto Nacional de Investigación Agropecuaria (INIA) Las Brujas, Uruguay in collaboration with Facultad de Agronomía of Universidad de La Republica (Udelar) Uruguay, contains the widest representativeness of *S. malmeanum* conserved ex-situ (36 accessions) followed by U.S. National Plant Germplasm System (NPGS) USDA (25 accessions), and Embrapa Potato Genebank (11 accessions), Brazil. The External Branch North of the Department Genebank of Germany, the Center for Genetic Resources (CGN) of the Netherlands, Instituto Nacional de Tecnología Agropecuária (INTA) of Argentina and the James Hutton Institute (JHI) of United Kingdom account together for a total of 17 accessions.

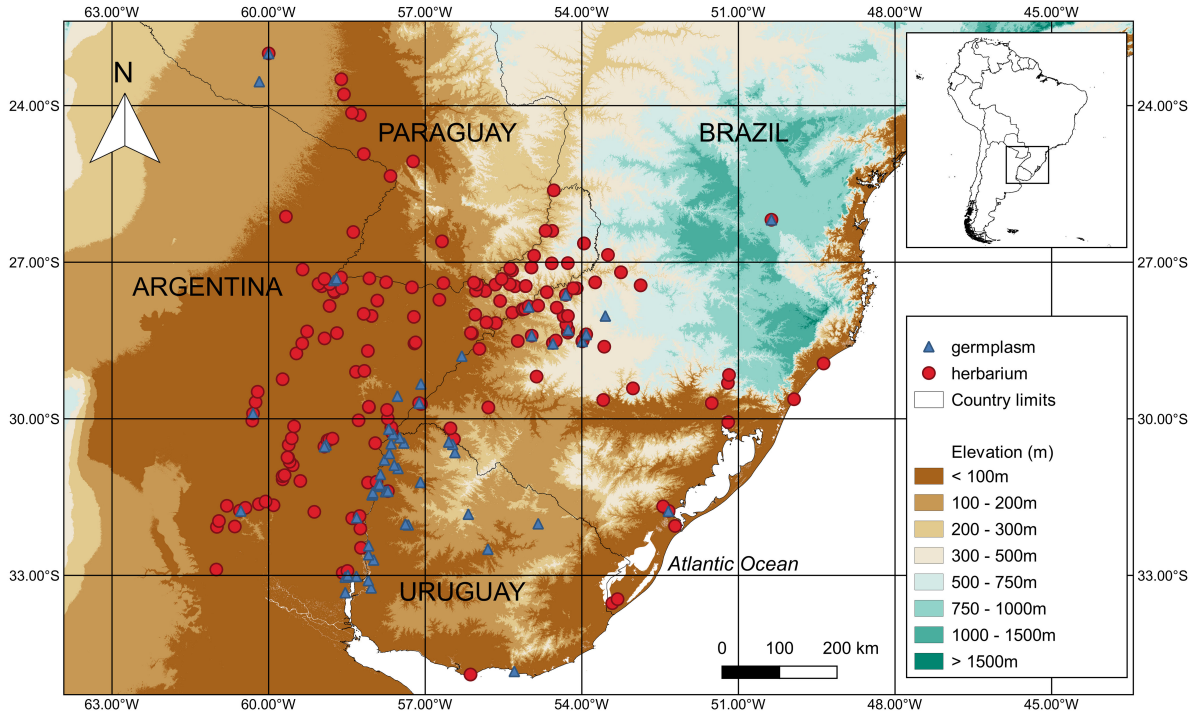


FIGURE 1
Geographical distribution of *Solanum malmeanum* (Solanaceae) in Southern South America. Occurrence points are represented by red circles for herbarium specimens and blue triangles for accessions in genebanks.

A total of 89 collected samples of *S. malmeanum* are ex-situ conserved in genebanks (G) (Table 2), whereas only four collected samples are conserved in duplicates (OKADA5071, OKADA7270, OKADA7292, and OKADA7254). From the 89 collected samples, 81 with geographic coordinates are represented in Figure 1. For eight collected samples from G it was not possible to retrieve geographic coordinates data.

3.4 Genetic variability characterization

Although there are only a few studies on genetic variability which include accessions from *Solanum malmeanum*, in general, the results point at a close relationship with *S. commersonii* but allow to distinguish both species from each other and from *S. chacoense* (Lester, 1965; Jacobs et al., 2008; Siri et al., 2009). Through immune-electrophoretic analysis, one shared tuber antigen found in *S. malmeanum* and *S. commersonii* was absent from *S. chacoense*, pointing to a closer relationship between these first two species (Lester, 1965).

In order to analyze genetic variability within the species and its relationship with closely related species, molecular markers have been applied among different accessions from *S. malmeanum* across its distribution. Partial sequences of a new retrotransposon *Tnt1* family member, Retrosol, were reported in several *Solanum* sect. *Petota* species, including *S. malmeanum* (Manetti et al., 2007). The accession BAL80009 presented three different fragment sequences of Tnt1 with an average of 423+/-(5) base pairs (bp) and 0.093 πJC nucleotide diversity with Jukes and Cantor (1969) corrections.

Phylogenetic analyses using AFLP markers confirmed that *S. malmeanum* can be genetically separated from *S. chacoense* and *S.*

TABLE 1 Representativeness of *Solanum malmeanum* (Solanaceae) accessions at genebanks (Genesys, 2022; WIEWS, 2022).

Country	Institution/genebank	Genebank code	Number of accessions
Brazil	Brazilian Agricultural Research Corporation (EMBRAPA)	BRA020	11
Netherlands	Centre for Genetic Resources, the Netherlands (CGN)	NDL037	6
Germany	External Branch North of the Department Genebank (IPK)	DEU159	8
Uruguay	Instituto Nacional de Investigación Agropecuaria (INIA)	URY008, URY032	36
Argentina	Instituto Nacional de Tecnología Agropecuaria (INTA)	–	1
United Kingdom	The James Hutton Institute (JHI)	GBR251	2
USA	The U.S. National Plant Germplasm System (NPGS)	USA004	25
	Total		89

10.3389/fpls.2022.1046702

TABLE 2 Summary of *Solanum malmeanum* (Solanaceae) accessions at genebanks following Country, Collection Code, Accession Code, Latitude and Longitude (Genesys, 2022).

Country	Collection Code	Accession Code	Latitude	Longitude
ARGENTINA	BAL	2015020	-	–
BRAZIL	BRA020	BRA 00183754-1	-31.772	-52.342
BRAZIL	BRA020	BRA 00167436-5	-31.772	-52.343
BRAZIL	BRA020	BRA 00167432-4	-28.563	-54.554
BRAZIL	BRA020	BRA 00167071-0	-28.517	-53.992
BRAZIL	BRA020	BRA 00167433-2	-28.408	-54.961
BRAZIL	BRA020	BRA 00167093-4	-28.299	-54.263
BRAZIL	BRA020	BRA 00183755-8	-27.856	-55.016
BRAZIL	BRA020	BRA 00183756-6	-27.856	-55.016
BRAZIL	BRA020	BRA 00167137-9	-27.626	-54.308
BRAZIL	BRA020	BRA 00183778-0	-26.187	-50.368
BRAZIL	BRA020	BRA 00167394-6	-26.187	-50.368
GERMANY	DEU159	WKS 32184	-33.020	-58.320
GERMANY	DEU159	WKS 35339	-23.540	-60.180
GERMANY	DEU159	WKS 32181	–	–
GERMANY	DEU159	WKS 35341	-28.030	-53.550
GERMANY	DEU159	WKS 35340	-23.000	-60.000
GERMANY	DEU159	AKS 23002	–	–
GERMANY	DEU159	WKS 32182	–	–
GERMANY	DEU159	WKS 32183	–	–
NETHERLANDS	NLD037	CGN22351	-33.331	-58.533
NETHERLANDS	NLD037	CGN22352	-30.533	-58.933
NETHERLANDS	NLD037	CGN18215	-30.200	-57.700
NETHERLANDS	NLD037	CGN21353	-29.700	-57.117
NETHERLANDS	NLD037	CGN18025	-29.567	-57.533
NETHERLANDS	NLD037	CGN18329	-32.500	-55.800
UNITED KINGDOM	GBR251	CPC 7520	–	–
UNITED KINGDOM	GBR251	CPC 7058	–	–
UNITED STATES	USA004	PI 472849	-30.500	-58.900
UNITED STATES	USA004	PI 472843	-33.033	-58.533
UNITED STATES	USA004	PI 472846	-33.033	-58.533
UNITED STATES	USA004	PI 472844	-33.000	-58.483
UNITED STATES	USA004	PI 472840	-31.894	-58.315
UNITED STATES	USA004	PI 472855	-31.766	-60.533
UNITED STATES	USA004	PI 472850	-30.533	-58.933
UNITED STATES	USA004	PI 472854	-30.533	-58.933
UNITED STATES	USA004	PI 472845	-30.200	-57.700
UNITED STATES	USA004	PI 498416	-29.900	-60.300
UNITED STATES	USA004	PI 472847	-29.700	-57.116

(Continued)

10.3389/fpls.2022.1046702

TABLE 2 Continued

Country	Collection Code	Accession Code	Latitude	Longitude
UNITED STATES	USA004	PI 458318	-29.566	-57.533
UNITED STATES	USA004	PI 472841	-29.566	-57.533
UNITED STATES	USA004	PI 472851	-29.566	-57.533
UNITED STATES	USA004	PI 472842	-29.333	-57.083
UNITED STATES	USA004	PI 472848	-29.333	-57.083
UNITED STATES	USA004	PI 414154	-28.800	-56.300
UNITED STATES	USA004	PI 472853	-27.350	-58.750
UNITED STATES	USA004	PI 472852	-27.316	-58.700
UNITED STATES	USA004	PI 320269	-28.383	-53.916
UNITED STATES	USA004	PI 498417	-28.383	-53.916
UNITED STATES	USA004	PI 498419	-28.383	-53.916
UNITED STATES	USA004	PI 498408	–	–
UNITED STATES	USA004	PI 498418	-23.000	-60.000
UNITED STATES	USA004	PI 320268	-31.387	-57.716
URUGUAY	INIA	M6P6	-34.840	-55.294
URUGUAY	INIA	RN9P2	-33.239	-58.037
URUGUAY	INIA	RN5P2	-33.098	-58.091
URUGUAY	INIA	RN4P1	-32.705	-57.989
URUGUAY	INIA	RN3P1	-32.604	-58.086
URUGUAY	INIA	P3P2	-32.432	-58.085
URUGUAY	INIA	P3P1	-32.432	-58.085
URUGUAY	INIA	P1P1	-32.032	-57.326
URUGUAY	INIA	P2P2	-32.021	-57.380
URUGUAY	INIA	P2P1	-32.021	-57.380
URUGUAY	INIA	R10P1	-32.011	-54.837
URUGUAY	INIA	T3P2	-31.828	-56.175
URUGUAY	INIA	S1P19	-31.458	-58.020
URUGUAY	INIA	S1P12	-31.456	-58.020
URUGUAY	INIA	S36P1	-31.445	-58.010
URUGUAY	INIA	S35P1	-31.415	-58.005
URUGUAY	INIA	S32P1	-31.396	-57.761
URUGUAY	INIA	S31P1	-31.390	-57.714
URUGUAY	INIA	S39P1	-31.295	-57.867
URUGUAY	INIA	S38P1	-31.272	-57.906
URUGUAY	INIA	S1P1	-31.257	-57.875
URUGUAY	INIA	S5P3	-31.217	-57.091
URUGUAY	INIA	S5P2	-31.217	-57.091
URUGUAY	INIA	S43P1	-31.065	-57.859
URUGUAY	INIA	S43P2	-31.065	-57.859
URUGUAY	INIA	S41P1	-30.951	-57.521

(Continued)

TABLE 2 Continued

Country	Collection Code	Accession Code	Latitude	Longitude
URUGUAY	INIA	S42P1	-30.889	-57.595
URUGUAY	INIA	S40P1	-30.789	-57.783
URUGUAY	INIA	A12P1	-30.665	-57.687
URUGUAY	INIA	A2P3	-30.646	-56.432
URUGUAY	INIA	A6P1	-30.490	-56.475
URUGUAY	INIA	A11P1	-30.478	-57.654
URUGUAY	INIA	A7P2	-30.464	-57.409
URUGUAY	INIA	A3P2	-30.445	-56.551
URUGUAY	INIA	A8P1	-30.364	-57.492
URUGUAY	INIA	A9P2	-30.292	-57.615

commersonii, because two of 12 accessions of *S. commersonii* that were allocated in the same group with seven *S. malmeanum* accessions (Jacobs et al., 2008), were later confirmed to in fact belong to *S. malmeanum* (CGN 22351, GLKS 35340). Siri et al. (2009) assessed the genetic diversity of *S. commersonii* and *S. malmeanum* from different geographical areas of Uruguay by using different PCR-based markers (three primers for AFLP, 10 primers for RAPD, and four primers for SSR). They found a high consistency of clustering with the geographic origin of accessions, with *S. malmeanum* (cluster B) from the north of Uruguay and *S. commersonii* (cluster A) from the south, previously characterized in morphology (see Siri et al., 2009). When associated with morphological features, the PCR-based markers were able to discriminate *S. malmeanum* from *S. commersonii*. Few studies on the geographic distribution of genetic diversity for *S. commersonii* and *S. malmeanum* are available. In addition, the entire potential area in which both taxa are geographically distributed has not yet been explored, and further studies are needed for a complete and more accurate coverage of the geographic distribution.

3.5 Agronomical traits

Many interesting traits have been found in *S. malmeanum* and are summarized in Table 3. Most of them are not found in the cultivated potato gene pool, for example, some accessions are resistant against bacterial (*Ralstonia solanacearum* (Smith) Yabuuchi) and verticillium wilt (*Verticillium dahliae* Kleb.); ring rot (*Corynebacterium michiganensis* subsp. *sepedonicus* Spieck. & Kotth.); late (*Phytophthora infestans* Mont.) and early blight (*Alternaria solani* Elis and Martin); fusarium dry rot (*Fusarium sambucinum* Fuckel); hapla (*Meloidogyne hapla* Chitwood) and cyst nematode (*Globodera rostochiensis* Wollenweber); colorado potato beetle (*Leptinotarsa decemlineata* Say); potato leaf hopper (*Empoasca fabae* Harris); green peach aphid (*Myzus persicae* Sulzer); potato aphid (*Macrosiphum euphorbiae* Thomas); and potato leafroll virus (PLRV) (*Polerovirus* sp.) (Radcliffe and Laufr, 1971; Flanders and Radcliffe, 1992; Laferriere et al., 1999; Micheletto et al., 1999; Micheletto et al., 2000; Siri et al., 2009; National Plant Germplasm System, USDA, 2022). *Solanum malmeanum* stands out because of its frost tolerance and ability to cold acclimate at field conditions (i.e., increase cold tolerance after gradually increasing exposure to low temperatures), but also the capacity to be tolerant to heat (Vega and Bamberg, 1995; Genesys, 2022). Valuable quality traits for industrial purposes such as high dry matter, high protein and low reducing sugar content in tubers were also reported (Rocha et al., 2000; Chalá et al., 2001; Jansen et al., 2001) (Table 4).

3.6 *Solanum malmeanum* as a source of traits for the generation of new potato cultivars

3.6.1 Frost and heat tolerance

Solanum malmeanum is particularly interesting due to its freezing tolerance and great capacity to cold acclimate (i.e., increase cold tolerance after exposure to low, non-freezing temperatures) (Hawkes, 1958; Hudson, 1961; Ross and Rowe, 1965; Hawkes and Hjerting, 1969; National Plant Germplasm System, USDA, 2022; Genesys, 2022). Twenty-three accessions were reported to be frost resistant and one accession is also reported to be moderately resistant to heat stress (Table 3). It has been reported to be capable of surviving at extremely low temperatures (−0.55°C to −5°C), with none or a relatively small percentage (0–20%) of the leaf area damaged (Ross and Rowe, 1965; Hawkes and Hjerting, 1969; Vega and Bamberg, 1995). Earlier observations reported that wild potatoes that have a rosette habit, such as *S. malmeanum*, generally are more highly resistant to cold and frost (Firbas and Ross, 1962). Because most potato cultivars are sensitive to low temperatures and are unable to cold acclimate at temperatures lower than −3°C (Chen and Li, 1980), great efforts are carried out to find sources of tolerance and to introgress this trait into cultivated potato. A *Solanum malmeanum* freezing tolerant accession (MLM266-2) was fused with freezing susceptible dihaploid *S. tuberosum* (AC142) by somatic hybridization. Shoots were regenerated from calli formed, excised and cultured on MS medium (Tu et al., 2021). Somatic hybrids submitted to a correlation analysis between freezing tolerance and tuberization capacity of progenies generated from the first backcrossing with tetraploid cultivars *S. tuberosum* indicated that these traits are controlled by independent genetic loci (Tu et al., 2021).

TABLE 3 Resistance against abiotic and biotic stresses reported for the wild potato *Solanum malmeanum* (Solanaceae).

Trait	Agent	Accession code or identification	Source
Abiotic stress	Frost	PI 320269; PI 414154; PI 472840; PI 472841; PI 472841; PI 472843; PI 472844; PI 472845; PI 472846; PI 472847; PI 472848; PI 472849; PI 472851; PI 472852; PI 472853; PI 472854; PI 472855; PI 498408; PI 498416; PI 498417; PI 498418; PI 498419; MLM266-2	(Tu et al., 2021; Genesys, 2022)
	Heat	PI 498416	(Genesys., 2022)
Biotic stress	Bacterial wilt (*R. solanacearum)*	Accession code or identification not provided	(Siri et al., 2009)
	Ring rot (*C. michiganensis* subsp. *sepedonicus*)	PI 458318; PI 472841	(Genesys, 2022)
	Early blight (*A. solanii)*	PI 472840	(Genesys, 2022)
	Late blight (*P. infestans)*	OKA 7256; OKA 7256.01; OKA 7256.07; OKA 7256.08; OKA 7292; OKA 7310; OKA 7310.01	(Micheletto et al., 1999; Micheletto et al., 2000)
		PI 472844; PI 472846; PI 472847; PI 472850; PI 472853; PI 472855	(Genesys, 2022)
	Verticillium wilt (*V. dahliae)*	PI 472851	(Genesys, 2022)
	Green peach aphid (*M. persicae*)	PI 320269; PI 458318; PI 472840	(Radcliffe and Laufr, 1971; Genesys, 2022)
	Potato aphid (*M. euphorbiae*)	PI 320269	(Radcliffe and Laufr, 1971)
	Colorado potato beetle (*L. decemlineata*)	PI 458318	(Flanders and Radcliffe, 1992)
	Potato leafhopper (*E. fabae*)	PI 472851; PI 472843	(Flanders and Radcliffe, 1992)
	Potato cyst nematode (*G. rostochiensis)*	Accession code or identification not provided	(Castelli et al., 2003)
	Root-knot nematode (*M. hapla)*	PI 472841	(Genesys, 2022)
	Potato leaf roll virus (*Polerovirus* sp.)	PI 458318	(Genesys, 2022)

3.6.2 Biotic stresses

3.6.2.1 Bacterial

3.6.2.1.1 Bacterial wilt

In the range of geographical distribution of *Solanum malmeanum* in Southern South America, the predominant *R. solanacearum* belong to phylotype II sequevars (Siri et al, 2009; Santiago et al., 2017). *Solanum malmeanum* is documented to possess interesting levels of resistance against *R. solanacearum* phylotype II sequevars 1-2 previously isolated from different potato fields in Uruguay (Siri et al., 2009), evaluated under controlled conditions. Two accessions were classified to be resistant (e.g. absence of wilting symptoms) and three accessions demonstrate moderate resistance. Resistant accessions are summarized in Table 3. Little is known about the resistance across the range of distribution, however, there are ongoing studies aiming at characterizing resistance levels of potato wild species accessions across the distribution in Uruguay. Considering that each strain of bacterial wilt is present at specific geographical regions, the knowledge of its variability and diversity could be an efficient guide to identify new potential sources of resistance from different geographical populations.

TABLE 4 Quality traits found in potato wild relative *Solanum malmeanum* (Solanaceae) accessions.

Organ	Specificities	Accession	References
Tubers	Dry matter content 28.6% – 34.62%	CL55=BRA 00167071-0; CL57=BRA 00167093-4; CL60=BRA 00183755-8; CL63=BRA 00183756-6; CL65=BRA 00167137-9	(Rocha et al., 2000; Jansen, 2001; AleloVegetal, 2022)
	Tuber starch 15.4 – 23%	Accession code or identification not provided	(Jansen, 2001)
	Starch granules size 45.8 um	Accession code or identification not provided	(Jansen, 2001)
	Reducing sugar 0.19 –0.33%	CL63= BRA 00183756-6; CL65= BRA 00167137-9	(Chalá et al., 2001; Alelo Vegetal, 2022)
	Protein 4.5% of dry matter	Accession code or identification not provided	(Jansen et al., 2001)
	Amylose 32.1% of amylose content in starch	Accession code or identification not provided	(Jansen et al., 2001)
Leaves	Glycoalkaloid (dehydrocommersonine) 23%	PI 320269	(Gregory et al., (1981)

3.6.2.1.2 Bacterial ring rot

Two accessions (PI 458318 and PI 472841) proved to be resistant against bacterial ring rot of potato (Table 3) (Genesys, 2022; National Plant Germplasm System, USDA, 2022). Ring rot is caused by *Corynebacterium michiganensis* subsp. *sepedonicus* (Spieckermann & Kotthoff).

3.6.2.2 Fungal

3.6.2.2.1 Potato early blight

The genotype PI 472840 from Argentina was evaluated and has been reported to be resistant against potato early blight (Table 3) (Wolters et al., 2021; Genesys, 2022; National Plant Germplasm System, USDA, 2022). Genotype MLM 266-2 was evaluated under greenhouse conditions and showed no signs of infection, but small dark spots occurred due to the hypersensitive response (Wolters et al., 2021). MLM 266-2 was crossed with diploid *S. tuberosum*, producing a triploid progeny with high levels of resistance inherited dominantly (Wolters et al., 2021). The disease is caused by *Alternaria solani* (Ellis. & G. Martin) Ser., a leaf-spotting and defoliation agent that is responsible for significantly reduced yields in many tropical and subtropical producing regions (Rotem, 1994; Xue et al., 2022).

3.6.2.2.2 Potato late blight

In total thirteen accessions were reported to be resistant to late blight (Table 3). The genotypes OKA 7310.01, OKA 7256.08, OKA 7256.07, OKA 7256.01 showed an incompatible reaction after inoculation with R0 in trials under greenhouse conditions. A second experiment aimed to reconfirm the presence of *R*-genes and the offspring produced by the cross between OKA 7310.01 (resistant to R0) and OKA 7282.06 (susceptible to R0) was inoculated with a complex race of late blight. Among the offspring of 300 plants, 70% (210) showed incompatible (resistance) responses, confirming the presence of the *R* genes in *S. malmeanum* (Micheletto et al., 1999; Micheletto et al., 2000). Micheletto and collaborators (2000) screened four Argentinean diploid accessions for quantitative resistance to late blight *Phytophthora infestans* virulent (R1, R3, R4, R5, R7, R10, R11) and non-virulent (R0) complex races under both greenhouse and field conditions during two seasons (1996/1997 and 1997/98) to assess the year × genotype interaction. Results showed high variability reaction of *P. infestans* among genotypes. Two accessions were susceptible (OKA 7282, OKA 7291), while two resistant accessions segregated for resistance (Table 3). OKA 7292 genotypes 5 and 9 and OKA 7256 genotypes 2, 7 and 9 performed as resistant. Furthermore, no genotype × year interaction was detected, and the behavior was consistent across years. Although Micheletto and collaborators (1999; 2000) treat the evaluated accessions as *S. commersonii*, they are currently to be considered as *S. malmeanum* (Spooner et al., 2014; Spooner et al., 2016; Global Biodiversity Information Facility, 2022). The disease caused by the oomycete *Phytophthora infestans* (Mont.) de Bary is considered the most important disease of potato crop in all producing regions worldwide (Santana et al., 2013; Göre et al., 2019).

3.6.2.2.3 Verticillium wilt

The accession PI 472850 is reported to be immune against *Verticillium dahliae* (Table 3) (National Plant Germplasm System, USDA, 2022). Verticillium wilt is one of the most important soilborne fungal diseases on potato, also known as potato early dying disease, caused by *V. dahliae* Klebahn and *V. albo-atrum* Reinke & Berthold (Frost et al., 2007; Simko and Haynes, 2017; Li et al., 2019).

3.6.2.2.4 Fusarium dry rot

One accession (PI 414154) was screened in a tuber disk essay for resistance to *Fusarium sambucinum* (Lynch et al., 2003), however it responded as susceptible. Fusarium dry rot of potato tubers is caused by many species such as *F. avenaceum*, *F. culmorum*, *F. equiseti*, *F. oxysporum*, *F. sambucinum*, *F. solani*, among others. It is a severe disease responsible for causing significant economic losses mainly during the post-harvest and storage periods (Peters et al., 2008; Du et al., 2012; Gachango et al., 2012).

3.6.2.2.5 Nematodes

Solanum malmeanum showed full resistance to potato cyst nematode *Globodera rostochiensis* and susceptibility to *G. pallida*, but the inheritance of resistance was not investigated (Castelli et al., 2003). Furthermore, it has also been reported to be resistant against root-knot nematode (*Meloidogyne hapla*) (Table 3) (National Plant Germplasm System, USDA, 2022). Nematodes, when the population density is high, negatively affect potato production from the field, restricting plant growth, to the post-harvest period, rendering tubers unmarketable (Contina et al., 2019; Koirala et al., 2020).

3.6.2.2.6 Insect pests

Solanum malmeanum was moderately resistant against potato leafhopper (*Empoasca fabae* Harris), although the defense mechanisms have not been elucidated, as well as fully resistant to potato aphid *Macrosiphum euphorbiae* (Thomas) and potato green peach aphid *Myzus persicae* (Sulzer) (Table 3) (Radcliffe and Laufr, 1971; Genesys, 2022; National Plant Germplasm System, USDA, 2022). Insect resistance in potato is mainly associated to morphological characteristics, density on leaves and exudates of secondary metabolites of glandular trichomes, which provide protection against herbivory (Flanders et al., 1992). Also, an important natural source of insect resistance in potato is related to the production of high levels of foliar glycoalkaloids (Coombs et al., 2002). Some results suggest that this is an important mechanism of insect resistance found in *S. malmeanum*.

3.6.3 Viruses

Accession PI 458318 from Argentina has been reported to be resistant against potato leaf roll virus (*Polerovirus* sp.) (Table 3) (Genesys, 2022; National Plant Germplasm System, USDA, 2022). *Solanum malmeanum* has also been reported to be susceptible and spontaneously infected by the systematic ringspot virus (TMV tobacco mosaic virus) that is transmitted from generation to generation *via* tubers (Hansen, 1960).

3.6.4 Quality traits for production and industry

3.6.4.1 Tuber dry matter, starch and protein contents

Solanum malmeanum has been reported (Table 4) to contain a range from 28.6% to 34.62% of dry matter content (Rocha et al., 2000; Jansen et al., 2001), 15.4 to 23% starch content (Jansen et al., 2001), 4.5% of protein in tuber dry matter content and 32.1% of amylose content in starch, with a mean particle diameter of starch granules of

45.8 µm (CV%=9.2) (Jansen et al., 2001). In addition, reducing sugars on its tubers have been evaluated from two different growing seasons, with an average of 0.22% in spring, and 0.33% in autumn, far below the levels presented in the reference cultivar 'Baronesa' (*S. tuberosum*) (Chala et al., 2001). A strong association in diploid *S. malmeanum* clones was verified between glutamate oxaloacetate transaminase (GOT) at relative mobility band 1.00 of electrophoretic standards and high dry matter content (average of 34.62 of dry matter content on tubers). The use of the GOT enzyme may be efficient in more accurately predicting this character in the early stages of selection in diploid than tetraploid species (Andreu and Pereira, 2004).

3.6.4.2 Glycoalkaloids

Steroidal glycoalkaloids are secondary metabolites toxic for human consumption, although they may also have advantages to improve resistance against viral, insect-pests or microbial diseases. *Solanum* ser. *Commersoniana* species mainly produces α-solanine and β-chaconine (Schreiber, 1963). *Solanum malmeanum* accession PI 320269 (cited as *S. commersonii*) has been reported to contain 23% (mg/g dry leaves) of glycoalkaloid content on leaves (Table 4), where 100% is dehydrocommersonine, with undetectable levels (<3%) of α-solanine and β-chaconine (Gregory et al., 1981) (Table 4). Dehydrocommersonine was associated in *Solanum oplocense* Hawkes with host resistance against *Leptinotarsa decemlineata* (Tai et al., 2014; Tai et al., 2015), then *S. malmeanum* could be predicted as a potential source of resistance against this pest. There is room for selection because they are controlled by genetic factors (Mensinga et al., 2005) and even though high levels of foliar steroidal glycoalkaloids are frequently associated with high contents in tubers (Sanford et al., 1992), they are accumulated during the development of each organ (Krits et al., 2007; Mweetwa et al., 2012; Paudel et al., 2019; Zhen et al., 2019). This differential accumulation behavior allows to change the distribution of accumulation in specific plant tissues across some generations of selection (Sanford et al., 1992; Umemoto et al., 2016; Nahar et al., 2017; Zhen et al., 2019), so it should be possible to select accessions from *S. malmeanum* for breeding to obtain cultivars with resistance against pests given by high foliar glycoalkaloids content and with low toxicity for human consumption due to low levels of glycoalkaloids in the tubers.

3.7 Cytogenetics

Solanum malmeanum is a diploid species (2n = 2x = 12) with occasional triploid plants (Hawkes and Hjerting, 1969; Tarn and Hawkes, 1986; Spooner et al., 2016). Diploid plants usually produce higher frequency of stainable pollen than triploids. The mean pollen diameter in equatorial view is 23 µm in diploid plants and varies from 18.5 to 24.9 µm in triploids (Tarn and Hawkes, 1986).

Some accessions of *S. malmeanum* showed regular meiotic abnormalities observed at microsporogenesis, including chromosomes out of the equatorial plate in metaphases I and II (MI and MII) and lagging chromosomes in anaphases I and II, and telophases I and II (Pandolfi, 1998; Tomé et al., 2007; Tomé et al., 2009). Furthermore, a range from 3.7% to 50.8% of pollen mother cells (PMC) with meiosis abnormalities was observed. The frequency of abnormal PMC decreased in the later stages of meiosis. This, together with a good pollen viability (60.33 – 96.5%), suggests the existence of selection against abnormal PMC cells during the microsporogenesis process with implications for the choice of pollen donors for crosses.

Pollen viability was significantly higher in *S. malmeanum* than in cultivated potatoes (Pandolfi, 1998; Tomé et al., 2007). It also produces a high pollen amount, ranging from 454.20 to 476.20 pollen grains per field, exceeding the means of *S. tuberosum* clones (Tomé et al., 2007). According to Hardigan et al. (2017), the reduction of pollen viability of cultivated potato (*S. tuberosum*) is a consequence of the domestication syndrome. Although the accessions SCM57 (BGB017) and SCM60 (BGB447) (AleloVegetal, 2022) did not show unreduced pollen grains, they had meiotic mechanisms (parallel spindles) that possibly can lead to the formation of unreduced pollen, presenting the highest percentages of cells with parallel spindles (27% and 22%). Furthermore, pollen grain size of *S. commersonii* (13.68–35.52 µm) tends to be larger than that of *S. malmeanum* (11.4 to 23.9 µm) (Tomé et al., 2009). This approach put forward by Tomé et al., (2009) raises the hypothesis that *S. malmeanum* has no pollen grain diameters satisfying the value of 25 µm suggested by Quin et al. (1974) and Ramanna (1974) and would imply the proposition of a new threshold for unreduced (2n) pollen grains. However, the selection of 2n pollen based only on diameter and shape of the pollen grain is feasible only in species with known differences between haploid and diploid pollen grains (Parrott and Smith, 1984; Parrott and Smith, 1986). According to Ordoñez et al. (2017), 2n pollen have a diameter 1.2 times larger than reduced gametes (n). 2n gametes production can be predicted by crossing the species of interest with another different EBN group species, and then observing pollen tube growth, fruit set, seed set and germination success. Two of 12 clones of *S. malmeanum* (PI 414154, cited as "*S. commersonii*") were reported to produce 2n pollen (den Nijs and Peloquin, 1977a). Therefore, it is expected that in natural populations, functional 2n gametes can be found.

3.8 Reproductive biology, breeding system and endosperm balance number

Solanum malmeanum is a tertiary gene pool (1EBN) wild relative of potato and is considered reproductively isolated from other tuber-bearing 2EBN and 4EBN species (Hanneman, 1994; Spooner et al., 2016). *Solanum malmeanum* reproduces vegetatively by tubers and stolons and sexually by botanical seeds. From five diploid accessions, four were characterized as self-incompatible (Nicolao, 2021) and one self-compatible accession of *S. malmeanum* was found (Nicolao, 2021).

Hanneman (1994) used a diploid female plant of *S. malmeanum* for the assignment of EBN in crossing tests. Although only one accession was used as female to determine its EBN, it produced an average of six seeds per fruit from any crosses with male standard species 2x (1EBN), less than one seed per fruit from 2x (2EBN) species, and no seed set from 4x (4EBN) crosses. Therefore, it was designated as 1EBN.

Crosses between different potato species with the same EBN number, regardless of ploidy, are usually successful (Johnston et al., 1980; Johnston and Hanneman, 1982). Theoretically, species with the same EBN may be

intercrossed in nature if sympatry and synchronized flowering occur and pre-zygotic barriers are absent. Tarn and Hawkes (1986) obtained vigorous and fertile hybrid progenies from reciprocal crosses between 2x *S. malmeanum* and 2x *S. commersonii*. The seed production per berry and mean weight per 100 seeds were equally variable. *Solanum commersonii* produces fewer (66.5), but heavier seeds (52.8 mg per 100 seeds) per berry than *S. malmeanum* (128.4 seeds per berry, with 12.8 mg per 100 seeds) (Tarn and Hawkes, 1986). Furthermore, the seedlings grown from reciprocal crosses between *S. commersonii* and *S. malmeanum* usually were vigorous, fertile, highly variable morphologically and generally displayed intermediate morphology between the two parental species (Tarn and Hawkes, 1986). It is interesting to note that parental traits as flower pigmentation (typically purple-pigmented corolla in *S. commersonii* and white (not-pigmented) corolla in *S. malmeanum*), always segregate in F1 progenies with intermediate phenotypes (Tarn and Hawkes, 1986).

Hawkes and Hjerting (1969) made significant efforts to perform reciprocal crosses with *S. chacoense* and obtained only one seed that germinated and generated a vigorous plant. Ehlenfeldt and Hanneman (1988) produced a small number of seeds from many hand pollinations between *S. malmeanum* (accession PI 320269, cited as "*S. commersonii*") (1EBN) and *S. chacoense* (2EBN) (Ehlenfeldt and Hanneman, 1988). Obtaining hybrid offspring of *S. malmeanum* × *S. chacoense* was possible, but the endosperm barrier between these two species greatly limits viable seed development. In reciprocal backcrosses of the F1 hybrids (1 ½ EBN) with both *S. chacoense* (2EBN) and *S. malmeanum* (1EBN), the results were consistent with a directional effect determining the difference in the ratio of viable seeds to aborted seeds, related to which parent is male and which parent is female in any given cross, but it does not seem to be a cytoplasmic effect, because the same female may produce different results depending on pollen source. This effect is most likely due to the dose-effect in the endosperm (Ehlenfeldt and Hanneman, 1988).

Crossability barriers between 2x 2EBN *S. chacoense* and 2x 1EBN *S. malmeanum* are well-established, but no sexual barriers were observed between 2x 1EBN *S. commersonii* and 2x 1EBN *S. malmeanum*, elucidating that EBN is a strong internal post-zygotic barrier (Johnston et al., 1980; Summers and Grun, 1981; Hanneman, 1994). Thus, it may predict that *S. malmeanum* could freely cross with other 1EBN sympatric wild potatoes such as 2x 1EBN *S. commersonii*, but not with 2EBN such as *S. chacoense*, except when 2n gametes are present (Hawkes and Hjerting, 1969; Hanneman, 1994). Thus, the mechanisms involved in keeping the two species appart in sympatric zones where *S. commersonii* and *S. malmenum* occur still deserve further explanations.

4 Discussion

Solanum malmeanum is a valuable genetic resource for many traits of importance to the potato industry. This 1EBN wild potato contains genes to combat pests and diseases which are lacking in the primary genepool of potato. However, using this 1EBN wild species to improve cultivated potatoes is challenging. *S. malmeanum* is cross-incompatible with cultivated potatoes due to the strong EBN post-zygotic barrier (Hanneman, 1994). To capture genetic diversity from 1EBN potato species into the cultivated gene pool, breeders have efficiently used the 2x 2EBN Mexican potato species *Solanum verrucosum* as bridge species (Jansky and Hamernik, 2009). Bamberg and collaborators (2021) show the ability of female parent *S. verrucosum* (2EBN) to form interspecific hybrids with 1EBN species as male parent without stylar barrier. Much like *S. verrucosum*, the lack of stylar barriers in cultivated tomato has allowed tomato breeders to access genetically distant wild relatives, even those of different effective ploidy (Fulton et al., 1997; Roth et al., 2019; Städler et al., 2021). Thus, a combination of strategies such as ploidy manipulation, somatic fusion, embryo rescue and transgenesis are expected not to be required to access the valuable 1EBN potato wild species genetic diversity.

Ploidy manipulation methods were established to introgress genetic diversity from diploids to tetraploids (Hermundstad and Peloquin, 1985a; Peloquin et al., 1989a; Jansky et al., 1990; Werner and Peloquin, 1991a; Carputo et al., 2003). Breeders usually use di-haploids (2x 2EBN) derived from crossing tetraploid *S. tuberosum* (Group Tuberosum or Group Phureja) with haploid inducers to enable intercrossing with wild diploid species (Hermundstad and Peloquin, 1985a; Peloquin et al., 1989a; Carputo et al., 2003), or anti-mitotic substances, such as colchicine and oryzalin, to increase the ploidy level of diploid species (Tomé et al., 2016). Diploid F1 offspring that produces 2n gametes are valuable genetic material for capturing allelic diversity from most of the diploid wild potatoes (Hermundstad and Peloquin, 1985a; Peloquin et al., 1989a; Jansky et al., 1990; Jansky and Peloquin, 2006). After many cycles of selection for adaptation and desired agronomic traits, superior diploid hybrids can be crossed with meiotic mutants producing 2n gametes to generate 4x progenies (Peloquin et al., 1989a; Peloquin et al., 1989b; Werner and Peloquin, 1991a; Masuelli et al., 1992; Carputo et al., 2003). Nodal segments of diploid *S. malmeanum* when treated with oryzalin (10-50μM for 24h) successfully induced tetraploid plants (Tomé et al., 2016).

Recently, diploid F1-hybrid breeding has increased attention of breeders due to the possibility of redomestication of potato crop by creating inbred lines from clones carrying mutations of the *Sli* self-incompatibility gene locus (Lindhout et al., 2011; Jansky et al., 2016; Eggers et al., 2021). Diploid inbred lines are outstanding materials for breeding and functional genetics (Marand et al., 2019; Jayakody et al., 2022), because they allow high throughput phenotyping and efficient QTL detection (Manrique-Carpintero et al., 2015; Hara-Skrzypiec et al., 2018), accelerating the breeding process. An obstacle still present in diploid potato breeding is the high level of inbreeding depression (De Jong and Rowe, 1971; Zhang et al., 2019).

The taxonomic history of *S. malmeanum* can hamper the exploration of its true potential for potato breeding. *Solanum malmeanum* has been genetically understudied compared to *S. chacoense* and *S. commersonii* that have their genomes sequenced (Aversano et al., 2015; Leisner et al., 2018). The utilization of *S. chacoense* is well established in diploid potato breeding programs (Lindhout et al., 2011; Jansky et al., 2014). *Solanum commersonii* is highly exploited mainly for its frost tolerance (Cardi et al., 1993; Esposito et al., 2020; Esposito et al., 2021) and bacterial wilt (*Ralstonia solanacearum*) resistance (Kim-Lee et al., 2005; Siri et al., 2009; Andino et al., 2022).

The morphological variability, wide geographic distribution, diverse array of habitats occupied and ecological preferences of *S. malmeanum* are not fully represented in current genebanks. Limited studies were performed to assess the genetic variability and the few

available ones do not fully cover these aspects. Important agronomical traits were documented, although comprehensive studies on this species are lacking and most of the references available for response to abiotic and biotic stresses (Flanders and Radcliffe, 1992; Micheletto et al., 1999; Micheletto et al., 2000; Castelli et al., 2003) and quality traits for production and industry (Rocha et al., 2000; Chalá et al., 2001; Jansen, 2001) are from occasional works or have evaluated *S. malmeanum* as "*S. commersonii*".

Next Generation Sequencing (NGS) technologies will allow a better understanding of the evolutionary relationships between *S. malmeanum* and *S. commersonii* to provide the required framework for conserving and using this potato wild genetic diversity (Fielder et al., 2015; Dempewolf et al., 2017). The reinstatement at the species level brings new light to a comeback and promising future for the applied uses of this wild potato in classic and cutting-edge techniques for potato breeding. Towards the true exploitation of this promising potato genetic resource, the current taxonomy resulting from the latest treatment must be updated in all genebanks and reference collections to allow better use of the conserved germplasm and enhance its applied uses. Furthermore, with the rise of new technologies, new studies are needed to evaluate the entire geographical distribution of *S. malmeanum* and compare to the other two sympatric potato wild species *S. chacoense* and *S. commersonii*.

Plant breeders require genetic variability information to speed up the search of each desirable agronomic trait, tolerance against abiotic stresses, as well as for the genes encoding for quality traits (Bethke et al., 2017; Swarup et al., 2021).

As a final remark, we highlight that the potential of *S. malmeanum* for potato breeding has been neglected, especially when compared to other close potato wild relatives such as *S. chacoense* and *S. commersonii* (Aversano et al., 2015; Leisner et al., 2018) and must be properly addressed with contemporary molecular tools to unlock its applied use as a promising potato wild relative to face the challenges imposed to the potato crop in the 21st century.

Author contributions

RN discussed the original idea, reviewed and checked literature data, compiled and wrote the manuscript. PG wrote partially, provided data on the Uruguayan references, and reviewed the text. CC provided advisory, wrote partially and reviewed the text. GH discussed the original idea, provided advisory, wrote partially and reviewed the text. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by CAPES/PROAP, Embrapa, FAPERGS (2551-0001703-0, 2019) and Conselho Nacional de Pesquisa (CNPq 429368/2016-0). RN acknowledges The International Association for Plant Taxonomy for the IAPT 2022 Research Grants Program. GH acknowledges Conselho Nacional de Pesquisa (CNPq 314590/2020-0) for the productivity research grant fellowship.

Acknowledgments

The authors acknowledge Antônio Costa Oliveira (UFPEL), Arione da Silva Pereira (Embrapa) and João Iganci (UFPEL) for their contributions on the earlier versions of the manuscript, Ariana Digilio (INTA, Argentina) for the information provided on genebank accession, and the referees Leonardo Lopes Bhering (UFV) and Leandro Giacomin (UFPB) for their contributions towards the improvement of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

AleloVegetal (2022). Embrapa Genetic Resources and Biotechnology. Available at: https://alelobag.cenargen.embrapa.br/AleloConsultas/Passaporte/buscaNcDetalhes.do?idn=6672&idb=475https://alelo.cenargen.embrapa.br/ (Accessed 31 March 2022).

Andino, M., Gaiero, P., Gonzalez-Barrios, P., Galvan, G., Vilaro, F., and Speranza, P. (2022). Potato introgressive hybridisation breeding for bacterial wilt resistance using *Solanum commersonii* Dun. as donor: Genetic and agronomic characterisation of a backcross 3 progeny. *Potato Res.* 65 (1), 119–136. doi: 10.1007/s11540-021-09512-1

Andreu, M. A., and Pereira, A. S. (2004). Estimação da qualidade industrial da batata (*Solanum tuberosum* L.) através do uso da isoenzima glutamato oxaloacetato transaminase. *RIA. Rev. Investigaciones Agropecuarias.* 33, 5–13.

Aversano, R., Contaldi, F., Ercolano, M. R., Grosso, V., Iorizzo, M., Tatino, F., et al. (2015). The *Solanum commersonii* genome sequence provides insights into adaptation to stress conditions and genome evolution of wild potato relatives. *Plant Cell.* 27 (4), 954–968. doi: 10.1105/tpc.114.135954

Bachmann-Pfabe, S., and Dehmer, K. J. (2020). Evaluation of wild potato germplasm for tuber starch content and nitrogen utilization efficiency. *Plants.* 9 (7), 833. doi: 10.3390/plants9070833

Bamberg, J., Kielar, A., del Rio, A., and Douches, D. (2021). Making hybrids with the wild potato *Solanum jamesii*. *Am. J. Potato Res.* 98 (3), 87–193. doi: 10.1007/s12230-021-09828-1

Base de Dados da Pesquisa Agropecuária (2022). Available at https://www.bdpa.cnptia.embrapa.br/consulta/ (Accessed 31 March 2022).

Bethke, P. C., Halterman, D. A., and Jansky, S. (2017). Are we getting better at using wild potato species in light of new tools? *Crop Science.* 57 (3), 1241–1258. doi: 10.2135/cropsci2016.10.0889

Biblioteca Digital Brasileira de Teses e Dissertações (2022). Available at: http://bdtd.ibict.br/vufind/ (Accessed 31 March 2022).

Birch, P. R., Bryan, G., Fenton, B., Gilroy, E. M., Hein, I., Jones, J. T., et al. (2012). Crops that feed the world 8: Potato: are the trends of increased global production sustainable? *Food Security.* 4 (4), 477–508. doi: 10.1007/s12571-012-0220-1

Bitter, G. (1913). "*Solanum (Tuberarium) malmeanum* Bitter, nov. spec," in *Repertorium specierum novarum regni vegetabilis.*, vol. 12. , 447–448.

Camadro, E. L., Carputo, D., and Peloquin, S. J. (2004). Substitutes for genome differentiation in tuber-bearing *Solanum*: interspecific pollen-pistil incompatibility, nuclear-cytoplasmic male sterility, and endosperm. *Theor. Appl. Genet.* 109, 1369–1376. doi: 10.1007/s00122-004-1753-2

Camadro, E. L., Erazzu, L. E., Maune, J. F., and Bedogni, M. C. (2012). A genetic approach to the species problem in wild potato. *Plant Biol.* 14 (4), 543–554. doi: 10.1111/j.1438-8677.2012.00563.x

Camadro, E. L., and Masuelli, R. W. (1995). A genetic model for the endosperm balance number (EBN) in the wild potato *Solanum acaule* Bitt. and two related diploid species. *Sexual Plant Reproduction.* 8 (5), 283–288. doi: 10.1007/BF00229385

Camadro, E. L., and Peloquin, S. J. (1981). Cross-incompatibility between two sympatric polyploid *Solanum* species. *Theor. Appl. Genet.* 60, 65–70. doi: 10.1007/BF00282417

Cardi, T., D'Ambrosio, E., Consoli, D., Puite, K. J., and Ramulu, K. S. (1993). Production of somatic hybrids between frost-tolerant *Solanum commersonii* and *S. tuberosum*: characterization of hybrid plants. *Theor. Appl. Genet.* 87 (1), 193–200. doi: 10.1007/BF00223764

Carputo, D., Frusciante, L., and Peloquin, S. J. (2003). The role of 2n gametes and endosperm balance number in the origin and evolution of polyploids in the tuber-bearing Solanums. *Genetics.* 163 (1), 287–294. doi: 10.1093/genetics/163.1.287

Castelli, L., Ramsay, G., Bryan, G., Neilson, S. J., and Phillips, M. S. (2003). New sources of resistance to the potato cyst nematodes *Globodera pallida* and *G. rostochiensis* in the Commonwealth Potato Collection. *Euphytica.* 129(3), 377–386. doi: 10.1023/A:1022264111096.

Chala, C. D. A., Pereira, A. D. S., Campos, A. D., Viegas, J., and Salamoni, A. T. (2001). Genetic variability for reducing sugar content in wild potatoes that ocurr in southern Brazil. *Ciênc. Rural.* 31, 43–47. doi: 10.1590/S0103-84782001000100007

Chen, H. H., and Li, P. H. (1980). Characteristics of cold acclimation and de-acclimation of tuber-bearing *Solanum* species. *Plant Physiol.* 65, 1146–1148. doi: 10.1104/pp.65.6.1146

Cipar, M. S., Peloquin, S. J., and Hougas, R. W. (1964). Variability in the expression of self-incompatibility in tuber-bearing diploid *Solanum* species. *American Potato Journal* 41 (6), 155–162. doi: 10.1007/BF02855317

Contina, J. B., Dandurand, L. M., and Knudsen, G. R. (2019). A predictive risk model analysis of the potato cyst nematode *Globodera pallida* in Idaho. *Plant Dis.* 103 (12), 3117–3128. doi: 10.1094/PDIS-04-19-0717-RE

Coombs, J. J., Douches, D. S., Li, W., Grafius, E. J., and Pett, W. L. (2002). Combining engineered (*Bt-cry3A*) and natural resistance mechanisms in potato for control of colorado potato beetle. *J. Am. Soc. Hortic. Sci.* 127 (1), 62–68. doi: 10.21273/JASHS.127.1.62

Correll, D. (1962). *The potato and its wild relatives.* (Renner, USA: Texas Res Found).

De Egea, J. I.M., Céspedes Domínguez, G. R., Peña-Chocarro, M. D.C., Mereles, M. F., and Rolón Mendoza, C. (2018). Recursos fitogenéticos del Paraguay: Sinopsis, Atlas y Estado de Conservación de los Parientes Silvestres de Especies de Importancia para la Alimentación y la Agricultura. *Parte I.* n4, 197–198.

De Jong, H., and Rowe, P. (1971). Inbreeding in cultivated diploid potatoes. *Potato Res.* 14, 74–83. doi: 10.1007/BF02355931

Dempewolf, H., Baute, G., Anderson, J., Kilian, B., Smith, C., and Guarino, L. (2017). Past and future use of wild relatives in crop breeding. *Crop Science.* 57 (3), 1070–1082. doi: 10.2135/cropsci2016.10.0885

Dempewolf, H., Eastwood, R. J., Guarino, L., Khoury, K. C., Müller, J. V., and Toll, J. (2014). Adapting agriculture to climate change: a global initiative to collect, conserve, and use crop wild relatives. *Agroecology Sustain. Food Systems.* 38 (4), 369–377. doi: 10.1080/21683565.2013.870629

den Nijs, T. P. M., and Peloquin, S. J. (1977a). 2n gametes in potato species and their function in sexual polyploidization. *Euphytica.* 26, 585–600. doi: 10.1007/BF00021684

Devaux, A., Kromann, P., and Ortiz, O. (2014). Potatoes for sustainable global food security. *Potato Res.* 57, 185–199. doi: 10.1007/s11540-014-9265-1

Douches, D., Schroeter, K., Ludlam, K., and Hicks, K. (1989). Allelic diversity among the *Solanum* species, section *Petota*. *Am. Potato J.* 66, 517. doi: 10.1007/BF02855442

Drapal, M., Lindqvist-Kreuze, H., Mihovilovich, E., Aponte, M., Bonierbale, M., and Fraser, P. D. (2020). Cooking dependent loss of metabolites in potato breeding lines and their wild and landrace relatives. *J. Food Composition Analysis.* 88, 103432. doi: 10.1016/j.jfca.2020.103432

Duan, Y., Liu, J., Xu, J., Bian, C., Duan, S., Pang, W., et al. (2019). DNA Fingerprinting and genetic diversity analysis with simple sequence repeat markers of 217 potato cultivars (*Solanum tuberosum* L.) in China. *Am. J. Potato Res.* 96 (1), 21–32. doi: 10.1007/s12230-018-9685-6

Du, M., Ren, X., Sun, Q., Wang, Y., and Zhang, R. (2012). Characterization of *Fusarium* spp. causing potato dry rot in China and susceptibility evaluation of Chinese potato germplasm to the pathogen. *Potato Res.* 55 (2), 175–184. doi: 10.1007/s11540-012-9217-6

Eggers, E. J., van der Burgt, A., van Heusden, S. A., de Vries, M. E., Visser, R. G., Bachem, C. W., et al. (2021). Neofunctionalisation of the *Sli* gene leads to self-compatibility and facilitates precision breeding in potato. *Nat. Commun.* 12 (1), 1–9. doi: 10.1038/s41467-021-24267-6

Ehlenfeldt, M. K., and Hanneman, R. E. (1988). Genetic control of endosperm balance number (EBN): three additive loci in a threshold-like system. *Theor. Appl. Genet.* 75 (6), 825–832. doi: 10.1007/BF00258041

Esposito, S., Aversano, R., Bradeen, J., D'Amelia, V., Villano, C., and Carputo, D. (2021). Coexpression gene network analysis of cold-tolerant *Solanum commersonii* reveals new insights in response to low temperatures. *Crop Science.* 61 (5), 3538–3550. doi: 10.1002/csc2.20473

Esposito, S., Aversano, R., Bradeen, J. M., Di Matteo, A., Villano, C., and Carputo, D. (2020). Deep-sequencing of *Solanum commersonii* small RNA libraries reveals riboregulators involved in cold stress response. *Plant Biol.* 22, 133–142. doi: 10.1111/plb.12955

Fielder, H., Brotherton, P., Hosking, J., Hopkins, J. J., Ford-Lloyd, B., and Maxted, N. (2015). Enhancing the conservation of crop wild relatives in England. *PloS One.* 10 (6), e0130804. doi: 10.1371/journal.pone.0130804

Firbas, H., and Ross, H. (1962). Züchtung auf frostresistenz bei der kartoffel. II: Über die frostresistenz der knolle und ihre beziehung zur frostresistenz des laubes. z. *Pflanzenzüchtung.* 47, 51–66.

Flanders, K. L., Hawkes, J. G., Radcliffe, E. B., and Lauer, F. I. (1992). Insect resistance in potatoes: sources, evolutionary relationships, morphological and chemical defenses, and ecogeographical associations. *Euphytica.* 61 (2), 83–111. doi: 10.1007/BF00026800

Flanders, K. L., and Radcliffe, E. B. (1992). Host plant resistance in *Solanum* germplasm. an appraisal of resistance to Colorado potato beetle, potato leafhopper and potato flea beetle. St. Paul, Minnesota. *Minnesota Agricultural Experiment Station Bulletin.* (599), 1–22.

Frost, K. E., Rouse, D. I., and Jansky, S. H. (2007). Considerations for *Verticillium* wilt resistance evaluation in potato. *Plant Disease.* 91 (4), 360–367. doi: 10.1094/PDIS-91-4-0360

Fulton, T. M., Beck-Bunn, T., Emmatty, D., Eshed, Y., Lopez, J., Petiard, V., et al. (1997). QTL analysis of an advanced backcross of *Lycopersicon peruvianum* to the cultivated tomato and comparisons with QTLs found in other wild species. *Theor. Appl. Genet.* 95 (5), 881–894. doi: 10.1007/s001220050639

Fu, Y. B., Peterson, G. W., Richards, K. W., Tarn, T. R., and Percy, J. E. (2009). Genetic diversity of Canadian and exotic potato germplasm revealed by simple sequence repeat markers. *Am. J. Potato Res.* 86 (1), 38–48. doi: 10.1007/s12230-008-9059-6

Gachango, E., Hanson, L. E., Rojas, A., Hao, J. J., and Kirk, W. W. (2012). *Fusarium* spp. causing dry rot of seed potato tubers in Michigan and their sensitivity to fungicides. *Plant Disease.* 96 (12), 1767–1774. doi: 10.1094/PDIS-11-11-0932-RE

Genesys (2022). Global portal on plant genetic resources. Available at: https://www.genesys-pgr.org (Accessed March 31, 2022).

GEOlocate (2022). Available at: https://www.geolocation.com/ (Accessed 31 March 2022).

Germplasm Resources Information Network - GRIN, Taxonomy (2022)USDA, Agricultural research service. Available at: https://doi.org/10.15482/USDA.ADC/1212393 (Accessed March 31, 2022).

Global Biodiversity Information Facility - GBIF (2022). Available at: https://www.gbif.org/species/2931000https://doi.org/10.15468/dl.zvvgga (Accessed June 15, 2022).

Global Biodiversity Information Facility (2022). Available at: https://www.gbif.org/ (Accessed 31 March 2022).

Göre, M. E., Altin, N., Yaman, T., Myers, K., Çagli, A., Cooke, D. E. L., et al. (2019). Severe outbreaks of *Phytophthora infestans* on potato in Turkey caused by recent changes in the pathogen population structure. *Phytoparasitica.* 47 (5), 693–709. doi: 10.1007/s12600-019-00768-5

Google Earth (2022). Available at: http://earth.google.com/ (Accessed 31 March 2022).

Google Maps (2022). Available at: https://www.google.com.br/maps (Accessed 31 March 2022).

Google Scholar (2022). Available at: https://scholar.google.com/ (Accessed 31 March 2022).

Gregory, P., Sinden, S. L., Osman, S. F., Tingey, W. M., and Chessin, D. A. (1981). Glycoalkaloids of wild, tuber-bearing *Solanum* species. *Journal of Agricultural and Food Chemistry.* 29(6), 1212–1215.

Gutarra, L., Herrera, J., Fernandez, E., Kreuze, J., and Lindqvist-Kreuze, H. (2017). Diversity, pathogenicity, and current occurrence of bacterial wilt bacterium *Ralstonia solanacearum* in Peru. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01221

Hanneman, R. E.Jr. (1985). Self fertility in *Solanum chacoense*. *Am. Potato J.* 62, 428–429.

Hanneman, R. E.Jr. (1989). The potato germplasm resource. *Am. Potato J.* 66, 655–667. doi: 10.1007/BF02853985

Hanneman, R. E.Jr. (1994). Assignment of endosperm balance number to the tuber-bearing solanums and their close non-tuber-bearing relatives. *Euphytica.* 74, 19–25. doi: 10.1007/BF00033762

Hansen, H. P. (1960). Tobacco mosaic virus carried in potato tubers. *American Potato Journal.* 37, 95–101. doi: 10.1007/BF02855138

Hara-Skrzypiec, A., Śliwka, J., Jakuczun, H., and Zimnoch-Guzowska, E. (2018). QTL for tuber morphology traits in diploid potato. *J. Appl. Genet.* 59 (2), 123–132. doi: 10.1007/s13353-018-0433-x

Hardigan, M. A., Bamberg, J., Buell, C. R., and Douches, D. S. (2015). Taxonomy and genetic differentiation among wild and cultivated germplasm of *Solanum* sect. *Petota*. *Plant Genome.* 8 (1), plantgenome2014–06. doi: 10.3835/plantgenome2014.06.0025

Hardigan, M. A., Laimbeer, F. P. E., Newron, L., Crisovan, E., Hamilton, J. P., Vaillancourt, B., et al. (2017). Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication in the cultivated potato. *Proc. Natl. Acad. Sci. United States America.* 114 (46), 9999–10008. doi: 10.1073/pnas.1714380114

Hawkes, J. G. (1958). Significance of wild species and primitive forms for potato breeding. *Euphytica.* 7 (3), 257–270. doi: 10.1007/BF00025267

Hawkes, J. G. (1990). *The potato: evolution, biodiversity and genetic resources.* (London: Belhaven Press).

Hawkes, J. G., and Hjerting, J. P. (1969). "The potatoes of Argentina, Brazil, Paraguay and Uruguay," in *A Biosystematic Study.* (London: Oxford University Press).

Hermundstad, S. A., and Peloquin, S. J. (1985). Germplasm enhancement with potato haploids. *J. Heredity.* 76, 463–467. doi: 10.1093/oxfordjournals.jhered.a110146

Hijmans, R. J., and Spooner, D. M. (2001). Geographic distribution of wild potato species. *Am. J. Botany.* 88, 2101–2112. doi: 10.2307/3558435

Hosaka, K., and Hanneman, R. E. (1998a). Genetics of self-compatibility in a self-incompatible wild diploid potato species *Solanum chacoense*. 1. detection of an S locus inhibitor (*Sli*) gene. *Euphytica.* 99 (3), 191–197. doi: 10.1023/A:1018353613431

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

Hudson, M. A. (1961). The limitations of a cut leaf test for assessing the frost resistance of the tuber-bearing solanums. *Euphytica.* 10 (2), 169–179. doi: 10.1007/BF00022208

Jacobs, M. M., van den Berg, R. G., Vleeshouwers, V. G., Visser, M., Mank, R., Sengers, M., et al. (2008). AFLP analysis reveals a lack of phylogenetic structure within *Solanum* section *Petota*. *BMC Evolutionary Biol.* 8 (145), 1–12. doi: 10.1186/1471-2148-8-145

Jansen, G., Flamme, W., Schüler, K., and Vandrey, M. (2001). Tuber and starch quality of wild and cultivated potato species and cultivars. *Potato Res.* 44, 137–146. doi: 10.1007/BF02410100

Jansky, S. (2000). Breeding for disease resistance in potato. *Plant Breed. Rev.* 19, 69–156. doi: 10.1002/9780470650172.ch4

Jansky, S. H., Charkowski, A. O., Douches, D. S., Gusmini, G., Richael, C., Bethke, P. C., et al. (2016). Reinventing potato as a diploid inbred line-based crop. *Crop Science.* 56, 1412–1422. doi: 10.2135/cropsci2015.12.0740

Jansky, S. H., Chung, Y. S., and Kittipadukal, P. I. Y. A. (2014). M6: a diploid potato inbred line for use in breeding and genetics research. *Journal of Plant Registrations.* 8(2), 195–199. doi: 10.3198/jpr2013.05.0024crg.

Jansky, S. H., Dempewolf, H., Camadro, E. L., Simon, R., Zimnoch-Guzowska, E., Bisognin, D. A., et al. (2013). A case for crop wild relative preservation and use in potato. *Crop Science.* 53 (3), 746–754. doi: 10.2135/cropsci2012.11.0627

Jansky, S., and Hamernik, A. (2009). The introgression of 2× 1EBN *Solanum* species into the cultivated potato using *Solanum verrucosum* as a bridge. *Genet. Resour. Crop Evolution.* 56 (8), 1107. doi: 10.1007/s10722-009-9433-3

Jansky, S. H., and Peloquin, S. J. (2006). Advantages of wild diploid *Solanum* species over cultivated diploid relatives in potato breeding programs. *Genet. Resour. Crop Evolution.* 53 (4), 669–674. doi: 10.1007/s10722-004-2949-7

Jansky, S. H., Yerk, G. L., and Peloquin, S. J. (1990). The use of potato haploids to put 2x wild species germplasm into a usable form. *Plant Breeding.* 104, 290–294. doi: 10.1111/j.1439-0523.1990.tb00438.x

Jayakody, T. B., Enciso-Rodríguez, F. E., Jensen, J., Douches, D. S., and Nadakuduti, S. S. (2022). Evaluation of diploid potato germplasm for applications of genome editing and genetic engineering. *Am. J. Potato Res.* 99 (1), 13–24. doi: 10.1007/s12230-021-09855-y

Johnston, S. A., den Nijs, T. P., Peloquin, S. J., and Hanneman, R. E. J. (1980). The significance of genic balance to endosperm development in interspecific crosses. *Theor. Appl. Genet.* 57, 5–9. doi: 10.1007/BF00276002

Johnston, S. A., and Hanneman, R. E. (1982). Manipulations of endosperm balance number overcome crossing barriers between diploid *Solanum* species. *Science.* 217: 4558, 446–448. doi: 10.1126/science.217.4558.446

Johnston, S., and Hanneman, R. (1996). Genetic control of endosperm balance number (EBN) in the Solanaceae based on trisomic and mutation analysis. *Genome.* 39 (2), 314–321. doi: 10.1139/g96-042

JSTOR (2022). Available at: https://www.jstor.org/ (accessed 31 March 2022).

Kim-Lee, H., Moon, J. S., Hong, Y. J., Kim, M. S., and Cho, H. M. (2005). Bacterial wilt resistance in the progenies of the fusion hybrids between haploid of potato and *Solanum commersonii*. *Am. J. Potato Res.* 82 (2), 129–137. doi: 10.1007/BF02853650

Koirala, S., Watson, P., McIntosh, C. S., and Dandurand, L. M. (2020). Economic impact of *Globodera pallida* on the Idaho economy. *Am. J. Potato Res.* 97, 214–220. doi: 10.1007/s12230-020-09768-2

Kolech, S. A., Halseth, D., Perry, K., Wolfe, D., Douches, D. S., Coombs, J., et al. (2016). Genetic diversity and relationship of Ethiopian potato varieties to germplasm from North America, Europe and the International Potato Center. *Am. J. Potato Res.* 93 (6), 609–619. doi: 10.1007/s12230-016-9543-3

Krits, P., Fogelman, E., and Ginzberg, I. (2007). Potato steroidal glycoalkaloid levels and the expression of key isoprenoid metabolic genes. *Planta.* 227 (1), 143–150. doi: 10.1007/s00425-007-0602-3

Laferriere, L. T., Helgeson, J. P., and Allen, C. (1999). Fertile *Solanum tuberosum* + *S. commersonii* somatic hybrids as sources of resistance to bacterial wilt caused by *Ralstonia solanacearum*. *Theor. Appl. Genet.* 98 (8), 1272–1278. doi: 10.1007/s001220051193

Leisner, C. P., Hamilton, J. P., Crisovan, E., Manrique-Carpintero, N. C., Marand, A. P., Newton, L., et al. (2018). Genome sequence of M6, a diploid inbred clone of the high-glycoalkaloid-producing tuber-bearing potato species *Solanum chacoense*, reveals residual heterozygosity. *Plant J.* 94 (3), 562–570. doi: 10.1111/tpj.13857

Lester, R. N. (1965). Immunological studies on the tuber-bearing solanums: 1. techniques and South American species. *Ann. Botany.* 29 (4), 609–624. doi: 10.1093/oxfordjournals.aob.a083975

Lindhout, P., Meijer, D., Schotte, T., Hutten, R. C. B., Visser, R. G. F., and Van Eck, H. J. (2011). Towards F1 hybrid seed potato breeding. *Potato Res.* 54, 301–312. doi: 10.1007/s11540-011-9196-z

Li, H., Wang, Z., Hu, X., Shang, W., Shen, R., Guo, C., et al. (2019). Assessment of resistance in potato cultivars to verticillium wilt caused by *Verticillium dahliae* and *Verticillium nonalfalfae*. *Plant Dis.* 103 (6), 1357–1362. doi: 10.1094/PDIS-10-18-1815-RE

Lynch, D. R., Kawchuk, L. M., Chen, Q., and Kokko, M. (2003). Resistance to *Fusarium sambucinum* in wild and cultivated *Solanum* species. *Am. J. Potato Res.* 80, 353–358. doi: 10.1007/BF02854246

Manetti, M. E., Rossi, M., Costa, A. P., Clausen, A. M., and Van Sluys, M. A. (2007). Radiation of the Tnt1 retrotransposon superfamily in three Solanaceae genera. *BMC Evolutionary Biol.* 7 (1), 34. doi: 10.1186/1471-2148-7-34

Manrique-Carpintero, N. C., Coombs, J. J., Cui, Y., Veilleux, R. E., Buell, C. R., and Douches, D. (2015). Genetic map and QTL analysis of agronomic traits in a diploid potato population using single nucleotide polymorphism markers. *Crop Science.* 55 (6), 2566–2579. doi: 10.2135/cropsci2014.10.0745

Marand, A. P., Jansky, S. H., Gage, J. L., Hamernik, A. J., de Leon, N., and Jiang, J. (2019). Residual heterozygosity and epistatic interactions underlie the complex genetic architecture of yield in diploid potato. *Genetics.* 212 (1), 317–332. doi: 10.1534/genetics.119.302036

Masuelli, R. W., and Camadro, E. L. (1997). Crossability relationships among wild potato species with different ploidies and endosperm balance numbers (EBN). *Euphytica.* 94 (2), 227–235. doi: 10.1023/A:1002996131315

Masuelli, R. W., Camadro, E. L., and Mendiburu, A. O. (1992). 2n gametes in *Solanum commersonii* and cytological mechanisms of triplandroid formation in triploid hybrids of *Solanum commersonii* × *Solanum gourlayi*. *Genome.* 35, 864–869. doi: 10.1139/g92-131

Matesevach, M., and Barboza, G. E. (2005). Flora fanerogámica argentina. Solanaceae, parte 13 tribo i. Solanaceae. parte 2. *Solanum* subgén. *Potatoe sect. Petota.* 1-39.

Matsubayashi, M. (1991). "Phylogenetic relationships in the potato and its related species," in *Developments in plant genetics and breeding.*, vol. 2). (Amsterdam: Elsevier Science Publishers), 93–118.

Maurice, S., Montes, M. S., Nielsen, B. J., Bødker, L., Martin, M. D., Jønck, C. G., et al. (2019). Population genomics of an outbreak of the potato late blight pathogen, *Phytophthora infestans*, reveals both clonality and high genotypic diversity. *Mol. Plant Pathol.* 20 (8), 1134–1146. doi: 10.1111/mpp.12819

Maxted, N., Ford-Lloyd, B. V., Jury, S. L., Kell, S. P., and Scholten, M. A. (2006). Towards a definition of a crop wild relative. *Biodiversity Conserv.* 15 (8), 2673–2685. doi: 10.1007/s10531-005-5409-6

Mensinga, T. T., Sips, A. J., Rompelberg, C. J., van Twillert, K., Meulenbelt, J., van den Top, H. J., et al. (2005). Potato glycoalkaloids and adverse effects in humans: an ascending dose study. *Regul. Toxicol. Pharmacol.* 41 (1), 66–72. doi: 10.1016/j.yrtph.2004.09.004

Mentz, L. A., and Oliveira, P. L. (2004). O gênero *Solanum* na Região Sul do Brasil. Pesquisas, série Botânica. 54, 1–327.

Micheletto, S., Andreoni, M., and Huarte, M. A. (1999). Vertical resistance to late blight in wild potato species from Argentina. *Euphytica.* 110 (2), 133–138. doi: 10.1023/A:1003793227338

Micheletto, S., Boland, R., and Huarte, M. (2000). Argentinian wild diploid *Solanum* species as sources of quantitative late blight resistance. *Theor. Appl. Genet.* 101, 902–906. doi: 10.1007/s001220051560

Morrone, J. J. (2014). Biogeographical regionalisation of the Neotropical region. *Zootaxa.* 3782(1):1–110. doi: 10.11646/zootaxa.3782.1.1

Mweetwa, A. M., Hunter, D., Poe, R., Harich, K. C., Ginzberg, I., Veilleux, R. E., et al. (2012). Steroidal glycoalkaloids in *Solanum chacoense*. *Phytochemistry.* 75, 32–40. doi: 10.1016/j.phytochem.2011.12.003

Nahar, N., Westerberg, E., Arif, U., Huchelman, A., Guasca, A. O., Beste, L., et al. (2017). Transcript profiling of two potato cultivars during glycoalkaloid-inducing treatments shows differential expression of genes in sterol and glycoalkaloid metabolism. *Sci. Rep.* 7, 43268. doi: 10.1038/srep43268

National Plant Germplasm System, USDA (2022). Agricultural research service. Available at: https://data.nal.usda.gov/dataset/national-plant-germplasm-system (Accessed March 31, 2022).

Nicolao, R. (2021). "*Solanum malmeanum*, a promising wild relative for potato breeding," in *Dissertation thesis.* (Brazil, Pelotas: Universidade Federal de Pelotas).

Ordoñez, B., Orrillo, M., and Bonierbale, M. (2017). *Manual on potato reproductive and cytological biology* (Lima, Peru: International Potato Center).

Pandolfi, V. (1998). "Microsporogênese em *Solanum commersonii*," in *Dissertation thesis.* (Brazil, Pelotas: Universidade Federal de Pelotas).

Parrott, W. A., and Smith, R. R. (1984). Production of 2n pollen in red clover. *Crop Science.* 24, 469–472. doi: 10.2135/cropsci1984.0011183X002400030009x

Parrott, W. A., and Smith, R. R. (1986). Recurrent selection 2n pollen formation red clover. *Crop Sci.* 26, 1132–1135. doi: 10.2135/cropsci1986.0011183X002600060009x

Paudel, J. R., Gardner, K. M., Bizimungu, B., De Koeyer, D., Song, J., and Tai, H. H. (2019). Genetic mapping of steroidal glycoalkaloids using selective genotyping in potato. *Am. J. Potato Res.* 96 (5), 505–516. doi: 10.1007/s12230-019-09734-7

Pavek, J. J., and Corsini, D. L. (2001). Utilization of potato genetic resources in variety development. *Am. J. Potato Res.* 78 (6), 433–441. doi: 10.1007/BF02896375

Peloquin, S. J., Jansky, S. H., and Yerk, G. L. (1989b). Potato cytogenetics and germplasm utilization. *Am. Potato J.* 66, 629–638. doi: 10.1007/BF02853983

Peloquin, S. J., Yerk, G. L., Werner, J. E., and Darmo, E. (1989a). Potato breeding with haploids and 2n gametes. *Genome.* 31 (2), 1000–1004. doi: 10.1139/g89-174

Perrino, E. V., and Perrino, P. (2020). Crop wild relatives: Know how past and present to improve future research, conservation and utilization strategies, especially in Italy: a review. *Genet. Resource Crop Evolution.* 67, 1067–1105. doi: 10.1007/s10722-020-00930-7

Peters, J. C., Lees, A. K., Cullen, D. W., Sullivan, L., Stroud, G. P., and Cunnington, A. C. (2008). Characterization of *Fusarium* spp. responsible for causing dry rot of potato in Great Britain. *Plant Pathol.* 57 (2), 262–271. doi: 10.1111/j.1365-3059.2007.01777.x

Portal de Periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (2022). Available at https://www-periodicos-capes-gov-br.ezl.periodicos.capes.gov.br/index.php? (Accessed 31 March 2022).

Pradheep, K., Singh, M., Sultan, S. M., Singh, K., Parimalan, R., and Ahlawat, S. P. (2019). Diversity in wild relatives of wheat: an expedition collection from cold-arid Indian Himalayas. *Genet. Resour. Crop Evolution.* 66, 275–285. doi: 10.1007/s10722-018-0706-6

QGIS Development Team (2022). QGIS Geographic Information System. Open Source Geospatial Foundation Project. Available at: http://qgis.osgeo.org.

Quin, A. A., Mok, D. W. S., and Peloquin, S. J. (1974). Distribution and significance of diplandroids among the diploid solanums. *Am. Potato J.* 51 (1), 16–21. doi: 10.1007/BF02852024

Radcliffe, E. B., and Laufr, F. I. (1971). Resistance to green peach aphid and potato aphid in introductions of wild tuber-bearing *Solanum* species. *J. Economic Entomology.* 64 (5), 1260–1266. doi: 10.1093/jee/64.5.1260

Ramanna, M. S. (1974). The origin of unreduced microspores due to aberrant cytokinesis in the meiocytes of potato and its genetic significance. *Euphytica.* 23 (1), 20–23. doi: 10.1007/BF00032736

Rocha, B. H. G., Augustin, E., Silva, J. B. D., and Pereira, A. D. S. (2000). Associação entre isoenzimas e matéria seca em batata silvestre. *Pesquisa Agropecuária Brasileira.* 35 (12), 2415–2421. doi: 10.1590/S0100-204X2000001200012

Ross, R. W., and Rowe, P. R. (1965). Frost resistance among the *Solanum* species in the IR-1 potato collection. *Am. Potato J.* 42 (7), 177–185. doi: 10.1007/BF02860948

Rotem, J. (1994). The genus *Alternaria*: biology, epidemiology, and pathogenicity. *Am. Phytopathological Soc.*

Roth, M., Florez-Rueda, A. M., and Städler, T. (2019). Differences in effective ploidy drive genome-wide endosperm expression polarization and seed failure in wild tomato hybrids. *Genetics.* 212 (1), 141–152. doi: 10.1534/genetics.119.302056

Sanford, L. L., Deahl, K. L., Sinden, S. L., and Ladd, T. L. (1992). Glycoalkaloid contents in tubers from *Solanum tuberosum* populations selected for potato leafhopper resistance. *Am. Potato J.* 69 (11), 693–703. doi: 10.1007/BF02853813

Santana, F. M., Gomes, C. B., Rombaldi, C., Bianchi, V. J., and Reis, A. (2013). Characterization of *Phytophthora infestans* populations of southern Brazil in 2004 and 2005. *Phytoparasitica.* 41, 557–568. doi: 10.1007/s12600-013-0316-y

Santiago, T. R., Lopes, C. A., Caetano-Anollés, G., and Mizubuti, E. S. G. (2017). Phylotype and sequevar variability of *Ralstonia solanacearum* in Brazil, an ancient centre of diversity of the pathogen. *Plant Pathol.* 66 (3), 383–392. doi: 10.1111/ppa.12586

Schreiber, K. (1963). Über die alkaloid glykoside knollen tragen der *Solanum*-arten *Solanum*-alkaloide. *XXVII. Mitteilung. Die Kulturpflanze.* 11 (1), 422–450. doi: 10.1007/BF02136126

Scientific Electronic Library Online (2022). Available at: https://scielo.org/pt/ (Accessed 31 March 2022).

ScientificDirect (2022). Available at: https://www.sciencedirect.com/ (Accessed 31 March 2022).

Sedighian, N., Krijger, M., Taparia, T., Taghavi, S. M., Wicker, E., van der Wolf, J. M., et al. (2020). Genome resource of two potato strains of *Ralstonia solanacearum* biovar 2 (phylotype IIB sequevar 1) and biovar 2T (phylotype IIB sequevar 25) isolated from lowlands in Iran. *Mol. Plant-Microbe Interactions.* 33 (7), 872–875. doi: 10.1094/MPMI-02-20-0026-A

Simko, I., and Haynes, K. G. (2017). Maturity-adjusted resistance of potato (*Solanum tuberosum* L.) cultivars to verticillium wilt caused by *Verticillium dahliae*. *Am. J. Potato Res.* 94 (2), 173–177. doi: 10.1007/s12230-016-9553-1

Siri, M. I., Galván, G. A., Quirici, L., Silvera, E., Villanueva, P., Ferreira, F., et al. (2009). Molecular marker diversity and bacterial wilt resistance in wild *Solanum commersonii* accessions from Uruguay. *Euphytica.* 165, 371–382. doi: 10.1007/s10681-008-9800-8

speciesLink (2022). Available at: https://specieslink.net/ (Accessed 31 March 2022).

Spooner, D. M., Alvarez, N., Peralta, I. E., and Clausen, A. M. (2016). Taxonomy of wild potatoes and their relatives in Southern South America (*Solanum* sects. *Petota* and *Etuberosum*). *Systematic Bot. Monogr.* 100, 1–240+10 plates.

Spooner, D. M., and Bamberg, J. B. (1994). Potato genetic resources: sources of resistance and systematics. *Am. Potato J.* 71 (5), 325–337. doi: 10.1007/BF02849059

Spooner, D. M., Ghislain, M., Simon, R., Jansky, S. H., and Gavrilenko, T. (2014). Systematics, diversity, genetics, and evolution of wild and cultivated potatoes. *Botanical Review.* 80, 283–383. doi: 10.1007/s12229-014-9146-y

Spooner, D. M., Jansky, S., Rodríguez, F., Simon, R., Ames, M., Fajardo, D., et al. (2019). Taxonomy of wild potatoes in Northern South America (*Solanum* section *Petota*). *Systematic Bot. Monographs.* 108, 1–305.

Spooner, D. M., Van den Berg, R. G., Rodrigues, A., Bamberg, J. B., Hijmans, R. J., and Lara-Cabrera, S. (2004). Wild potatoes (*Solanum* section *Petota*; Solanaceae) of North and Central America. *Am. Soc. Plant Taxonomists.* 68, 1–209. doi: 10.2307/25027915

Städler, T., Florez-Rueda, A. M., and Roth, M. (2021). A revival of effective ploidy: the asymmetry of parental roles in endosperm-based hybridization barriers. *Curr. Opin. Plant Biol.* 61, 102015. doi: 10.1016/j.pbi.2021.102015

Summers, D., and Grun, P. (1981). Reproductive isolation barriers to gene exchange between *Solanum chacoense* and *S. commersonii* (Solanaceae). *Am. J. Botany.* 68 (9), 1240–1248. doi: 10.1002/j.1537-2197.1981.tb07831.x

Swarup, S., Cargill, E. J., Crosby, K., Flagel, L., Kniskern, J., and Glenn, K. C. (2021). Genetic diversity is indispensable for plant breeding to improve crops. *Crop Science.* 61 (2), 839–852. doi: 10.1002/csc2.20377

Tai, H. H., Worrall, K., De Koeyer, D., Pelletier, Y., Tai, G. C., and Calhoun, L. (2015). Colorado Potato beetle resistance in *Solanum oplocense* × *Solanum tuberosum* intercross hybrids and metabolite markers for selection *American Journal of Potato Research.* 92, 6, 684–696. doi: 10.1007/s12230-015-9484-2

Tai, H. H., Worrall, K., Pelletier, Y., De Koeyer, D., and Calhoun, L. A. (2014). Comparative metabolite profiling of *Solanum tuberosum* against six wild *Solanum* species with Colorado potato beetle resistance. *J. Agric. Food Chem.* 62 (36), 9043–9055. doi: 10.1021/jf502508y

Tarn, T. R., and Hawkes, J. G. (1986). Cytogenetic studies and the occurrence of triploidy in the wild potato species *Solanum commersonii* Dun. *Euphytica.* 35 (1), 293–302. doi: 10.1007/BF00028567

Tomé, L. G. O., Alves, A. A., Salgado, C. C., Davide, L. C., and Pinto, C. A. B. P. (2009). Pólen 2n e mecanismos meióticos de formação em *Solanum commersonii* ssp. *Cienc. e Agrotecnologia.* 33, 473–477. doi: 10.1590/S1413-70542009000200017

Tomé, L. G. O., Davide, L. C., Pinto, C. A. B. P., Alves, A. A., and Salgado, C. C. (2007). Pollen viability and meiotic analysis of *Solanum commersonii commersonii* Dun., *Solanum commersonii malmeanum* Bitt. and *Solanum tuberosum* L. *Crop Breed. Appl. Biotechnol.* 7, (4).

Tomé, L. G. O., Silva, A. B. D., Pinto, C. A. B. P., Davide, L. C., Pereira, D. S., and Carvalho, C. R. D. (2016). Colchicine and oryzalin effects on tetraploid induction and leaf anatomy of *Solanum commersonii* ssp. *Ciência Rural.* 46, 1973–1979. doi: 10.1590/0103-8478cr20150477

Thiers, B. (2022, continuously updated). Index Herbariorum: A global directory of public herbaria and associated staff. *New York Botanical Gardens Virtual Herbarium.* Available at: http://sweetgum.nybg.org/ih (Accessed 31 March 2022).

Tu, W., Dong, J., Zou, Y., Zhao, Q., Wang, H., Ying, J., et al. (2021). Interspecific potato somatic hybrids between *Solanum malmeanum* and *S. tuberosum* provide valuable resources for freezing-tolerance breeding. *Plant Cell Tissue Organ Culture.* 147 (1), 73–83. doi: 10.1007/s11240-021-02106-2

Umemoto, N., Nakayasu, M., Ohyama, K., Yotsu-Yamashita, M., Mizutani, M., and Seki, H. (2016). Two cytochrome P450 monooxygenases catalyze early hydroxylation steps in the potato steroid glycoalkaloid biosynthetic pathway. *Plant Physiol.* 171 (4), 2458–2467. doi: 10.1104/pp.16.00137

Vega, S. E., and Bamberg, J. B. (1995). Screening the U.S. potato collection for frost hardiness. *Am. Potato J.* 72, 13–21. doi: 10.1007/BF02874375

Vleeshouwers, V. G. A. A., Raffaele, S., Vossen, J. H., Champouret, N., Oliva, R., Segretin, M. E., et al. (2011). Understanding underexploiting late blight resistance in the age of effectors. *Annu. Rev. Phytopathology.* 49, 507–531. doi: 10.1146/annurev-phyto-072910-095326

Warschefsky, E., Penmetsa, R. V., Cook, D. R., and Von Wettberg, E. J. (2014). Back to the wilds: Tapping evolutionary adaptations for resilient crops through systematic hybridization with crop wild relatives. *Am. J. Botany.* 101 (10), 1791–1800. doi: 10.3732/ajb.1400116

Werner, J. E., and Peloquin, S. J. (1991a). Potato haploid performance in 2x x 4x crosses. *Am. Potato J.* 68, 801–811. doi: 10.1007/BF02853854

WIEWS (2022). World Information and Early Warning System on Plant Genetic Resources for Food and Agriculture. Available at: https://www.fao.org/wiews/en/ (Accessed 31 March 2022).

Wolters, P. J., Wouters, D., Kromhout, E. J., Huigen, D. J., Visser, R. G., and Vleeshouwers, V. G. (2021). Qualitative and quantitative resistance against early blight introgressed in potato. *Biology.* 10(9), 892. doi: 10.3390/biology10090892

Xue, W., Haynes, K. G., Clarke, C. R., and Qu, X. (2022). Genetic dissection of early blight resistance in tetraploid potato. *Front. Plant Science.* 13. doi: 10.3389/fpls.2022.851538

Zhang, H., Mittal, N., Leamy, L. J., Barazani, O., and Song, B. H. (2017). Back into the wild–apply untapped genetic diversity of wild relatives for crop improvement. *Evolutionary Applications.* 10, 5–24. doi: 10.1111/eva.12434

Zhang, C., Wang, P., Tang, D., Yang, Z., Lu, F., Qi, J., et al. (2019). The genetic basis of inbreeding depression in potato. *Nat. Genet.* 51, 374–378. doi: 10.1038/s41588-018-0319-1

Zhen, P., Wang, P., Tang, D., Sang, Y., Li, C., Huang, S., et al. (2019). Inheritance of steroidal glycoalkaloids in potato tuber flesh. *J. Integr. Agriculture.* 18 (10), 2255–2263. doi: 10.1016/S2095-3119(19)62718-8

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 26 April 2023
DOI 10.3389/fpls.2023.1046400

Check for updates

OPEN ACCESS

EDITED BY
Manuela Nagel,
Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Germany

REVIEWED BY
Dorcus Chepkesis Gemenet,
International Potato Center, Peru
Stephan Weise,
Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Germany

*CORRESPONDENCE
Roxana Yockteng
ryockteng@agrosavia.co

†PRESENT ADDRESS
Norma C. Manrique-Carpintero,
Alliance of Bioversity International and The International Center for Tropical Agriculture (CIAT), Cali, Colombia

SPECIALTY SECTION
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

RECEIVED 16 September 2022
ACCEPTED 14 March 2023
PUBLISHED 26 April 2023

CITATION
Manrique-Carpintero NC, Berdugo-Cely JA, Cerón-Souza I, Lasso-Paredes Z, Reyes-Herrera PH and Yockteng R (2023) Defining a diverse core collection of the Colombian Central Collection of potatoes: a tool to advance research and breeding. *Front. Plant Sci.* 14:1046400. doi: 10.3389/fpls.2023.1046400

COPYRIGHT
© 2023 Manrique-Carpintero, Berdugo-Cely, Cerón-Souza, Lasso-Paredes, Reyes-Herrera and Yockteng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Defining a diverse core collection of the Colombian Central Collection of potatoes: a tool to advance research and breeding

Norma Constanza Manrique-Carpintero[1†], Jhon A. Berdugo-Cely[1,2], Ivania Cerón-Souza[1], Zahara Lasso-Paredes[1], Paula H. Reyes-Herrera[1] and Roxana Yockteng[1,3*]

[1]Corporación Colombiana de Investigación Agropecuaria-AGROSAVIA, Centro de Investigación Tibaitatá, Mosquera, Colombia, [2]Corporación Colombiana de Investigación Agropecuaria-AGROSAVIA, Centro de Investigación Turipaná, Montería, Colombia, [3]Institut de Systématique, Evolution, Biodiversité-UMR-CNRS 7205, National Museum of Natural History, Paris, France

The highly diverse Colombian Central Collection (CCC) of cultivated potatoes is the most important source of genetic variation for breeding and the agricultural development of this staple crop in Colombia. Potato is the primary source of income for more than 100.000 farming families in Colombia. However, biotic and abiotic challenges limit crop production. Furthermore, climate change, food security, and malnutrition constraints call for adaptive crop development to be urgently addressed. The clonal CCC of potatoes contains 1,255 accessions —— an extensive collection size that limits its optimal assessment and use. Our study evaluated different collection sizes from the whole clonal collection to define the best core collection that captures the total genetic diversity of this unique collection, to support a characterization more cost-effectively. Initially, we genotyped 1,141 accessions from the clonal collection and 20 breeding lines using 3,586 genome-wide polymorphic markers to study CCC's genetic diversity. The analysis of molecular variance confirmed the CCC's diversity with a significant population structure (Phi=0.359; *p-value*=0.001). Three main genetic pools were identified within this collection (CCC_Group_A, CCC_Group_B1, and CCC_Group_B2), and the commercial varieties were located across the pools. The ploidy level was the main driver of pool identification, followed by a robust representation of accessions from Phureja and Andigenum cultivar groups based on former taxonomic classifications. We also found divergent heterozygosity values within genetic groups, with greater diversity in genetic groups with tetraploids (CCC_Group_B1: 0.37, and CCC_Group_B2: 0.53) than in diploid accessions (CCC_Group_A: 0.14). We subsequently generated one mini-core collection size of 3 percent (39 entries) and three further core collections sizes of 10, 15, and 20 percent (i.e., 129, 194, and 258 entries, respectively) from the total samples genotyped. As our results indicated that genetic diversity was similar across the sampled core collection sizes compared to the main collection, we selected the smallest core collection size of 10 percent. We expect this 10 percent core collection to be an optimal

tool for discovering and evaluating functional diversity in the genebank to advance potato breeding and agricultural-related studies. This study also lays the foundations for continued CCC curation by evaluating duplicity and admixing between accessions, completing the digitalization of data, and ploidy determination using chloroplast count.

KEYWORDS

genetic diversity, molecular markers, population structure, mini-core collection, polyploid

Introduction

Crop diversity is the primary source of genetic variation for crop improvement and the development of cultivated species. Potato (*Solanum tuberosum* L.) is a species with great diversity in the primary gene pool. As domestication took place, several cultivated species co-evolved, from the most ancient cultivated potato, *S. stenotomum*, to the modern and most globally commercialized potato (Spooner et al., 2014). Based on the latest taxonomic classification, there are four cultivated potato species, *Solanum tuberosum* L., with two cultivar groups, Andigenum and Chilotanum, *S. ajanhuiri*, *S. juzepczukii*, and *S. curtilobum* (Spooner et al., 2007). The Andigenum group collapsed several former species from the previously accepted Hawkes (1990) classification, combining *S. stenotomum* Juz. & Bukasov, *S. phureja* Juz. & Bukasov, *S. chaucha* Juz. & Bukasov, and *S. tuberosum* subsp. *andigenum* Hawkes. Various dynamic and evolutionary factors influencing the traditional potato cropping of Andean communities generated the broad genetic diversity of cultivated potatoes. Those factors are a continuous hybridization of cultivated diversity mediated by the outcrossing nature of most diploids, the production of unreduced gametes, the presence of cultivated diversity with different ploidy levels in the same field, simultaneous sexual and asexual reproduction of which vegetative propagation is the main system, and possible crosses with wild germplasm growing close to cultivated fields. All of these factors promoted the generation and maintenance of the enormous variation of cultivated potato varieties over time (Dodds and Paxman, 1962; Huamán and Spooner, 2002; Spooner et al., 2010).

Potato landraces originated in South America; most of them are from the highlands of the Andes (found at 3000 – 4000 m of elevation), from western Venezuela to the south in southern Bolivia and northern Argentina, except for the Chilotanum group landraces found in the lowlands of the central and south-central area of Chile (Spooner et al., 2010). The landraces grown in Central America and Mexico came from Post-Columbian introductions. Based on the analysis of the eco-geographical distribution of potato landraces, the *S. tuberosum* Andigenum group has the most distant geographic coverage, with the tetraploids most widespread, followed by diploid and triploid ploidy levels. Like the Chilotanum group, *S. ajanhuiri*, *S. curtilobum*, and *S. juzepczukii* have a much smaller distribution, restricted to central Peru and Bolivia. From this ample distribution, several thousand potato landraces have been reported with broad genetic diversity that can be described at various levels: morphological, physiological, and agronomic are just some to be highlighted. Some authors have reported extensive diversity in tuber skin and flesh color, tuber and leaf shape, floral colors, growth habit, maturity, dormancy, and photoperiod length needed for tuberization, in addition to pest and disease resistance, environmental responsiveness, and tuber yield (Spooner et al., 2010; Ellis et al., 2020; Bradshaw, 2021). This extensive diversity represents a reservoir of genes for crop improvement, food security, and adaptability to climate change (ibid).

Colombia is considered a center of cultivated potato diversity, as it is one of the Andean countries where the potato was domesticated. Potato is an important staple for food security in Colombia and the primary source of income for around 100.000 farming families (Ministerio de Agricultura y Desarrollo Rural, 2020). Potato diversity in Colombia is conserved *in situ* and *ex situ*. The *in situ* conservation depends on some isolated indigenous communities and smallholder farmers who still conserve traditional landraces (Tinjacá and Rodríguez, 2015). This is because potatoes cultivation in most of the territory uses modern commercial practices. On the other hand, *ex situ* potato conservation is carried out by the Colombian Corporation for Agricultural Research-AGROSAVIA (Corporación Colombiana de Investigación Agropecuaria in Spanish) and universities such as Universidad Nacional de Colombia, among others, using three different systems: seeds, field, and *in vitro*. Farmers and private companies also try to conserve and promote the commercialization of native potatoes. The Colombian Central Collection (CCC) of potato was started in the 1950s as the basis for a genetic platform for potato research and breeding initiative promoted by several Colombian plant scientists, agronomists, and farmers, supported by the Agriculture Minister of Colombia, and guided by Dr. Jack Hawkes (Luján, 1976; Pineda Colorado and Hernández Castillo, 1996) who established the national potato breeding program. Today, AGROSAVIA manages the CCC and conserves 2,499 wild and cultivated potato accessions. The clonal collection of cultivated germplasm has 1,255 accessions, most of which belong to the *S. tuberosum* Andigenum group and a few accessions to the *S. tuberosum* Chilotanum group based on Spooner et al. (2014). However, in the CCC, the germplasm is classified as *S. phureja*, *S. tuberosum* subsp. *andigenum*, and *S. tuberosum* subsp. *tuberosum*, following Hawkes (1990) taxonomic classification.

The CCC diversity has been important for advancing research and breeding of potato in Colombia. The assessment of genetic diversity in part of the clonal CCC showed that it is highly diverse, with high and significant population structure and greater percentage of genetic variation within than between populations (Juyó et al., 2015; Berdugo-Cely et al., 2017). This collection has been used as a source for late blight resistance caused by *Phytophthora infestans*, Guatemalan tuber moth *Tecia solanivora*, and drought stress (Santa Sepúlveda, 2018; Díaz-Valencia et al., 2021). These are some of the most limiting constraints of potato crop production. Moreover, the CCC of potatoes has been used for the generation and release of new varieties over the last 70 years in Colombia (Pineda Colorado and Hernández Castillo, 1996; Ñustez, 2011). Similarly, in 1995 AGROSAVIA scientists initiated the morphological characterization of the CCC, together with different evaluations of essential traits for breeding, such as late blight resistance, maturity, and quality attributes (Moreno and Valbuena, 2006; Berdugo-Cely et al., 2017). Most of the evaluations were done by groups of accessions over an extended timeline because of the collection size and limited resources at different levels, such as trained personnel, funding, infrastructure, etc. The potato crop in Colombia is subject to several biotic and abiotic challenges, besides food security, malnutrition, and climate-change constraints. Thus, the CCC has been and will remain the primary source of genetic diversity for crop improvement in Colombia.

The size of a collection, particularly extensive collections, could limit the possibility of fully documenting and using the germplasm because of the high logistics and infrastructure resources required to undertake research and evaluation of the collection. Thus, Frankel (1984) proposed rationalizing the collections by reducing redundancy and generating core collections, a representative subset of the whole collection diversity. In general, the purpose of a core collection is to maximize the representation of full genetic diversity. However, a core collection can also be assembled to capture the diversity in contrasting or extreme trait phenotypes, or to represent patterns of diversity based on different criteria such as geographic distribution, the proportion of diversity by clusters, and diversity of a trait, among others (Odong et al., 2013). In a core collection —— designed to represent the full diversity of the primary collection —— each selected accession or entry represents all the accessions in the entire collection that are similar to it. Thus, each entry in the core collection could represent one or more accessions in the entire collection. This is possible because the selection occurs from the center of unique weights or clusters that conform to a uniform representation across the whole diversity. Odong et al. (2013) recommended using genetic distance-based metrics to select entries in a core collection representing genetic diversity, precisely the minimal average distance between each Accession and the Nearest Entry (A–NE). Thus, to guarantee a good selection of a core collection, the value of the A–NE distance should be the smallest possible, considering that itself (A–NE=0) is the maximum representation of one accession. Since a genetic distance-based metric is the best criterion for generating a diverse core collection, molecular markers rather than phenotypic or passport data should be used to find a core collection that maximizes the representation of all genetic diversity in the main collection. Most molecular markers are selectively neutral, unaffected by the environment, and are more suitable for statistical genetic diversity analysis (Brown, 1989; Chapman and Crawford, 1990). Developing many markers with genome-wide coverage that ensure the best genetic diversity representation by the selected entries is straightforward today.

This study evaluated the CCC´s population structure and genetic diversity using genome-wide single nucleotide polymorphisms (SNPs). Similarly, we assembled a core and mini-core collection using the Average Accession to Nearest Entry (A-NE) method and validation recommended by Odong et al. (2013). The best core and mini-core collections were identified by comparing their genetic diversity measurements with the one in the whole CCC. Then, the quality and usefulness of the selected core collection was validated by comparing the phenotypic diversity representing the whole CCC with the core and mini core collections for three different agronomic traits (Average Tuber Weight in g (ATW), Number of Tubers per Plant (NTP), and Total Tuber Yield (TTY) in kg/plant). Finally, in this study, we set the basis to continue the CCC curation by evaluating duplicity and admixing between accessions, completing the characterization, data digitalization, and ploidy determination using chloroplast count.

Materials and methods

Plant material

For this study, we used 1,141 accessions from the field and *in vitro* copies taken from the clonal CCC of potatoes genebank managed by AGROSAVIA and 20 commercial varieties and advanced breeding lines obtained from different sources. This CCC contains mainly landraces and a small proportion of released varieties. Based on Spooner et al. (2007) taxonomy, most accessions belong to the species *S. tuberosum* group Andigenum, and few accessions to *S. tuberosum* group Chilotanum. However, the classification of the material in the potato collection follows Hawkes (1990) previously accepted taxonomic classification. According to this classification, 143 accessions are *S. phureja*, 681 *S. tuberosum* subsp. *andigenum*, 67 *S. tuberosum* subsp. *tuberosum*, one (1) *S. rybinii* (older classification), while for 269 accessions we currently do not have the information because it is in the curation process and digitalization of data from hard copy records. Based on the original classification, this study named the accessions Andigena, Phureja, and Tuberosum types. The geographic origin of the accessions is diverse; we have records for 802 accessions, from which 670 are from Colombia, 95 from other Latin American countries, 26 from Europe, and 11 from the United States (Table 1).

Sampling and genotyping

We collected leaf tissue samples for each accession from 1–2 month-old *in vitro* plantlets or field grown plants at vegetative stage. In the latter case, the tissue was collected from one randomly

TABLE 1 Geographic origin of accessions held in the Colombian Central Collection of potatoes conserved at the Corporación Colombiana de Investigación Agropecuaria (AGROSAVIA).

Country of origin	Number of accessions
Argentina	1
Belgium	1
Bolivia	14
Brazil	1
Colombia	670
Ecuador	13
Germany	2
Netherlands	11
Mexico	5
Perú	60
Scotland	4
United Kingdom	8
United States	11
Venezuela	1
Total	802

selected plant out of the 20 clones grown per accession. The CCC potato curators regenerate the field collection annually in a highland area of Zipaquirá municipality, department of Cundinamarca, Colombia, at 2.950 m altitude, with an average temperature of 15°C and relative humidity of 75 percent. We extracted DNA from the leaf tissue using the DNeasy Plant Mini Kit (Qiagen, Valencia, CA, USA), following the manufacturer's technical recommendations. Then, we genotyped each accession with the Illumina single nucleotide polymorphism (SNP) Infinium Potato Array version 8303 (Felcher et al., 2012) and scanned the fluorescent signals in the Illumina HiScan SQ system (Illumina, San Diego, CA) at AGROSAVIA. A set of 809 samples already had a genotyping profile processed by Berdugo-Cely et al. (2017). Therefore, we genotyped 482 additional samples to obtain a total of 1,291 samples genotyped and used in this study (Supplementary Table S1). Of these samples, 1,061 are unique accessions and 130 are biological repetitions. The accessions in the clonal CCC did not have unique identification numbers, and samples were collected from field and *in vitro* sources using another nomenclature. As part of the CCC curation process, the unique identifier numbers were assigned during this genotyping process; this allowed us to identify several samples that had been genotyped multiple times. Due to limitations in budget, samples were genotyped in batches in order to complete the samples evaluated in this study, which contributed to accumulating several duplicates in the genotyping process. We obtained the raw values from the fluorescent signals for all samples using GenomeStudio 2.0 software (Illumina, San Diego, CA). Clustercall package (Schmitz Carley et al., 2017) in the R platform (R Core Team, 2021) transformed these signal values to obtain the tetraploid genotype calls (AAAA, AAAB, AABB, ABBB, BBBB coded as 0, 1, 2, 3, 4).

Population structure and genetic diversity

The final SNP genotype matrix was obtained from the ClusterCall output matrix, filtered to remove markers with minor allele frequency of less than 0.03 and missing data greater than 5 percent at the genotype and marker level. We assessed population structure using this dataset with tetraploid genotype calls through two strategies: i) a Bayesian model implemented in the software STRUCTURE 2.3 (Pritchard et al., 2000) without *a priori* population information, and ii) a Principal Component Analysis (PCA) calculated using the R-Adegenet package (Jombart and Ahmed, 2011). For the Bayesian model, the number of populations (in this study referred to as genetic groups) K was estimated from one to ten using an admixture model with frequencies of correlated alleles, a burn-in set to 150.000 interactions, and ten independent repetitions per run. Then, we found the best number of populations in the Structure Harvester program (Earl and vonHoldt, 2012) using the Evanno et al. method (Evanno et al., 2005). The NbClust R-package (Charrad et al., 2014) was used to determine the K for the PCA loading as input the information of the first three components. We tested if there was population differentiation by calculating the F_{ST} and the percentages of differentiation between and within identified subpopulations using the analysis of molecular variance (AMOVA) in the R-packages dartR (Mijangos et al., 2022) and Poppr (Kamvar et al., 2014), respectively. Finally, we manually estimated the genetic diversity by calculating the observed heterozygosity (Ho) per genotype and genetic group, using the methodology described by Berdugo-Cely et al. (2021).

Ploidy-level prediction

We predicted the ploidy level of each accession using the proportion of simplex and triplex genotype scores (ABBB, AAAB) across all SNPs per accession, according to Alsahlany et al. (2019). This method sets up the parameters to predict the ploidy using a set of reference samples. In that case, when the sum of simplex and triplex proportions was close to zero, the sample would be diploid, and when the frequency was greater than 0.2 it would be tetraploid. In this study, we validated the accuracy of this method by comparing the ploidy of 112 reference accessions from the CCC assessed with SNP genotype score frequencies and the ploidy values obtained using chloroplast and chromosome counts from three publications (Guevara, 2011; Uribe Gaviria, 2011; Sánchez, 2017). For this analysis, we implemented a one-way analysis of variance (ANOVA) and Tukey test in R software to find significant differences among the ploidy prediction strategies. Finally, the parameters were established to assess the ploidy in the samples used in this study.

Core collections

We used the Core Hunter 3 program (de Beukelaer et al., 2018) to define core collections representing the total genetic diversity present in the CCC of potatoes. The Modified Roger's genetic distances (MRD) were calculated and used to identify the core sets with the A–NE distance function. This function optimizes the entry selection to minimize the average distance between each accession in the whole collection and the closest-selected entry, as expected when creating a diverse core collection. We tested three collection sizes of 10, 15, and 20 percent of the total samples evaluated in this study (i.e., 129, 194, and 258 entries, respectively) and one mini-core collection with 3 percent of the original size (39 entries).

We assessed the quality of the A-NE method for core collection selection by calculating and comparing the average criterion distance for A-NE and two additional distance-based metrics evaluated by Odong et al. (2013) for each of the core and mini-core collections identified: (1) average distance between each entry and the nearest neighboring entry (E-NE), and (2) average genetic distances between entries (E-E). The EvaluateCore R package v.0.1.3 (Aravind et al., 2022) and two additional core sizes (50 and 80 percent) were used for this comparison.

To validate that the genetic diversity representation was kept across collections and select the most suitable core collection for breeding and research, we compared the average observed heterozygosity (Ho) measurements from the core and mini-core collections to the average in the whole CCC. Additionally, a PCA was carried out for each collection to show the distribution of the samples selected in each proposed core collection compared to the whole CCC. Finally, for the 3 percent mini-core collection identified using Core Hunter, we replaced some selected entries to match and contain all accessions within the 10 percent core collection. For this, we identified the entries that differed between the 3 percent mini-core and 10 percent core collections from the initial Core Hunter entry selection. Then, using the MRD matrix, we calculated the difference between the genetic distance of each entry identified in the 3 percent mini-core and the 10 percent entries to identify the 10 percent entry with the minor genetic distance difference. Finally, the 3 percent entries were replaced with the corresponding 10 percent core entry identified, which were instead selected for 19 accessions.

Validation of core and mini-core collections' usefulness

We validated the effectiveness of the accessions in the core and mini-core collections to represent the whole spectrum of phenotypic variation for three different traits: i) Average Tuber Weight (ATW) in grams, ii) Number of Tubers per Plant (NTP), and iii) Total Tuber Yield (TTY) in kg/plant. For this we used the phenotypic data of 846 genotyped accessions (206 for CCC_Group_A, 494 for CCC_Group_B1, and 146 for CCC_Group_B2) for which preliminary field evaluations had been conducted for ATW, NTP, and TTY. Data were collected between 2013 and 2015. Accession data collected across these years depends on the accessions evaluated during each field campaign. Due to financial and operational capacity limitations, we prioritized accessions and traits to be evaluated by year. Andigenum and later the Phureja accessions were evaluated in three consecutive years. The evaluations were conducted in a highland area of Zipaquirá municipality, in the department of Cundinamarca. Field data for each year was collected from an average of 20 clones planted in one tier per accession, using a complete randomized design for the entire CCC in the field. Thus, each year's data represents a repetition. We used the mean values from all-year evaluations per trait and accession for the analyzes. We expected that, independently of the size, a good core collection would show a similar phenotypic distribution as the one for the 846 accessions. Moreover, we analyzed to see if the identified entries for the different core and mini-core collections represented the phenotypic diversity of all accessions by comparing their average criterium distance for A-NE, E-NE, and E-E metrics. The Gower distance (Pavoine et al., 2009) between accessions was calculated using the phenotypic variable matrix. Then, like in the above evaluation for SNPs, the EvaluateCore R program was loaded with this information to calculate the different distance-based metrics. We used selected entries in the core and mini-core collections that contained phenotypic data.

Results

Population structure and genetic diversity

After quality controls and filters, we identified 3,586 genome-wide SNP markers to conduct the analyzes (Supplementary Table S2). The population structure analysis using the STRUCTURE and NbClust approaches suggested two and three genetic groups within the CCC germplasm, respectively (Figures 1A, B, 2A, B and Supplementary Figures S1A, B). The ploidy level was the main driver of pool identification, followed by the taxonomic classification initially assigned to the different accessions in the collection (Table 2). The population genetic analyzes supported these results, showing a significant solid genetic structure for the two (Phi=0.370; $p\text{-}value$=0.001) and three (Phi=0.359; $p\text{-}value$=0.001) genetic groups detected in the CCC. The PCA also showed similar results (Figures 1B, 2B). The first and second components explained 19.94 percent and 8.04 percent of the sample variance, endorsing the differentiation of accessions by ploidy level and taxonomy with some level of admixture.

In the latest taxonomic classification, the most traditional and modern cultivated potatoes are classified within a single potato species. However, the Hawkes taxonomy originally assigned to the CCC aligns with a clear genetic differentiation inherent to the germplasm. Thus, for this paper, we designated the accessions as Phureja, Andigenum, and Tuberosum types. The population structure analysis conducted using STRUCTURE discriminated the CCC into two genetic groups, the CCC_Group_A with mainly diploid (2n=2x=24) and CCC_Group_B with mostly tetraploid (2n=4x=48) accessions. The second analysis conducted using Nbclust detected three genetic groups, the same diploid

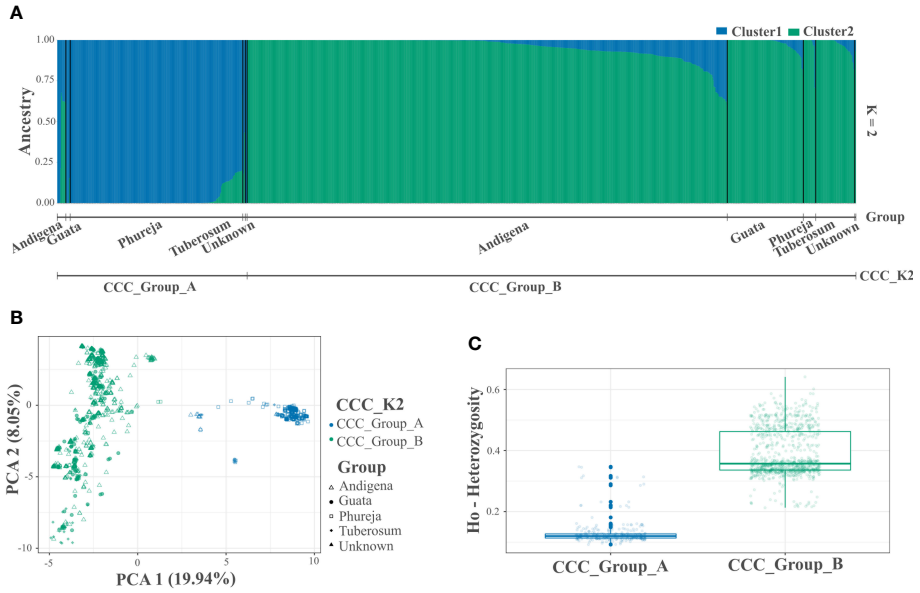


FIGURE 1
Population structure and genetic diversity analyzes of the Colombian Central Collection (CCC) of potatoes for two genetic groups (CCC_K2). **(A)** Bar plot of genetic structure analysis for two genetic groups (K=2) identified using the STRUCTURE software and designated as CCC_Group_A and CCC_Group_B (lower line). The AGROSAVIA in-house classification (Group) is indicated in the upper line (Andigena, Phureja, Tuberosum, "Guata" and Unknown), **(B)** Scatter plot of the Principal Component Analysis (PCA) color coded by genetic group and shaped following AGROSAVIA in-house biological status classification (Group), and **(C)** Box plot with distribution (box representing the second and third quartiles, line across the median, and vertical line the range of data) with distribution of observed heterozygosity (Ho) for each genetic group.

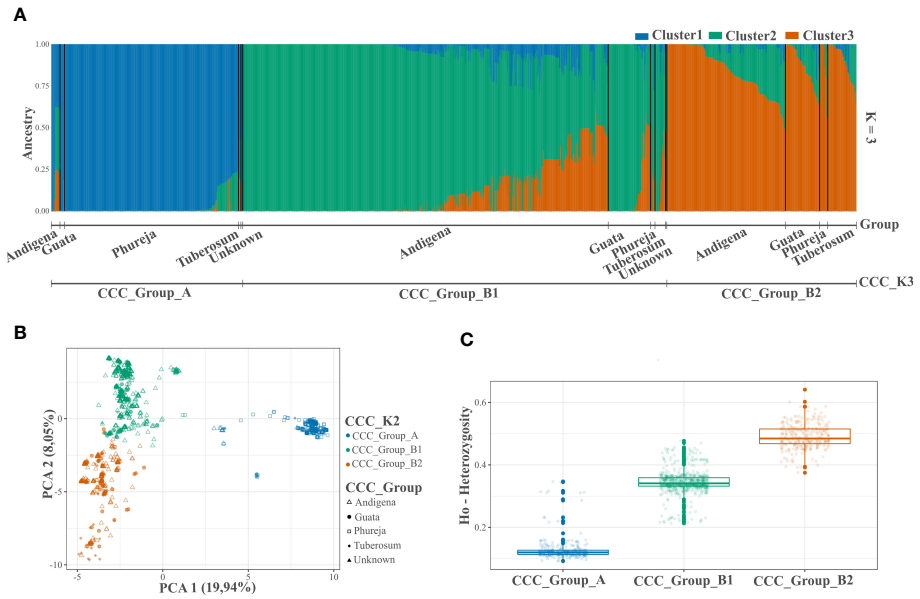


FIGURE 2
Population structure and genetic diversity analyzes of the Colombian Central Collection (CCC) of potatoes for three genetic groups (CCC_K3). **(A)** Bar plot of genetic structure analysis for three genetic groups (K=3) identified using the NbClust R-package and designated as CCC_Group_A, CCC_Group_B1, and CCC_Group_B2 (lower line). The AGROSAVIA in-house biological status germplasm classification (CCC_Group) is indicated in the upper line (Andigena, Phureja, Tuberosum, "Guata" and Unknown), **(B)** Scatter plot of the Principal Component Analysis (PCA) color coded by genetic group and shaped following AGROSAVIA in-house biological status classification (CCC_Group), and **(C)** Box plot with distribution (box representing the second and third quartiles; line across the median, and vertical line representing the range of data) with distribution of observed heterozygosity (Ho) for each genetic group.

TABLE 2 Taxonomic and ploidy characteristics of three genetic groups identified in the Colombian Central Collection (CCC) of potatoes, based on population structure and genetic diversity analyzes.

Genetic group	Taxonomic classification	Ploidy*			Total	Percentage
		2*x*=2n=24	2*x*=4n=48	Unknown		
CCC_Group_A		239	9	–	248	21.3%
	S. rybinii	1	–	–	1	0.4%
	S. tuberosum subsp. *andigenum*	5	2	–	7	2.8%
	S. tuberosum subsp. *phureja*	136	1	–	137	55.2%
	S. tuberosum subsp. *tuberosum*	4	–	–	4	1.6%
	S. tuberosum sp. (In the curation process)	93	6	–	99	39.9%
CCC_Group_B1		–	619	21	640	55.1%
	S. tuberosum subsp. *andigenum*	–	505	16	521	81.4%
	S. tuberosum subsp. *phureja*	–	2	–	2	0.3%
	S. tuberosum subsp. *tuberosum*	–	16	1	17	2.7%
	S. tuberosum sp. (In the curation process)	–	96	4	100	15.6%
CCC_Group_B2		–	273	–	273	23.5
	S. tuberosum subsp. *andigenum*	–	153	–	153	56.0%
	S. tuberosum subsp. *phureja*	–	4	–	4	1.5%
	S. tuberosum subsp. *tuberosum*	–	46	–	46	16.8%
	S. tuberosum sp. (In the curation process)	–	70	–	70	25.6%
Total		239	901	21	1161	100%

*Ploidy assessment base on the frequency of simplex and triplex Single Nucleotide Polymorphism (SNP) marker scores per accession.

cluster (CCC_Group_A) previously identified, while the tetraploid cluster was separated into two new genetic groups (CCC_Group_B1 and CCC_Group_B2). As shown in Table 2, the CCC_Group_A genetic group contained 248 accessions. Of these, 96.4 percent (n=239) were diploids, and 3.6 percent (n=9) were tetraploids. At the taxonomic level, most of the accessions (n=137, 55.2 percent) were Phureja type, while 2.8 percent (n=7) Andigenum, 1.6 percent (n=4) Tuberosum, 0.4 percent (n=1) was *S. rybinni*, and for 39.9 percent (n=99) we have no information. There has been no digitalization of hard copy records of the taxonomic classification assigned to these accessions when the CCC material was acquired. Therefore, this documentation is currently in the process of curation. The CCC_Group_B1 had 640 accessions, 96.7 percent (n=619) tetraploid accessions, and 3.2 percent (n=21) with unknown ploidy. Taxonomically, most of the accessions, 81.4 percent (n=521), were Andigenum, and a lower proportion Tuberosum, 2.7 percent (n=17), and Phureja, 0.3 percent (n=2). The remaining 15.6 percent (n=100) are currently undergoing taxonomic classification. Finally, the CCC_Group_B2 genetic group had 100 percent (n=273) tetraploid accessions. At the taxonomic level, 56 percent (n=153) corresponded to Andigenum, 16.8 percent (n=46) Tuberosum, 1.5 percent (n=4) Phureja, and 25.6 percent (n=70) are pending taxonomic classification. Finally, the accessions evaluated in this study that are released varieties or breeding program germplasm —— some of which are part of the CCC —— were evenly located across genetic groups. Accordingly, some were diploids and other tetraploids.

The genetic differentiation is significant between the CCC_Group_A and the CCC_Group_B genetic groups (F_{ST}=0.146, *p-value*<0.0001). The CCC_Group_A shows lower genetic differentiation with CCC_Group_B1 than with CCC_Group_B2 genetic groups (F_{ST}=0.164, *p-value*<0.0001 and 0.206, *p-value*<0.0001 respectively), while the tetraploid groups are more similar (F_{ST}=0.071, *p-value*<0.0001). Thus, the CCC_Group_B1 has an intermedium differentiation level between CCC_Group_A and CCC_Group_B2 genetic groups. The results show high genetic variation in the CCC germplasm, confirmed by a higher significant variation within genetic groups [62.95 percent for genetic structure with two groups (K2) and 64.08 percent for three groups (K3)] than between groups (37.04 percent for K2 and 35.91 percent for K3). Based on divergent heterozygosity values (Ho) for the genetic structure of two (CCC_Group_A: 0.14 ± 0.04 and CCC_Group_B: 0.42 ± 0.08) and three genetic groups (CCC_Group_A: 0.14 ± 0.04; CCC_Group_B1: 0.37 ± 0.04, and CCC_Group_B2: 0.53 ± 0.04), the genetic variation within the CCC_Group_B, B2, and B1 groups was greater than that observed in CCC_Group_A group (Figures 1C, 2C).

Ploidy assessment

We used a reference sample of 112 accessions to confirm the accuracy of predicting the ploidy level of the CCC accessions using the proportion of simplex and triplex SNP marker scores per sample, according to Alsahlany et al. (2019). This assessment was compared to another indirect ploidy assessment method, the average number of chloroplasts per guard cell, and —— a direct approach —— the number of chromosomes counted in root cells from previously published data (Guevara, 2011; Uribe Gaviria, 2011; Sánchez, 2017). In general, the reference set of samples showed a pattern for the number of chloroplasts per guard cell and the proportion of simplex and triplex scores of all SNPs per sample associated with the number of chromosomes (Figure 3 and Supplementary Table S1). The ploidy level based on chromosome count, the direct and precise method of ploidy assessment, for the 112 accessions in the reference set included 83 diploids (2n=2*x*=24), 27 tetraploids (2n=4*x*=48), and two triploids (2n=3*x*=36). As expected, the *Solanum* genus' primary number of chromosomes (12) was duplicated 2–4 times according to each accession's ploidy level (Figure 3A and Supplementary Table S1). Thus, this ploidy assessment was used to compare the accuracy of the indirect methods. The chloroplast counts in guard cells matched an average of 7.6 ± 0.83 chloroplasts/guard cells for diploid accessions and 13.2 ± 1.06 chloroplasts/guard cells for tetraploid accessions. The tetraploid group differed significantly from the diploid group (*p-value* <0.0001). The chloroplast count in guard cells could not predict triploid accessions. Most diploid accessions (74/83, 89.2 percent) had 7–8 chloroplasts/guard cell average, and a few (9/83, 10.8 percent) were in a grey area between 9–10 chloroplasts/guard cell average. The tetraploids were more consistent with 12–14 chloroplast/guard cells for 96 percent of accessions (26/27) and one in a grey area of 10 chloroplasts/guard cells.

When using the proportion of SNP genotype scores as an indirect method for assessing ploidy, this analysis showed an average of 4.8 percent ±9.04 simplex and triplex SNP genotype scores for diploid accessions and a 27.1 percent ±1.53 for tetraploid accessions. Moreover, the ANOVA confirmed significant differences between the average for diploid and tetraploid samples (*p-value* < 0.001), as shown in Figure 3C and Supplementary Table S1. Similarly to the chloroplast count method, this method could not discriminate triploid samples. From the total diploid accessions based on chromosome count, for 77 the proportion of simplex and triplex SNP scores was 0–15 percent, and for six accessions the frequency of simplex and triplex SNP scores was off-type with 24–41 percent. For the tetraploid samples, the proportion of simplex and triplex SNP genotype scores per accession was accurate for 26 tetraploid samples with values of 22–42 percent and one sample off-type with 0 percent. The mismatch between methods could be due to sample admixture between the first and current evaluations since the sampling for each assessment occurred at separate times. Unlike for chloroplast and chromosome counts, this mismatch did not occur, and the methods were assessed using samples from the same source of plant materials. If we consider them real method errors, ploidy prediction accuracy using simplex and triplex SNP genotype scores per accession was 92.8 percent (77/83) for diploids and 96.3 percent for tetraploids (26/27).

Based on these results, we predicted the ploidy of all samples used in this study using a threshold of 0–15 percent frequency of simplex and triplex SNP genotype scores per accession for diploids and greater than 20 percent for tetraploids. The unknown ploidy corresponded to accessions with a frequency of simplex and triplex in a grey area of 15–20 percent. Of the 1,291 genotyped samples, 239 unique accessions were diploids, 901 tetraploids, and 21 unknowns (Table 2). The ploidy for the 130 biological repetitions included 59 diploids, 70 tetraploids, and one unknown. For 101 accessions with biological replicates, the ploidy assessment matched

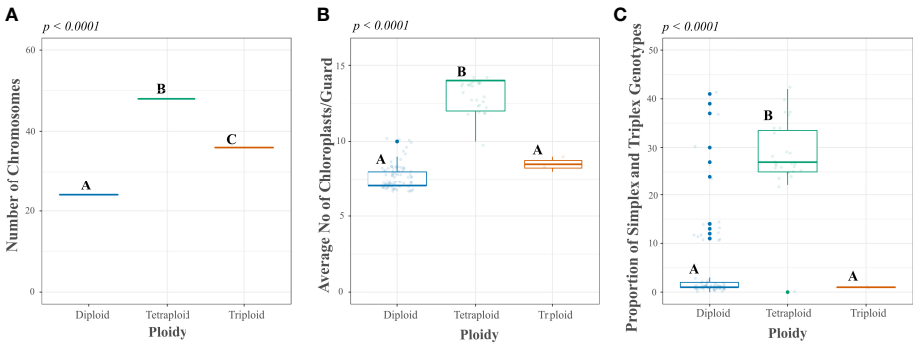


FIGURE 3
Comparison of direct and indirect methods of ploidy assessment in a reference set of accessions from the Colombian Central Collection (CCC) of potatoes. Y-axes show the measurement scales of direct and indirect methods: **(A)** chromosome count in the root tips, **(B)** chloroplast count per guard cell, and **(C)** proportion of simplex and triplex (ABBB, AAAB) Single Nucleotide Polymorphic (SNP) genotype scores per sample, compared to the x-axes with the ploidy base on the direct method of ploidy assessment. The *p-values* correspond to the ANOVA significance, while the different letters correspond to test results using a Tukey test in R software.

among repetitions, while for 12 accessions, some repetitions did not have the same results.

Core collection assembly

Using the Core Hunter 3 program, we identified three core collections for 10 (CCC_10), 15 (CCC_15), and 20 percent (CCC_20) of the total number of samples used in this study (i.e., 129, 194, and 258 entries, respectively), plus one mini-core collection with 3 percent (CCC_3) of the original size (39 entries). The accuracy of the A-NE algorithm used to identify diverse core collections was confirmed. Lower values of average A-NE distance were observed for the core and mini-core collections identified using the A-NE. In contrast, the same entries in the different core collections had an average distance near the optimum maximum values for the E-E or E-NE distance criteria (Figure 4). Similarly, as expected, the greater the collection size, the smaller the average A-NE distance, since each accession represents itself in the whole collection. Then, we compared the genetic diversity representation across collections identified using the A-NE method. We found no differences between the core and mini-core collections compared to the primary collection using the PCA and Ho for this comparison. The PCA showed that all the core and mini-core sizes represented a similar pattern of genetic diversity distribution (Figures 5A–D) comparable to the entire collection (Figures 1B, 2B). For the whole collection (CCC_100), the Ho mean was 0.33 (± 0.13), 0.36 (± 0.12) for the 20 percent (CCC_20) core collection, 0.38 (± 0.12) for the 15 percent (CCC_15), and 0.38 (± 0.13) for the 10 percent (CCC_10) core collections, while the Ho for the 3 percent (CCC_3) mini-core collection was 0.37 (± 0.13). In general, Ho mean was greater in the core collections (Ho = 0.36-0.38) compared to the whole CCC (Ho = 0.33), and the Ho density distribution from accessions in each collection follow similar patterns to the CCC and were contained within it (Figure 6A and Table 3). Therefore, we propose a 10 percent core collection for research and breeding in AGROSAVIA. The 10 percent core collection is a suitable size that captures the total diversity of the CCC and has a manageable number of accessions for research and breeding projects. The sample proportion in the three genetic groups (CCC_Group_A, CCC_Group_B1, and CCC_Group_B2) was 24 percent, 53 percent, and 24 percent for all genotyped samples, and 15 percent, 52 percent, and 33 percent for the 10 percent core collection.

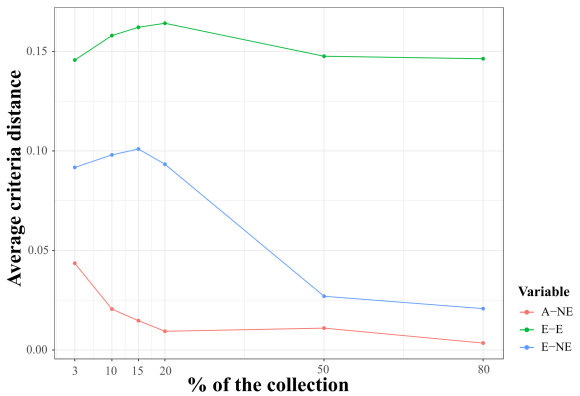


FIGURE 4
Qualitative evaluation of average Accession to Nearest Entry (A-NE) metrics used for diverse core collection identification and quality of core collections. The average criteria distance (y-axes) for three distance-based metrics A-NE, E-E (average Entry to Entry distance), and E-NE (average Entry to Neatest Entry distance) calculated using the core entries identified with A-NE criterion and single nucleotide (SNP) markers for six core collection sizes of 3, 10, 15, 20, 50, and 80 percent of whole samples genotyped in this study (x-axes).

Finally, we used the available phenotypic data of three agronomical traits (ATW, NTP, and TTY) for 846 genotyped accessions of the CCC to validate the representation of phenotypic variation in the core and mini-core collections. The validation strategy demonstrated that the 846 genotyped and phenotyped accessions had a similar distribution range of values for central and dispersion statistics in the whole collection (CCC_100) and across the different core (CCC_20-10) and mini-core (CCC_3) collections for the three different traits (Figures 6B–D and Table 3). In contrast to the values of genetic diversity in terms of Ho across collections, that showed a trend towards increasing from the whole collection to the smaller size of the core collection (Figure 6A), for the ATW trait, the means in the core collections decreased (CCC_20 = 18.27 g - CCC_3 = 15.54 g) with respect to the whole CCC (CCC_100 = 23.04 g) (Table 3). However, the NTP (CCC_20 = 11.59 tubers - CCC_3 = 14.99 tubers; Figure 6C and Table 3) and TTY (CCC_20 = 0.73 kg/plant - CCC_3 = 0.92 kg/plant; Figure 6D and Table 3) traits presented a similar tendency in mean values as Ho; here, the mean values in each of the core collections were greater than in the whole CCC (in CCC_100: NTP = 11.64 tubers and TTY = 0.77 kg/plant).

Moreover, we compared the behavior of three distance-based metrics used for core collection identification to validate —— using phenotypic data —— the quality of the core and mini-core collections identified with SNP markers and the A-NE metric (Figure 7). As observed for the genotypic data, small values were obtained for the average A-NE distance for the different core collection sizes. Even though the average distances for E-NE and EE metrics were greater than A-NE, they did not show a strong maximization of values for small collection sizes. As expected, the average A-NE values decreased in the same way that the core size was closer to the full collection size.

Core collection description

The 10 percent core and 3 percent mini-core collections contain accessions from different taxonomic groups and geographic origins (Supplementary Table S1). Of the 129 core collection accessions contained in the 10 percent collection, 82 accessions are *S. tuberosum* subsp. *andigenum*, 18 are *S. phureja*, 17 are *S. tuberosum* subsp. *tuberosum*, and 12 are currently undergoing curation. This 10 percent core collection contains 70 accessions from ten collection sites in Colombia, the larger representation from Nariño (22), Boyacá (14), and Cundinamarca (14)

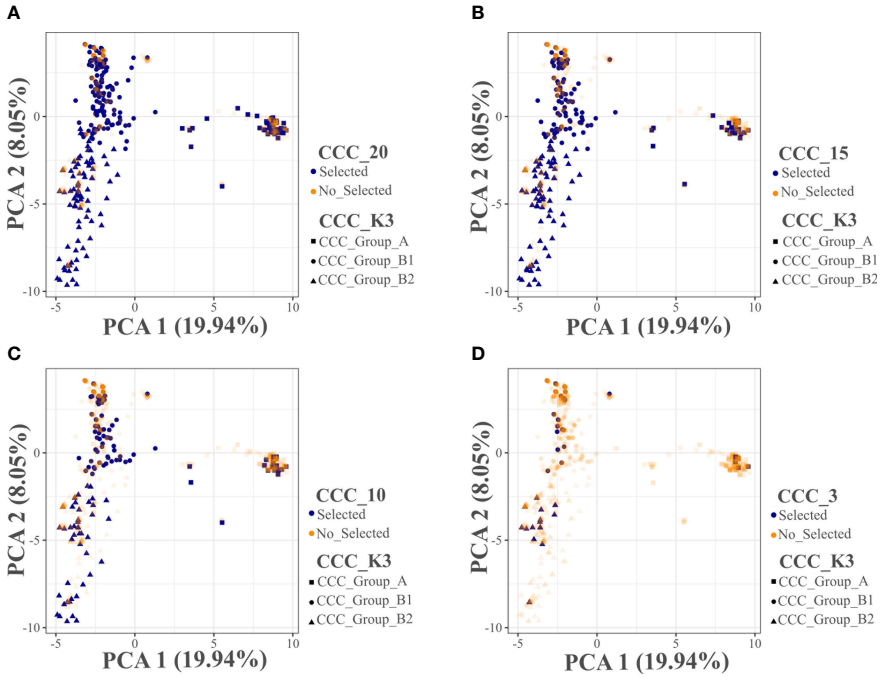


FIGURE 5
Principal Component Analyses (PCA) distribution of the selected entries in three core and one mini-core collections evaluated for the Colombian Central Collection (CCC) of potatoes. **(A)** Core collection size of 20 percent (CCC_20), **(B)** Core collection size of 15 percent (CCC_15), **(C)** Core collection size of 10 percent (CCC_10), and **(D)** Mini-core collection size of 3 percent (CCC_3). The CCC_K3 information in **(A–D)** corresponds to the three genetic groups suggested by the CCC population structure analysis using the NbCluster R-package.

departments, followed by Cauca (6), Antioquia (5), Valle del Cauca (5), and one accession from Caldas, Quindío, Santander, and Tolima. Moreover, the core collection has accessions from Peru (7), Bolivia (3), Ecuador (2), Mexico (2), the USA (2), the Netherlands (2), and Germany (2). Thirty-nine of these accessions are of unknown origin. Of the 39 accessions in the mini-core collection (3 percent), 25 are *S. tuberosum* subsp. *andigenum*, six are *S. phureja*, two are *S. tuberosum* subsp.

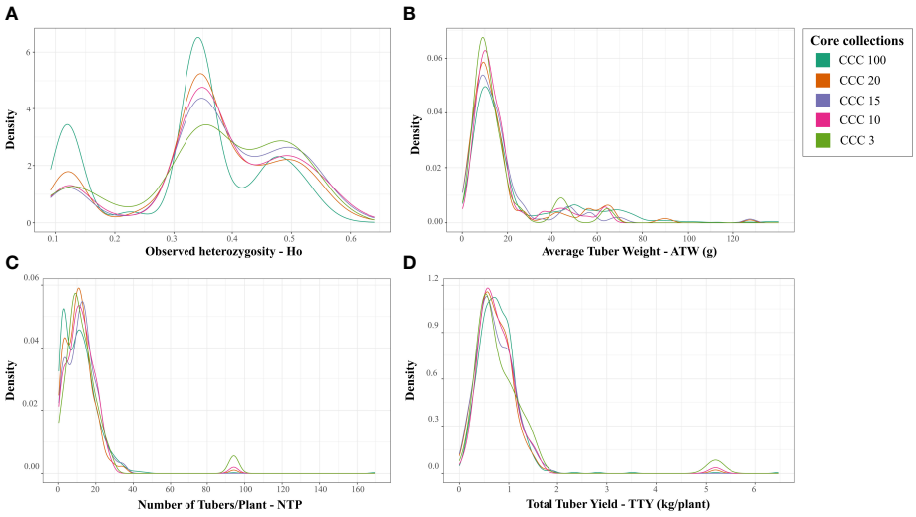


FIGURE 6
Density distribution of the genetic diversity and three different agronomic traits values of samples compared across main and proposed core and mini-core collections for the Colombian Central Collection (CCC) of potatoes. Distribution of values for **(A)** observed heterozygosity (Ho), **(B)** Average Tuber Weight in g (ATW), **(C)** Number of Tubers per Plant (NTP), and **(D)** Total Tuber Yield in kg/plant (TTY) for the whole CCC (CCC_100), and three core collection sizes of 20 percent (CCC_20), 15 percent (CCC_15), and 10 percent (CCC_10), and one mini-core collection size of 3 percent (CCC_3).

TABLE 3 Measures of central tendency (mean and median) and dispersion [coefficient of variation (CV), standard deviations (Std Dev), minimal (min) and maximum (Max) values] for the observed Heterozygosity (Ho), Average Tuber Weight (ATW), Number of Tubers per Plant (NTP), and Total Tuber Yield (TTY) traits evaluated in the Colombian Central Collection (CCC_100) of potatoes and Core Collections sizes of 20 percent (CCC_20), 15 percent (CCC_15), 10 percent (CCC_10) and 3 percent (CCC_3) of the CCC proposed in this study.

Core Collection	NGGD	NGPD	Measure	Ho	ATW in g	NTP	TTY in kg/plant
CCC_100	1291	846	CV	39.91	100.94	88.7	55.37
			Mean	0.33	23.04	11.64	0.77
			Std dev	0.13	23.26	10.21	0.43
			Median	0.34	13.65	10.32	0.73
			Min	0.09	0.17	0.35	0.0
			Max	0.64	140.0	169.69	6.47
CCC_20	258	172	CV	34.42	108.94	80.59	64.8
			Mean	0.36	18.27	11.59	0.73
			Std dev	0.12	19.91	9.34	0.47
			Median	0.35	11.25	10.58	0.67
			Min	0.09	0.17	0.48	0.02
			Max	0.64	127.5	94.23	5.19
CCC_15	194	115	CV	32.39	103.01	59.89	48.8
			Mean	0.38	16.91	12.51	0.73
			Std dev	0.12	17.42	7.49	0.35
			Median	0.37	12.07	12.21	0.69
			Min	0.1	0.5	0.82	0.08
			Max	0.64	127.5	34.48	1.77
CCC_10	129	81	CV	33.35	85.58	87.53	73.32
			Mean	0.38	16.82	12.9	0.8
			Std dev	0.13	14.4	11.29	0.59
			Median	0.38	16.82	12.9	0.8
			Min	0.1	1.45	0.82	0.1
			Max	0.64	64.5	94.23	5.19
CCC_3	39	26	CV	34.84	91.64	116.24	102.07
			Mean	0.37	15.54	14.99	0.92
			Std dev	0.13	14.24	17.43	0.94
			Median	0.37	15.54	14.99	0.92
			Min	0.12	3.67	1.24	0.21
			Max	0.56	64.5	94.23	5.19

NGGD, Number of Genotypes with Genotypic Data; NGPD, Number of Genotypes with Phenotypic Data.

tuberosum, and six have yet to be identified and are in the process of curation. Regarding their origins, the mini-core collection has 22 accessions from Colombia collected from seven departments: Nariño, Cundinamarca, Boyacá, Antioquia, Cauca, Quindío, and Valle del Cauca. The mini-core collection has four accessions from other countries, two from Peru, one from Mexico, and one from the Netherlands. Finally, the mini-core collection has 13 accessions of unknown origin.

Discussion

Our study genotyped 90.9 percent (1141/1255) of the clonal collection of the CCC of potatoes and a representation of released varieties. These genomic data allowed us to study the population structure and genetic diversity, determine ploidy, and define core and mini-core collections representing the genetic diversity of the whole CCC. Below we discuss our study results and how we can use

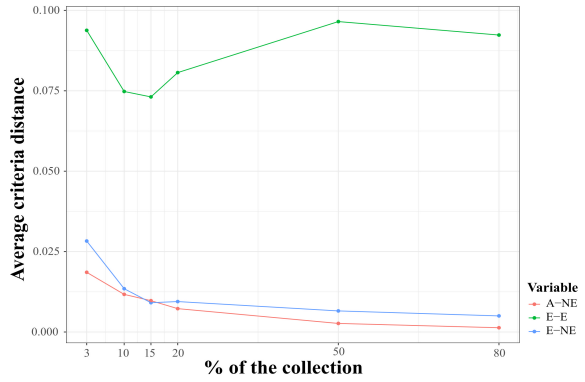


FIGURE 7
Evaluation of core quality using three different distance-based metrics for core identification and phenotypic data. The average criteria distance (y-axes) for three distance-based metrics A-NE (average Accession to Nearest Entry distance), E-E (average Entry to Entry distance), and E-NE (average Entry to Neatest Entry distance) calculated using phenotypic distance among core entries identified with A-NE criterion and genomic markers for six core collection sizes: 3, 10, 15, 20, 50 and 80 percent of whole samples genotyped in this study (x-axes).

them to curate and improve the Colombian potato collection in agreement with the Global Conservation Strategy for potato (Nagel et al., 2022).

Population structure and genetic diversity

The pattern of population genetic structure and diversity observed suggests that potato accessions held at the CCC fit the evolutionary patterns of farming selection that separated diploid from tetraploid landraces, and this from the most-modern cultivated potato (*Solanum tuberosum* L.). According to our previous study that used 809 accessions (Berdugo-Cely et al., 2017), the CCC has a population structure that contains two main genetic groups of mostly diploid and tetraploid accessions represented in general by the Phureja and Andigenum types, respectively (CCC_Group_A and CCC_Group_B). This study corroborated this pattern, but separated the tetraploid genetic group into two well-differentiated genetic groups (CCC_Group_B1 and CCC_Group_B2). The CCC_Group_B1 has a greater number of accessions and levels of admixture than the two other genetic groups. Most accessions are of Andigenum type, probably with more primitive ancestry than CCC_Group_B2. Thus, the Andigenum accessions that resulted in a more differentiated and divergent CCC_Group_B2 genetic group probably come from a more recent selection pressure by farmers and breeders for some modern traits. A higher number of Tuberosum accessions were also in CCC_Group_B2 group, supporting the hypothesis that this group suffered substantial selection pressure. This concurs with the fact that most of the Tuberosum accessions in the CCC are modern varieties acquired as part of potato breeding efforts in Colombia (Moreno and Valbuena, 2006). Differences in the agricultural selection patterns have been reported in potatoes. The modern North American and traditional Andean cultivars shared only 14–16 percent of genes under selection; thus, the adaptation of highland Andigenum and lowland Tuberosum had different selection strategies (Hardigan et al., 2017). Besides this conclusion, we consider that modern breeding practices in North America and Europe also strengthened the differentiation between most modern/commercialized potato and Andean landraces. The CCC_Group_B1 and CCC_Group_B2 groups also showed divergence from Phureja, probably due to polyploidization, a more comprehensive distribution range, and the selection pressure for some culinary or crop traits. In general, potatoes have been selected by ancient farmers and, more recently, by breeders for tuber size, tuber carbohydrates and glycoalkaloids content, photoperiod adaptation, and reduced sexual fertility (Hardigan et al., 2017). The CCC diversity study results support divergence between diploid and tetraploid accessions from Phureja, Andigenum, and Tuberosum types, probably mediated by selection patterns.

The CCC is one of the Latin American clonal native potato collections that can be compared to the International Potato Center (CIP) collection in Peru, with around 4,500 landrace accessions (Ellis et al., 2020). Genetic diversity analysis of part of CIP's potato collection agreed with our analysis results for the patterns of genetic differentiation of the Phureja and Andigenum groups observed in the CCC. Thus, Andigenum has a more significant number of accessions, genetic admixture, and genetic diversity. In contrast, Phureja is more genetically homogeneous with fewer accessions (Ellis et al., 2018). As documented by different authors, the genetic diversity of native tetraploid potatoes from the Andigenum group has broader geographic coverage and diversity compared to the Phureja group distributed in a more restricted area (Hawkes, 1990; Huamán and Spooner, 2002; Spooner et al., 2010; Bradshaw, 2021), which suggests that polyploidy gave advantages of yield, robustness, and adaptation to different environments in the Andean Highlands. Although the Phureja diploids are also derived from *S. stenotomum*, they were selected due to a lack of tuber dormancy, shorter crop cycles, and adaptability to warmer, lower, and eastern valleys of the Andes. Besides the less cosmopolitan adoption of diploid Phureja, some levels of self-compatibility could also have contributed to the genetic homogeneity of this group. While the self-incompatibility nature of diploid potatoes is widely acknowledged, recent reports have revealed the existence of self-compatibility sources in cultivated potatoes (Kaiser et al., 2021). Notably, the Phureja clone 1S1 has been identified as a source of self-compatibility, leading us to the assumption that self-compatibility contributes to the genetic diversity of Phureja (ibid). In summary, both the CIP and CCC potato collections exhibited similar patterns of genetic diversity differentiation for Phureja and Andigenum accessions.

In different population structure and genetic diversity studies of genebank collections of traditional cultivars and breeding programs germplasm of potato, the material has been discriminated following ploidy, taxonomy, and characteristics of breeding selection (Hirsch et al., 2013; Berdugo-Cely et al., 2017; Hardigan et al., 2017; Deperi et al., 2018; Ellis et al., 2018; Pandey et al., 2021). In general, it is observed that the genetic diversity studies of germplasm obtained from genebanks have shown that the discrimination among accessions is mainly associated with ploidy and taxonomy (Berdugo-Cely et al., 2017; Hardigan et al., 2017; Ellis et al., 2018; Berdugo-Cely et al., 2021). This discrimination agrees with our results, in which ploidy and taxonomy explained the identified

genetic groups. In contrast, genetic diversity studies of germplasm from breeding programs have shown that discrimination of accessions depends on potato market class selection, indicating that the diversity is shaped by the strong selection pressure for breeding traits (Hirsch et al., 2013; Deperi et al., 2018; Pandey et al., 2021). Overall, the studies on population structure and genetic diversity in potato showed that trait introgression from wild germplasm and the selection pressure of breeding contributed to increasing heterozygosity and diversification of cultivated lineages.

Ploidy assessment

This study used the frequency of simplex and triplex scores of all SNPs in each accession as a ploidy level predictor, following Alsahlany et al. (2019). This methodology is possible because the SNP genotyping using the Illumina array technology allows the dosage of biallelic markers in an autotetraploid species such as a potato to be identified. Alsahlany et al. (2019) proposed assessing the ploidy based on the frequency of simplex and triplex SNP scores (ABBB, AAAB), in which one allele would be present in a higher proportion than the other. This approach is objective because diploid accessions would only have two alleles, giving the same proportion for each; thus, there is a low or null probability of detecting a simplex or triplex SNP score in a diploid. Paralogous genes or mismatches between the sample DNA and the SNP probe could explain the detection of simplex or triplex scores in diploids, but this has a low probability of occurring.

In this study, the frequency of simplex and triplex SNP scores identified in the reference set of samples varied compared to Alsahlany et al. (2019); tetraploids have a proportion of 27.1 percent ±1.53 compared to 36 percent ±14, and diploids 4.8 percent ±9.04 compared to 2 percent ±1. The separation threshold to predict our samples' ploidy was <15 percent for diploids and >20 percent for tetraploids. Accessions with frequencies between 15 and 20 percent were considered unknown; this classification was tight compared with the ample separation reported by Alsahlany et al. (2019). The simplex/triplex frequency was up to 4 percent for most diploid samples (263/297, 88.6 percent), and a few samples were 9–15 percent (34/297, 11.4 percent). The genetic diversity of this collection probably interfered with the DNA hybridization during SNP genotyping for those samples and affected the proportion of simplex and triplex SNP genotype scores. In contrast, Alsahlany et al. (2019) used North American breeding germplasm similar to the varieties used to develop the potato SNP array (Hamilton et al., 2011). Even though for a few samples, the parameters for classification used to assess the ploidy of accessions in the CCC based on the frequency of simplex and triplex SNP scores were tight, this approach helped estimate the ploidy in the CCC.

Heterozygosity values estimated for potato samples genotyped with the Infinium Potato Array have also been related to ploidy levels. Ellis et al. (2018) reported that each ploidy group of a panel of 500 samples of potato landraces was associated with a percentage of heterozygosity pattern. However, they also detected that heterozygosity overlaps in narrow ranges making the separation difficult. They found higher heterozygosity in tetraploid germplasm (>30 percent) compared to diploids (<20 percent), and triploids have 20–30 percent of heterozygosity. In the present study, we found that the ranges of heterozygosity for samples in the reference set, excluding potential errors, were 10–23 percent for diploids, 13 percent for triploids, and 33–58 percent for tetraploids. However, there was a continuous data distribution of 9–66 percent heterozygosity for the complete samples with no clear separation. We expected mainly diploids, tetraploids, and a low number of triploids in the CCC (Figures 2C, 3C). These results showed that the separation thresholds should be defined for each case based on the germplasm evaluated. Similarly, the frequency of simplex and triplex scores of all SNPs in one accession is a much better approach to predicting ploidy level than the heterozygosity values.

A reference set of samples with ploidy information generated by other indirect (chloroplast count per guard cell or flow cytometry) and/or direct (chromosome count) methods would better define the threshold separation in ploidy assessment. In our case, we compared data from SNP scores, the number of chloroplast/guard cells, and chromosome counts. The threshold separation for the number of chloroplast/guard cells (6–8 for diploids and 12–14 for tetraploids) was consistent with previous reports (Rasmussen and Rasmussen, 1995; Gebhardt et al., 2006; Ordoñez et al., 2014; Alsahlany et al., 2019). Although assessing the ploidy using only SNP scores is valid, we recommend confirming this assessment with a low-cost and time-efficient technique such as the chloroplast count in guard cells.

Core collection

We proposed the 10 percent subset as the core collection because it already captured the representation of the total diversity of the CCC. The reduced but highly diverse core collection has a more manageable and versatile number of accessions for trait discovery, evaluations, and use in Colombia's potato breeding and crop improvement. The representation of phenotypic diversity for three agronomic traits (ATW, NTP, and TTY) within the core and mini-core collections confirmed their potential use. In general, the evaluation of the CCC for sources of new alleles for different traits has taken several years because of the size of this collection. Now, faster screening can be conducted, and the mini-core collection, which is even smaller in size, represents an alternative when evaluating the core collection is not feasible.

As expected, after eliminating redundancy in the collection, the genetic diversity measured using the Ho average increased in the core and mini-core collections. Similar results have been reported for recently developed core collections of potato (Bamberg and del Rio, 2004; Pandey et al., 2021). The PCA analysis also showed that the sample dispersion in the scatter plots has a similar distribution across core collection sizes. Likewise, our behavior analysis of the average distance-based metrics A-NE, EE, and E-NE used to identify core and mini-core collections confirmed that the A-NE metric is robust for obtaining a uniform representation of the original genetic space, as proposed by Odong et al. (2013). A-NE algorithm aims to obtain a small average distance between accession

and the nearest entry in the core for the whole collection; this guarantees the representativity of all accessions in the core collection. By contrast, the E-E and E-NE methods maximize the average genetic distances between entries and entry to nearest entry to identify core collections representing extreme values or a distribution, respectively. The average A-NE also decreased as the collection size increased because, as expected, the maximum representation of each accession by an entry is itself. The identified core and mini-core collections followed the expected average A-NE performance using phenotypic data.

The selected 10 percent core collection size was within the range of 5–20 percent core collection size reported and corresponded with the most recommended size (Van Hintum et al., 2000). We will present this core collection to breeders and scientists in AGROSAVIA to obtain their feedback and contributions to ensure the selected entries are the most suitable for research and breeding. If needed, the accessions of research and breeding interest or with long historical evaluation records could be added to the core collection by replacing initially selected entries with those that have the most similar genetic distance.

Practical use of diversity and ploidy assessment in the management of CCC

The CCC has been curated with the support of diversity studies and ploidy assessment. The clonal CCC of potatoes has been under AGROSAVIA´s management since 1994. At that time, the collection handled by the former curator categorized the accessions into different groups: Phureja, Andigena, Tuberosum, "Guata," "Chaucha," and unknown. This classification is based on a combination of morphological patterns specific to Andigenum, Phureja, and Tuberosum taxonomic classification, and a non-taxonomic, ethnobotanical classification for certain native accessions or breeding germplasm, referred to as "Guata," and "Chaucha". This AGROSAVIA categorization has been adjusted for each accession during the CCC curation process to assign and/or verify the classification according to the taxonomic classification of Hawkes (1990). Initially, the collection was morphologically characterized to document the discriminatory descriptors in the CCC and the characteristics of each accession. Then, the digitalization of passport data with the originally assigned taxonomic classification information, genetic diversity analysis, and ploidy assessment supported curation of the classification. In previous years, morphological descriptors and genetic diversity analysis were used to reassign some Chaucha accessions into the Phureja group. For the accessions evaluated in this study, we found that the Guata classification corresponds to accessions that are mostly tetraploids (83/88, 94,3 percent) and, based on a few recovered passport data (23/88), the accessions also correspond to *S. tuberosum* subsp. *andigenum*. The assignment to genetic groups, ploidy, morphology, and historical records should be considered when reassigning taxonomic classification. Some diploid accessions can be Andigenum, and some tetraploid accessions Phureja. In potatoes, the germplasm ploidy can be increased due to crosses involving 2n gametes (polyploidization) or reduced by pickle pollination (haploidization). Phureja accessions are mainly diploids, but triploids and tetraploids can be found in a low proportion (Ghislain et al., 2006). Similarly, Andigena accessions are mainly tetraploids; however, di-haploids could have been generated as part of breeding program strategies. Thus, the morphological description, passport data, analysis of diversity, and ploidy assessment need to be continued to support the curation of this collection.

In Colombia and Ecuador, indigenous communities and traditional farmers categorize the native cultivars between Chauchas and Guatas. In general, cultivars with a short crop cycle and early sprouting potato are known as Chauchas, which means *soft or easy*, and cultivars with an annual crop cycle and white tuber flesh color as Guatas, a Quechua word meaning *potato of one year* (Monteros-Altamirano, 2017; Rosero Alpala et al., 2020). From a taxonomic perspective, the indigenous or ethnobotanical classification of Chaucha contains germplasm that mainly corresponds to the formerly accepted classification by Hawkes (1990) of *S. phureja* and in lesser proportion to *S. chaucha*, while the Guatas are mainly associated with the *S. tuberosum* subsp. *andigenum* (Navarro et al., 2010; Monteros-Altamirano, 2017; Rosero Alpala et al., 2020). This homologation using ethnobotanical and taxonomic classification is fully aligned with the results obtained in the curation of the CCC collection (Figure 1A), thus this type of information could support taxonomic classification.

Part of the curation process is keeping all the information related to the germplasm organized by attributes or categories. The database should contain information on common names, taxonomic classification based on different authors (Hawkes, Spooner), biological status (wild, native/landrace, and breeding \research germplasm), and ethnobotanical classification (Chauchas and Guatas). Regarding the different taxonomic classifications, it is important to maintain the information of Hawkes' classification, which facilitates the management and use of germplasm in the conservation and breeding programs. Even though the classification based on Spooner et al. (2007) that regroups four former species (*S. phureja*, *S. tuberosum* subsp. *andigenum*, *S. stenotomum*, and *S. chaucha*) into a single species is well supported, a suite of morphological descriptors, length of plant cycle, photoperiod adaptation, and cytoplasm type support their differentiation (Hawkes, 1990; Spooner et al., 2014). If this distinction is unclear, scientists and breeders can encounter issues with fertility, dormancy, tuber attributes, tuberization, and maturity traits.

Future perspectives for CCC management

This study successfully genotyped 90.9 percent of the clonal collection of the CCC of potatoes and a representation of released varieties. This result is an outstanding achievement considering this collection's size (2,499 wild and cultivated potato accessions) and its importance for the region. Moreover, we are interested that this collection become in a regional reference. Therefore, we would recommend aligning the CCC conservation strategy with the ten actions suggested for improving the global-scale conservation of potato resources, as proposed by several institutions and

organizations (Nagel et al., 2022). Some of the actions that can be taken after this study include finding duplicates, clarifying the possible accession admixture exposed in this study, determining the ploidy using the chloroplast count in guard cells, verifying the taxonomy and biological status for accessions in the curation process, completing the digitalization of documentation and centralization of all historical and new evaluation data in the GRIN-Global platform, prioritize future evaluations in field experiments and data collection using the core collection, and completing the evaluation of core accessions that have missing information from previous biotic and abiotic evaluations. This collection's passport data, morphological characterization, and genetic profiling can be used to compare the *in vitro* and field collection to verify the identity of germplasm. Ellis et al. (2018) showed that some accession admixing could happen over time in the genebanks. Therefore, through an identity verification process, the collection can be revised, documented, and organized to correct errors and mitigate future mistakes by digitalizing data and implementing barcoding and quality management standards developed for genebanks.

The genotyping data generated in this study can also be compared with other collections, such as CIP's collection, genotyped using the same technology. This comparison could serve to identify signatures of selection, materials with potential use in breeding, duplicity, genetic gaps, homologate accessions between the collections, and assess the extent to which CCC diversity is representative of the world's largest collection of native potatoes. This paper is also available in Spanish to continue the discussion of this topic within Colombia and more broadly in the region (Supplementary Data Sheet 1).

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

RY, JB-C, IC-S, and PR-H conceived the study. JB-C, ZL-P, and PR-H organized the CCC databases and designed the analysis. ZL-P and IC-S collected leaves from selected accessions. JB-C and IC-S worked in the wet lab. JB-C and PR-H performed the analysis, including descriptive graphics. NM-C led the manuscript writing and performed part of the analyses. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by the Ministerio de Agricultura y Desarrollo Rural of Colombia under the funds TV19-21 in the project "Diseño e implementación de un Plataforma de genotipificación para el Banco de Germoplasma Vegetal de Colombia conservado por AGROSAVIA" with 1001386 code.

Acknowledgments

We especially thank L.P. Delgadillo for supporting DNA extractions for a part of the potato accessions used in this study. Thanks to B. Coronel and R.T. Torres for collecting leaves from some accessions of the CCC in the field. We also thank R.T. Torres for locating and organizing all the CCC passport data using the oldest books, systematized by hand. We would also like to thank M.S. Cerón for sharing the studies about the ploidy for the CCC. We thank the Molecular Laboratory at C.I. Tibaitatá of AGROSAVIA for their support in genotyping the CCC accessions. Finally, we thank the former curator, Ivan Valbuena, for his dedication to the CCC that motivated this work and Olga Spellman, Science Writer, Editor and Coordinator of the Alliance Science Writing Service (Alliance of Bioversity International and CIAT) for English and copy editing of this manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1046400/full#supplementary-material

DATA SHEET 1
Contains translated material.

SUPPLEMENTARY FIGURE 1
Number of genetic clusters suggested by **(A)** STRUCTURE Harvester **(B)** and NBClust R-Package test for 1,291 genotyped samples of the Colombian Central Collection (CCC) of potatoes and breeding material using 3,586 Single Nucleotide Polymorphism (SNP) markers.

SUPPLEMENTARY TABLE 1
List of potato genotype accessions of the Colombian Central Collection (CCC) and breeding material analyzed in this study.

SUPPLEMENTARY TABLE 2
Genotype matrix that included the 3,586 genome-wide single nucleotide polymorphic (SNP) markers identified and filtered for 1,291 potato samples of the Colombian Central Collection (CCC) and breeding material analyzed in this study.

References

Alsahlany, M., Zarka, D., Coombs, J., and Douches, D. S. (2019). Comparison of methods to distinguish diploid and tetraploid potato in applied diploid breeding. *Am. J. Potato Res.* 96, 244–254. doi: 10.1007/S12230-018-09710-7

Aravind, J., Kaur, V., Wankhede, D. P., and Nanjundan, J. (2022)EvaluateCore: Quality evaluation of core collections. In: *R package version 0.1.3.9000*. Available at: https://aravind-j.github.io/EvaluateCore/authors.html#citation (Accessed 30, 2022).

Bamberg, J. B., and del Rio, A. H. (2004). Genetic heterogeneity estimated by RAPD polymorphism of four tuber-bearing potato species differing by breeding system. *Amer J of Potato Res* 81, 377–383. doi: 10.1007/BF02870198

Berdugo-Cely, J., Martínez-Moncayo, C., and Lagos-Burbano, T. (2021). Genetic analysis of a potato (Solanum tuberosum l.) breeding collection for southern Colombia using single nucleotide polymorphism (SNP) markers. *PLoS ONE* 16 (3). doi: 10.1371/journal.pone.0248787

Berdugo-Cely, J., Valbuena B, R. I., Sánchez-Betancourt, E., Stella Barrero, L., and Yockteng, R. (2017). Genetic diversity and association mapping in the Colombian central collection of solanum tuberosum l. andigenum group using SNPs markers. *PLoS ONE* 12 (3). doi: 10.1371/journal.pone.0173039

Bradshaw, J. E. (2021). Domestication to twenty-first-century potato cultivars. *Potato Breeding: Theory Pract.* (Cham: Springer), 3–51. doi: 10.1007/978-3-030-64414-7_1

Brown, A. H. D. (1989). Core collections: A practical approach to genetic resources management. *Genome*, 818–824. doi: 10.1139/g89-144

Chapman, R. L., and Crawford, D. J. (1990). Plant molecular systematics: macromolecular approaches. *undefined* 98, 635. doi: 10.2307/3243598

Charrad, M., Ghazzali, N., Laval, U., and Niknafs, A. (2014) *NbClust: An r package for determining the relevant number of clusters in a data set véronique boiteau*. Available at: http://www.jstatsoft.org/.

de Beukelaer, H., Davenport, G. F., and Fack, V. (2018). Core hunter 3: Flexible core subset selection. *BMC Bioinf.* 19. doi: 10.1186/s12859-018-2209-z

Deperi, S. I., Tagliotti, M. E., Bedogni, M. C., Manrique-Carpintero, N. C., Coombs, J., Zhang, R., et al. (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, e0194398. doi: 10.1371/JOURNAL.PONE.0194398

Díaz-Valencia, P., Melgarejo, L. M., Arcila, I., Mosquera-Vásquez, T., and Co, I. A. (2021). Physiological, biochemical and yield-component responses of solanum tuberosum l. group phureja genotypes to a water deficit. *Plants (Basel)* 10 (4). doi: 10.3390/plants10040638

Dodds, K. S., and Paxman, G. J. (1962). The genetic system of cultivated diploid potatoes. *Evol. (N Y)* 16, 154. doi: 10.2307/2406194

Earl, D. A., and vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the evanno method. *Conserv. Genet. Resour* 4, 359–361. doi: 10.1007/S12686-011-9548-7

Ellis, D., Chavez, O., Coombs, J., Soto, J., Gomez, R., Douches, D., et al. (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. doi: 10.1139/gen-2017-0201

Ellis, D., Salas, A., Chavez, O., Gomez, R., and Anglin, N. (2020). "Ex situ conservation of potato [Solanum section petota (solanaceae)] genetic resources in genebanks," in *The potato crop: Its agricultural, nutritional and social contribution to humankind* (Springer International Publishing), 109–138. doi: 10.1007/978-3-030-28683-5_4/FIGURES/5

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x

Felcher, K. J., Coombs, J. J., Massa, A. N., Hansey, C. N., Hamilton, J. P., Veilleux, R. E., et al. (2012). Integration of two diploid potato linkage maps with the potato genome sequence. *PLoS One* 7, 36347. doi: 10.1371/JOURNAL.PONE.0036347

Frankel, O. H. (1984). Genetic perspectives of germplasm conservation. *Genetic Manipulation: Impact on Man and Society*, 161–170. doi: 10.3/JQUERY-UI.JS

Gebhardt, C., Bellin, D., Henselewski, H., Lehmann, W., Schwarzfischer, J., and Valkonen, J. P. T. (2006). Marker-assisted combination of major genes for pathogen resistance in potato. *Theor. Appl. Genet.* 112, 1458–1464. doi: 10.1007/S00122-006-0248-8

Ghislain, M., Andrade, D., Rodríguez, F., Hijmans, R. J., and Spooner, D. M. (2006). Genetic analysis of the cultivated potato solanum tuberosum l. phureja group using RAPDs and nuclear SSRs. *Theor. Appl. Genet.* 113, 1515–1527. doi: 10.1007/S00122-006-0399-7

Guevara, R. (2011). Determinación y comprobación del nivel de ploidía y conteo de cromosomas en cincuenta accesiones de papa chaucha (Solanum tuberosum grupo chaucha), procedentes del banco de germoplasma vegetal que administra CORPOICA. (Tesis de pregrado, Universidad de Cundinamarca).

Hamilton, J. P., Hansey, C. N., Whitty, B. R., Stoffel, K., Massa, A. N., van Deynze, A., et al. (2011). Single nucleotide polymorphism discovery in elite north American potato germplasm. *BMC Genomics* 12, 1–12. doi: 10.1186/1471-2164-12-302/TABLES/5

Hardigan, M. A., Laimbeer, F. P. E., Newton, L., Crisovan, E., Hamilton, J. P., Vaillancourt, B., et al. (2017). Genome diversity of tuber-bearing solanum uncovers complex evolutionary history and targets of domestication in the cultivated potato. *Proc. Natl. Acad. Sci. USA* 114, E9999–E10008. doi: 10.1073/PNAS.1714380114/SUPPL_FILE/PNAS.1714380114.SD11.XLSX

Hawkes, J. G. (1990). The potato: evolution, biodiversity and genetic resources. (London, UK: Belhaven Press).

Hirsch, C. N., Hirsch, C. D., Felcher, K., Coombs, J., Zarka, D., van Deynze, A., et al. (2013). Retrospective view of north American potato (Solanum tuberosum l.) breeding in the 20th and 21st centuries. *G3: Genes Genomes Genet.* 3, 1003–1013. doi: 10.1534/G3.113.005595/-/DC1

Huamán, Z., and Spooner, D. M. (2002). Reclassification of landrace populations of cultivated potatoes (Solanum sect. petota). *Am. J. Bot.* 89, 947–965. doi: 10.3732/AJB.89.6.947

Jombart, T., and Ahmed, I. (2011). Adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinf. Appl. NOTE* 27, 3070–3071. doi: 10.1093/bioinformatics/btr521

Juyó, D., Sarmiento, F., Álvarez, M., Brochero, H., Gebhardt, C., Mosquera, T., et al. (2015). Genetic diversity and population structure in diploid potatoes of solanum tuberosum group phureja. *Crop Sci.* 55, 760–769. doi: 10.2135/CROPSCI2014.07.0524

Kaiser, N. R., Jansky, S., Coombs, J. J., Collins, P., Alsahlany, M., and Douches, D. S. (2021). Assessing the contribution of sli to self-compatibility in north American diploid potato germplasm using KASP^{TM} markers. *Am. J. Potato Res.* 98, 104–113. doi: 10.1007/S12230-021-09821-8

Kamvar, Z. N., Tabima, J. F., and Grünwald, N. J. (2014). Poppr: an r package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *Peer J* 4 (2). doi: 10.7717/peerj.281

Luján, J. (1976) *Evolución del cultivo de papa en Colombia*. Available at: http://hdl.handle.net/20.500.12324/23444.

Mijangos, J. L., Gruber, B., Berry, O., Pacioni, C., and Georges, A. (2022). dartR v2: an accessible genetic analysis platform for conservation, ecology, and agriculture. *Methods Ecol. Evol* 13 (10), 215–2158. doi: 10.1111/2041-210x.13918

Ministerio de Agricultura y Desarrollo Rural (2020) *Cadena de la papa*. Available at: https://sioc.minagricultura.gov.co/Papa/Documentos/2020-06-30%20Cifras%20Sectoriales.pdf (Accessed 28, 2022).

Monteros-Altamirano, A. (2017). Ecuadorian Potato landraces trad nam and genetic ident. *Rev. Fitotec* 40, 481–489. Available at: http://repositorio.iniap.gob.ec/handle/41000/4797 [Accessed March 28, 2023].

Moreno, J. D., and Valbuena, R. I. (2006). Colección central colombiana de papa : riqueza de variabilidad genética para el mejoramiento del cultivo. *Rev. Innovación Cambio Tecnológico* 4, 16–24. Available at: http://hdl.handle.net/20.500.12324/563 [Accessed August 29, 2022].

Nagel, M., Dulloo, E., Bissessur, P., Gavrilenko, T., Ellis, D., and Peter, G. (2022). *Global strategy for the conservation of potato* (Bonn, Germany: Global Crop Diversity Trust).

Navarro, C., Carlos Bolaños, L., César Lagos, T., and Nariño, D. (2010). Caracterización morfoagronómica y molecular de 19 genotipos de papa guata y chaucha (Solanum tuberosum l. y solanum phureja juz et buk) cultivados en el departamento de nariño. *Rev. Cienc. Agrícolas* 27, 27–39. Available at: https://dialnet.unirioja.es/servlet/articulo?codigo=5104152&info=resumen&idioma=SPA [Accessed September 1, 2022].

Ñustez, C. E. (2011) *Variedades colombianas de papa*. Available at: https://es.scribd.com/document/423589454/Variedades-colombianas-de-papa-Carlos-Eduardo-Nustez-Lopez (Accessed 28, 2022).

Odong, T. L., Jansen, J., van Eeuwijk, F. A., and van Hintum, T. J. L. (2013). Quality of core collections for effective utilisation of genetic resources review, discussion and interpretation. *Theor. Appl. Genet.* 126, 289–305. doi: 10.1007/S00122-012-1971-Y/FIGURES/9

Ordoñez, B., Mihovilovich, E., and Bonierbale, M. (2014). Altering self-incompatibility using sli overcomes reproductive barriers between cultivated diploid potato and wild species of piurana series. in *Embrapa*. Federal University of Lavras (UFLA) State University of Santa Cruz (UESC) International Consortium in Advanced Biology (CIBA) Solanaceae: Book of abstracts., 90. Available at: http://www.sol2014.com.br/images/stories/book_of_abstracts-2014.pdf [Accessed August 30, 2022].

Pandey, J., Scheuring, D. C., Koym, J. W., Coombs, J., Novy, R. G., Thompson, A. L., et al. (2021). Genetic diversity and population structure of advanced clones selected over forty years by a potato breeding program in the USA. *Sci. Rep.* 11, 1–18. doi: 10.1038/s41598-021-87284-x

Pavoine, S., Vallet, J., Dufour, A. B., Gachet, S., and Daniel, H. (2009). On the challenge of treating various types of variables: Application for improving the measurement of functional diversity. *Oikos* 118, 391–402. doi: 10.1111/J.1600-0706.2008.16668.X

Pineda Colorado, R., and Hernández Castillo, E. (1996) *Genética y mejoramiento de la papa*. Available at: https://repository.agrosavia.co/handle/20.500.12324/32459#.Yw0cs5OVJFY.mendeley (Accessed 28, 2022).

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000) *Inference of population structure using multilocus genotype data*. Available at: http://www.stats.ox.ac.uk/pritch/home.html.

Rasmussen, J. O., and Rasmussen, O. S. (1995). Characterization of somatic hybrids of potato by use of RAPD markers and isozyme analysis. *Physiol. Plant* 93, 357–364. doi: 10.1111/J.1399-3054.1995.TB02240.X

R Core Team (2021). R: A language and environment for statistical computing. *R Foundation for statistical computing* (Vienna, Austria) Available at: https://www.R-project.org/. [Accessed August 28, 2022].

Rosero Alpala, M. G., Tapie, W. A., Rosero Alpala, D. A., Rosero Alpala, M. G., Tapie, W. A., and Rosero Alpala, D. A. (2020). Diversidad fenotípica de papas nativas en las comunidades indígenas de la etnia de los pastos (Nariño, colombia): Agricultura ecológica para la seguridad alimentaria y el desarrollo rural. *Rev. Peru Biol.* 27, 509–516. doi: 10.15381/RPB.V27I4.18020

Sánchez, M. A. (2017). Determinación de la ploidía en accesiones de papas nativas (Solanum tuberosum grupo andigenum) procedentes del banco de germoplasma vegetal de CORPOICA. *Respuestas* 22, 6–13. doi: 10.22463/0122820X.1166

Santa Sepúlveda, J. D. (2018). Análisis de QTL revela loci de resistencia cuantitativa a phytophthora infestans y tecia solanivora en papa tetraploide.

Santa, J. D., Berdugo-Cely, J., Cely-Pardo, L., Soto-Suárez, M., Mosquera, T., and Galeano, C. H. M. (2018). QTL analysis reveals quantitative resistant loci for Phytophthora infestans and Tecia solanivora in tetraploid potato (Solanum tuberosum L.). *PLoS One* 13. doi: 10.1371/journal.pone.0199716.

Schmitz Carley, C. A., Coombs, J. J., Douches, D. S., Bethke, P. C., Palta, J. P., Novy, R. G., et al. (2017). Automated tetraploid genotype calling by hierarchical clustering. *Theor. Appl. Genet.* 130, 717–726. doi: 10.1007/S00122-016-2845-5

Spooner, D. M., Gavrilenko, T., Jansky, S. H., Ovchinnikova, A., Krylova, E., Knapp, S., et al. (2010). Ecogeography of ploidy variation in cultivated potato (Solanum sect. petota). *Am. J. Bot.* 97, 2049–2060. doi: 10.3732/ajb.1000277

Spooner, D. M., Ghislain, M., Simon, R., Jansky, S. H., and Gavrilenko, T. (2014). Systematics, diversity, genetics, and evolution of wild and cultivated potatoes. *Botanical Rev.* 80, 283–383. doi: 10.1007/s12229-014-9146-y

Spooner, D. M., Nú, J., Trujillo, G., del Rosario Herrera, M., Guzmá, F., and Ghislain, M. (2007) *Extensive simple sequence repeat genotyping of potato landraces supports a major reevaluation of their gene pool structure and classification*. Available at: www.pnas.orgcgidoi10.1073pnas.0709796104.

Tinjacá, S., and Rodríguez, L. (2015) *Catálogo de papas nativas de nariño, Colombia* (Bogotá, Colombia: Universidad Nacional de Colombia). Available at: http://www.seguridadalimentarianarino.unal.edu.co/sites/default/files/pdf-componentes/C.Mej-Gen_Catalogo%20papas%20Nativas.pdf (Accessed 13, 2022).

Uribe Gaviria, A. F. (2011). Comprobación del nivel de ploidía en acceciones de papa criolla (Solanum tuberosum) grupo phureja, pertenecientes al banco de germoplasma vegetal de CORPOICA. (Tesis de pregrado, Universidad de Cundinamarca).

van Hintum, T., Brown, A. H. D., Spillane, C., and Hodgkin, T. (2000). "Colecciones núcleo de recursos fitogenéticos," in *Boletín técnico no. 3 del IPGRI* (Roma, Italia: Instituto Internacional de Recursos Fitogenéticos). Available at: https://www.researchgate.net/publication/268000676.

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 11 September 2023
DOI 10.3389/fpls.2023.1210046

Check for updates

OPEN ACCESS

EDITED BY
Iris Edith Peralta,
National University of Cuyo, Argentina

REVIEWED BY
Clizia Villano,
University of Naples Federico II, Italy
Benoit Bizimungu,
Agriculture and Agri-Food Canada (AAFC), Canada

*CORRESPONDENCE
Rodomiro Ortiz
rodomiro.ortiz@slu.se
Edgar Neyra
edgar.neyra@upch.pe

RECEIVED 21 April 2023
ACCEPTED 10 August 2023
PUBLISHED 11 September 2023

CITATION
De la Cruz G, Blas R, Pérez W, Neyra E and Ortiz R (2023) Foliar transcriptomes reveal candidate genes for late blight resistance in cultivars of diploid potato *Solanum tuberosum* L. Andigenum Group. *Front. Plant Sci.* 14:1210046. doi: 10.3389/fpls.2023.1210046

COPYRIGHT
© 2023 De la Cruz, Blas, Pérez, Neyra and Ortiz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Foliar transcriptomes reveal candidate genes for late blight resistance in cultivars of diploid potato *Solanum tuberosum* L. Andigenum Group

Germán De la Cruz[1], Raúl Blas[2], Willmer Pérez[3], Edgar Neyra[4,5]* and Rodomiro Ortiz[6]*

[1]Laboratorio de Genética y Biotecnología Vegetal, Facultad de Ciencias Agrarias, Universidad Nacional de San Cristóbal de Huamanga (UNSCH), Ayacucho, Peru, [2]Instituto de Biotecnologia (IBT), Facultad de Agronomia, Universidad Nacional Agraria La Molina (UNALM), Lima, Peru, [3]Plant Pathology Laboratory, Crop and Systems Sciences Division, International Potato Center, Lima, Peru, [4]Unidad de Genómica, Laboratorios de Investigación y Desarrollo, Facultad de Ciencias e Ingeniería, Universidad Peruana Cayetano Heredia, Lima, Peru, [5]Departamento Académico de Tecnología Médica, Facultad de Medicina, Universidad Peruana Cayetano Heredia, Lima, Peru, [6]Department of Plant Breeding, Swedish University of Agricultural Sciences, Lomma, Sweden

Characterization of major resistance (*R*) genes to late blight (LB) –caused by the oomycete *Phytophthora infestans*– is very important for potato breeding. The objective of this study was to identify novel genes for resistance to LB from diploid *Solanum tuberosum* L. Andigenum Group (StAG) cultivar accessions. Using comparative analysis with a edgeR bioconductor package for differential expression analysis of transcriptomes, two of these accessions with contrasting levels of resistance to LB were analyzed using digital gene expression data. As a result, various differentially expressed genes ($P \leq 0.0001$, $Log_2FC \geq 2$, FDR < 0.001) were noted. The combination of transcriptomic analysis provided 303 candidate genes that are overexpressed and underexpressed, thereby giving high resistance to LB. The functional analysis showed differential expression of *R* genes and their corresponding proteins related to disease resistance, NBS-LRR domain proteins, and specific disease resistance proteins. Comparative analysis of specific tissue transcriptomes in resistant and susceptible genotypes can be used for rapidly identifying candidate *R* genes, thus adding novel genes from diploid StAG cultivar accessions for host plant resistance to *P. infestans* in potato.

KEYWORDS

gene expression, genome, host plant resistance, *Phytophthora infestans*, *Solanum tuberosum*

Introduction

The oomycete *Phytophthora infestans* (Mont.) de Bary (Peronosporaceae) causes late blight (LB) in potato and other crops in the family Solanaceae. *Phytophthora infestans* is one of the most devastating pathogens worldwide and instigated the tragic Irish famine in the 19th century (Fry et al., 2015). It remains a challenge due to its rapid adaptation to climate change (Wu et al., 2020), early appearances, and its ability to cope with high temperatures (Lehsten et al., 2017; Lurwanu et al., 2021) and ultraviolet radiation (Wu et al., 2019). In Europe, this pathogen accounts for at least US$ 6 billion per year of losses in potato harvests (Haverkort et al., 2016). More than 2,000 MT of fungicide was used to control LB in the USA in 2001 (Magarey et al., 2019). In the Peruvian and Bolivian Andes, which is the primary center of the origin and diversity of the potato crop, control of LB is difficult for poor farmers, who also lack rural extension services. The *P. infestans* population from Peru includes the lineages EC-1, US-1, US-1, PE-3, PE-5, PE-6 and PE-7(Lindqvist-Kreuze et al., 2020). The EC-1 clonal lineage, which shows a complex virulence, is dominant in this country and is characterized by isolates with resistance to metalaxyl. Native potato germplasm are found in high frequency where its crop wild relatives thrive and released cultivars are grown (Lindqvist-Kreuze et al., 2020) including diploid *Solanum tuberosum* L. Andigenum Group (StAG) accessions (formerly known as *S. goniocalyx* Juz. & Bukasov [Ovchinnikova et al., 2011], but it is a synonym no longer accepted), whose center of origin and distribution ranges from central Peru to northern Bolivia. Preventative application of fungicides is being used to control LB. However, their extensive use poses detrimental risks to human health and to the environment (Jiang et al., 2018). Identifying *Solanum* sources of resistance to LB and introducing them into the elite cultigen pool through crossbreeding could become an environmentally benign alternative to fungicides (Enciso-Maldonado et al., 2022).

Research on potato gene expression for LB resistance using different methods (Northern and Southern blots) allows for studying a few genes individually. Next-generation sequencing technology (NGS) (UCLA, 2017) has driven an increase in genetic and genomic research based on DNA or RNA sequences (RNAseq). The deciphering of the potato genome sequences (Xu et al., 2011) enabled the assembling of a reference genome and became a database for differential expression research, thereby promoting the use of methods based on RNAseq (Li et al., 2012; Cevallos Hidalgo, 2015). Using this technology, the leaflet transcriptome of tetraploid *S. tuberosum* 'Yungay' when interacting with *P. infestans* (POX-107 and POX-067 strains) was elucidated (Izarra and Lindqvist-Kreuze, 2016). Likewise, high differential expression of genes was noticed in tubers of transgenic plants of tetraploid *S. tuberosum* 'Russet Burbank' after performing RNAseq analysis of their tubers interacting with the pathogen *P. infestans* at 48 hours after inoculation (hai) (Gao and Bradeen, 2016). High differential expression at 15 hai was also noticed in tetraploid *S. tuberosum* 'Bintje' when interacting with isolates 80029, 88133, and T30-4 of *P. infestans* (Avrova et al., 2003). The same results related to high differential expression at 48 hai under different environments were reported elsewhere (Wang et al., 2005; Ali et al., 2014; Stefańczyk et al., 2017).

Differential expression research with RNAseq in cultivated native potatoes inoculated with *P. infestans* was undertaken with diploid *S. phureja* Juz. & Bukasov [a synonym not accepted of StAG (Ovchinnikova et al., 2011)] (Evers et al., 2006; Massa et al., 2013). Nonetheless, RNAseq research on the other cultivated diploid StAG has not been pursued until this study, the main goal of which was to identify new candidate genes for resistance to *P. infestans* in two transcriptomes differentially expressed in the leaf, using accessions of two diploid-resistant and susceptible StAG cultivars.

Materials and methods

Plant materials

Accessions of two diploid ($2n = 2x = 24$ chromosomes) *Solanum tuberosum* L. Andigenum Group cultivars that are resistant and susceptible ['Wira Pasña' (CIP 704270)] and 'Sumaq Perqa' (CIP 703777), respectively] to *P. infestans* from Peru (Pérez et al., 2014; De la Cruz et al., 2020) were used in this study. The International Potato Center (CIP, Lima, Perú) provided the *in vitro* plantlets for this research. They were micro-propagated to increase the number of plantlets by tissue culture using Murashigue and Skoog media and environmental conditions according to Espinoza et al. (1989) and subsequently transplanted into pots with a moss: sand:agricultural soil substrate (2:1:2 v/v) for growth and phenotyping in a greenhouse of the CIP.

Phenotyping of potato landraces to *Phytophthora infestans*

The *P. infestans* isolate POX-67, which belongs to the clonal lineage EC-1 isolated from Oxapampa (Perú) (Pérez et al., 2014; Izarra and Lindqvist-Kreuze, 2016), previously stored in liquid nitrogen was propagated in potato slices according to an established protocol (Vleeshouwers et al., 1999). Plants of the resistant and susceptible accessions were inoculated before flowering with a concentration of 3,000 sporangia/ml using two methods, detached leaves and whole plants (Forbes et al., 2014; Pérez et al., 2014; Chang-Kee, 2016), to confirm their response to the pathogen (Figure 1). For the first method, three leaflets per accession were detached from the middle part of 6-week-old potato plants or before flowering initiation. Leaflets were placed in the lids of inverted Petri dishes with 1.5% water agar in the base and inoculated by placing one 20 µl drop of inoculum containing $3x10^3$ sporangia/ml on the midrib, thus producing one lesion per leaflet. Inoculated leaflets were incubated at 18 °C with a 14 h light/day cycle for 5 days after that the percentage of foliage area that is infected by *P. infestans* was determined visually according to Forbes et al. (2014). In the second method for the whole-plant assays, three 6-week-old plants, or before flowering initiation of each accession, were inoculated with a sporangial suspension of 40 ml of inoculum containing $3x10^3$ sporangia/ml until run-off (Eshraghi et al., 2011).

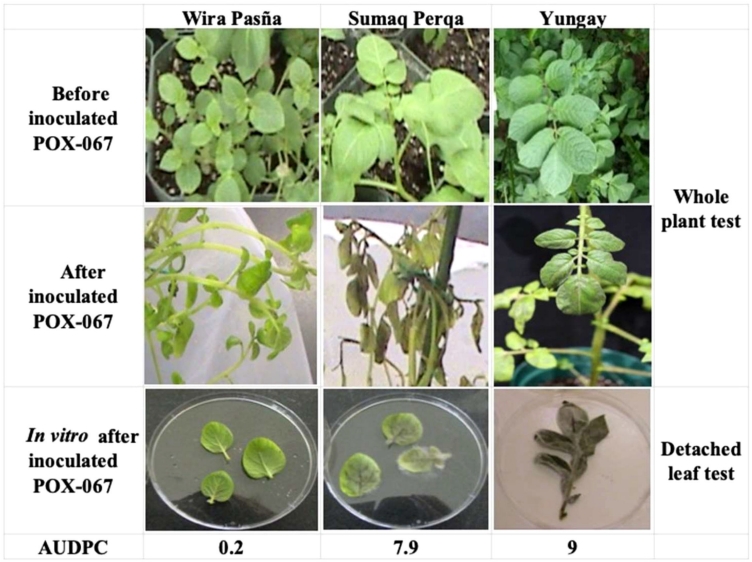


FIGURE 1
Phenotypic response of diploid *S. tuberosum* cultivar accessions 'Wira Pasña' (CIP 704270) and 'Sumaq Perqa' (CIP 703777) inoculated with *P. infestans* POX-067. Whole plant of 'Wira Pasña' and 'Sumaq Perqa' before and after 6 days after inoculation (dai). Detached leaf test in Petri dishes after 6 (dai), the tetraploid *S. tuberosum* 'Yungay' was used as a susceptible control. AUDPC = area under disease progress curve.

After inoculation, the plants were incubated at 16–18 ˚C with a relative humidity of 90% for 7 days. Severity (%) was evaluated once as described by Forbes et al. (2014).

Late blight symptoms caused by *P. infestans* were evaluated in the host potato cultivar using the detached leaf assay. The percentage of diseased leaf area was determined visually according to Forbes et al. (2014) every 48 hai for 6 days. In whole plants, with the severity data (%), the area under the disease progress curve (AUDPC) was calculated (Forbes et al., 2014). The accessions were also assessed through whole plant inoculation, as a second definitive test, which included the tetraploid ($2n = 4x = 48$ chromosomes) *Solanum tuberosum* 'Yungay' (CIP 720064) as a susceptible control (Forbes et al., 2014; Pérez et al., 2014; Izarra and Lindqvist-Kreuze, 2016). This is the most popular potato cultivar in Peru and is grown in 22% of the total potato planting area (Pradel et al., 2017).

RNAseq

The transcriptomes were from two leaf sampling times in each cultivar. At 0 hours they were inoculated with water without the pathogen and 48 hai with 3,000 sporangia/ml of *P. infestans* strain POX-67. The experiment included three replications labeled as R48A, R48B, R48C for 48 hai and R0D, R0E, R0F for 0 hai in the resistant cultivar, as well as S48G, S48H, S48I for 48 hai and S0J, S0K, S0L for 0 hai for the susceptible cultivar. In total, 12 libraries were analyzed. *Solanum tuberosum* 'Wira Pasña' and 'Sumaq Perqa' do not have strictly the same genome because they have evolved into two native Andean cultivars, and they produce two different transcriptomes in each accession. The differential expression test for the comparison of the transcriptome was always from the resistant to the susceptible (R/S) to separate the host plant resistance genes to *P. infestans* only in 'Wira Pasña' without the effect of the accessions. Hence, the transcriptome comparisons were between the resistant and susceptible cultivars stimulated by the pathogen at 48 hai (R48/S48); without the stimulus of the pathogen at 0 hai (R0/S0); and by the differential, i.e., (R48/S48) versus (R0/S0) to only obtain the genes differentially expressed by the stimulation of the pathogen without the accession effect in 'Wira Pasña'.

RNA extraction and sequencing

Immediately after extracting the leaf samples, they were introduced to liquid nitrogen and transferred to the laboratory, where they were preserved in an ultra-cold chamber (-80°C). The RNA extraction of the 12 leaf samples of both accessions was done in the laboratory of the Genomics Unit of the Universidad Peruana Cayetano Heredia (UPCH, Lima, Perú). The extraction protocol used was pre-tested and standardized (Espinoza, 2017). Leaf tissue (1 to 2 g) was crushed using a mortar and pestle with liquid nitrogen. Thereafter, Tri®Reagent (Sigma) (1ml/100g of tissue) and chloroform (200ul/1ml of Tri®Reagent) were used to continue the centrifugation series without breaking the cold chain (4°C) and to avoid RNA degradation. The precipitation of the "pellet" was achieved with isopropanol and washing with 75% ethanol. The samples were treated with a DNA-freeTM Kit (Ambion, USA) to clean contaminating DNA, ensuring optimal quality and quantity of the total RNA. The quality of total RNAs was confirmed through the absorbance ratio on a Thermo Scientific Spectrophotometer NanoDrop 2000, the RNA Integrity Number (RIN) on an Agilent 2000 Bioanalyzer (Plant RNA Nano Chip, Agilent), and the absence of smear.

The concentration of total RNA extracted ranged from 93.13 to 468.01 ng/ul, with an average of 297.63 ng/ul, in an average extraction volume of 69.66 ul. The purity values of the RNA evaluated by spectrophotometry were on average 2.14 (absorbance A260/A280) and 2.23 (absorbance A260/A230). The RIN integrity values ranged from 7.7 to 8.6, with an average of 8.13. The results of the amount of total extracted RNA (ng/ul) and the absorbance and integrity of the RNA of the 12 samples are given in Supplementary Table 1.

Sequencing was performed on the Illumina Hi-Seq 2500 platform, with Q30 (i.e., below 1/1000 of sequenced nucleotides) and a sequencing accuracy of 99.9%. The size of the reading fragments was 150 bp. Twelve cDNA paired-ends (PE) libraries were constructed. This process was achieved through the service of NovoGene Corporation Inc. (California, USA, https://en.novogene.com/).

Bioinformatic analysis of RNAseq data

The bioinformatic analysis (Gollery, 2006) was performed using the server and platform in a Linux environment (Carreño, 2017). FastQC (Leggett et al., 2013) (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used to assess the quality of the sequences considering only libraries with Q ≥ 33. Trimmomatix (Bolger et al., 2014) (https://github.com/kbaseapps/kb_trimmomatic) was used to clean the sequences of the Illumina adapters, considering a score of 33 and filtering the libraries with a minimum length of 145 bp in the reads. STAR (Dobin and Gingeras, 2015) was used for the mapping of the libraries to the reference genome of *S. phureja*, which was annotated by the Potato Genome Sequencing Consortium (PGSC) DM v4.03 (Xu et al., 2011). The *.sam files were converted to *.bam using SAMtools (http://www.htslib.org) (Li et al., 2009). SUBREAD (featureCounts) facilitated the quantification of the levels of expression of transcripts between treatments, obtaining a "count table" with the number of reads mapped to each segment of the genome ("transcripts") and their corresponding values in count-per-million (CPM), which were further used to detect differentially expressed genes (DEGs) among resistant and susceptible accessions.

The predictive analysis of DEGs when comparing the two transcriptomes of both cultivar accessions in response to stress with *P. infestans* was performed using the bioconductor package *edgeR* in the Rstudio environment. The input (table count data) was extracted to the R environment through *read.delim* and principal component analysis (PCA) of the 12 libraries with unfiltered data was performed with the *DGEList* and *plotMDS* packages. The filtering was performed with CPM rather than filtering the counts directly. CPMs of differentially expressed transcripts were then normalized using the *cpm* and *calcNormFactors* packages. The model was defined considering the experimental design using the *model.matrix* and *estimateDisp* packages so that the dispersion of the transcripts was only due to the effect of the treatments. The adjustment of the values was made to the quasi-likelihood (QL) F-tests, since the number of biological replicates was below five (Law et al., 2018), using the *glmQLFTest* and *TopTags* packages. Differentially expressed transcripts or candidate genes for host plant resistance were considered if they had a greater difference in abundance with $Log_2FC \geq 2$ up-expressed and $Log_2FC \leq -2$ down-expressed, at a probability of significance below or equal to 0.0001, and with a probability of false positive readings below 0.001 with the Benjamini-Hochberg method. "complexHeatmap" and "ggplot2" were used to make the smear, volcano, and heatmap plots. Likewise, the software Venny 2.1 (Oliveros, 2016) was used to organize the DEG sets in a Venn Euler diagram.

Functional analysis of the over and underexpressed transcripts was performed separately. The gene ontology (GO) and the enrichment of DEGs in metabolic pathways noted in the Kyoto Encyclopedia of Genes and Genomes (KEGG) was facilitated by the g:Profiler software (Raudvere et al., 2019) (https://biit.cs.ut.ee/gprofiler/gost) using the database of *S. tuberosum* PGSC_GENE, providing the identifiers (IDs) of the genes and their corresponding significant value (padj < 0.05) over and underexpressed peptides. To calculate the padj, the Benjamini-Hochberg FDR (false discovery rate) method was used, using g:Profiler and g:GOSt (functional profiling), with the user threshold at 0.05.

Results

Phenotyping of diploid StAG cultivar accessions

This experiment used whole plants of two diploid StAG cultivar accessions, namely the LB-resistant 'Wira Pasña' (CIP 704270) and the susceptible 'Sumaq Perqa' (CIP 703777), which were previously evaluated using a detached leaf assay. The susceptible tetraploid *S. tuberosum* 'Yungay' was used as a control (Figure 1). The resulting area under the disease progress curve (AUDPC) was 0.2 and 7.9 respectively for the resistant and susceptible diploid StAG cultivar accessions and 9 for 'Yungay'.

Sequencing

The results of the Illumina sequencing, in the *.fasta files of the 12 libraries, were reported considering paired reads (PE). The readings read in direction 5' to 3', "forward", were identified by adding _1 to the library code, those that were read in direction 3' to 5', "reverse", were identified by adding _2. Therefore, 24 libraries resulted whose sequencing yields (Supplementary Table 2) resulted in the range of 22´426,271 to 32´315,289 reads per library with an average of 27´370,780. All of them exhibited good quality (Q30) sequencing as analyzed with FastQC. In total, 656´898,720 readings were obtained in the 12 paired libraries. Approximately 94% of the readings (on average 26´270,256 readings per bookstore) "survived" the "cleaning" with *Trimmomatic* (Supplementary Table 3).

Mapping of reads to the reference genome

Table 1 lists the results of the alignment of the reads to the indexed reference genome (*S. tuberosum* Soltub.3.0.dna.toplevel.fa.gz).

Approximately 92% of the total reads were mapped, of which 88% of the reads were at single sites, 4% of the reads were at more than one site, and 8% of the reads were not mapped to the reference genome. There were 40,336 transcripts (i.e., total genes that are annotated) in the 12 libraries. After filtering the counts with *edgeR*, 19,666 aligned transcripts were obtained without distorting the modeling in all repetitions and in at least one treatment.

Differential expression analysis

Principal component analysis (PCA) was performed on all expressed transcripts (40,336 transcripts) from the 12 libraries (Figure 2, left, all treatments away from the center). The result of the normalization of the 12 libraries is shown in the boxplot of Figure 2 (right), in which the averages of each library were standardized and normalized with homogeneous distributions and averages of the CPM values.

First Analysis between resistant and susceptible transcriptomes without the stimulus of the pathogen at 0 hai (R0/S0)

The transcriptomes of the resistant 'Wira Pasña' and the susceptible 'Sumaq Perqa' were differentially expressed at 0 hai. In total, 400 genes in the transcriptome of 'Wira Pasña' (Supplementary Table 5; https://docs.google.com/spreadsheets/d/1paCEZkQXS6kSgNEnTb5p9r6FJJimoFYF/edit?usp=sharing&ouid=110777050509080518982&rtpof=true&sd=true) had an accession (A) effect. These genes are the ones that transcriptomically differentiate this resistant cultivar from the susceptible 'Sumaq Perqa' when they are not stimulated by the pathogen (0 hai). They were used in the following analysis as baseline differentially expressed genes (DEGs). The distribution of said differentially expressed genes is shown in Figure 3.

Second analysis between resistant and susceptible transcriptomes stimulated by the pathogen (P) at 48 hai (R48/S48)

There were 303 genes differentially expressed 48 hai in the transcriptome of the resistant cultivar after being stimulated with *P. infestans* (Supplementary Table 4; https://docs.google.com/spreadsheets/d/1LGJqSFDtgcoRl65pg8ei94tLq1pUbZqF/edit?usp=sharing&ouid=110777050509080518982&rtpof=true&sd=true). These genes transcriptomically differentiate 'Wira Pasña' from 'Sumaq Perqa', in addition to the effect of the accessions. Figure 4 shows the distribution of the differentially overexpressed and underexpressed genes (Log_2FC >2 and <-2, respectively, and FDR ≤ 0.001).

Third analysis based on previous differential transcriptomes (R48/S48 versus R0/S0)

There were 400 differential genes obtained from the previous basal differential quantification between the resistant cultivar

TABLE 1 Number of reads and percentage of reads, mapped to the *Solanum tuberosum* Soltub.3.0.dna.toplevel.fa.gz reference genome, from 12 StAG diploid libraries.

Sample Name	Total reads	Total mapped		Multiple mapped		Uniquely mapped		Unmapped
		N	%			N	%	(%)
R48A	47,398,026	43,353,448	91.47	1,654,926	3.49	41,698,522	87.98	8.53
R48B	51,218,548	46,818,700	91.41	1,847,371	3.61	44,971,329	87.80	8.59
R48C	57,910,202	53,204,824	91.87	1,976,333	3.41	51,228,491	88.46	8.13
S48G	59,893,568	54,666,092	91.27	2,045,387	3.42	52,620,705	87.86	8.73
S48H	54,708,012	49,917,835	91.24	1,964,538	3.59	47,953,297	87.65	8.76
S48I	72,287,734	66,225,093	91.61	2,615,413	3.62	63,609,680	88.00	3.89
R0D	41,740,904	38,532,626	92.31	1,607,250	3.85	36,925,376	88.46	7.69
R0E	52,261,774	48,450,142	92.71	2,062,379	3.95	46,387,763	88.76	7.29
R0F	63,342,836	58,632,077	92.56	2,716,340	4.29	55,915,737	88.27	7.44
S0J	42,225,972	39,194,192	92.82	1,901,763	4.50	37,292,429	88.32	7.18
S0K	52,638,960	48,959,396	93.01	3,788,764	7.20	45,170,632	85.81	6.99
S0L	58,355,444	53,596,973	91.85	2,381,674	4.08	51,215,299	87.76	8.15
Total	653,981,980	601,551,398		26,562,138		574,989,260		

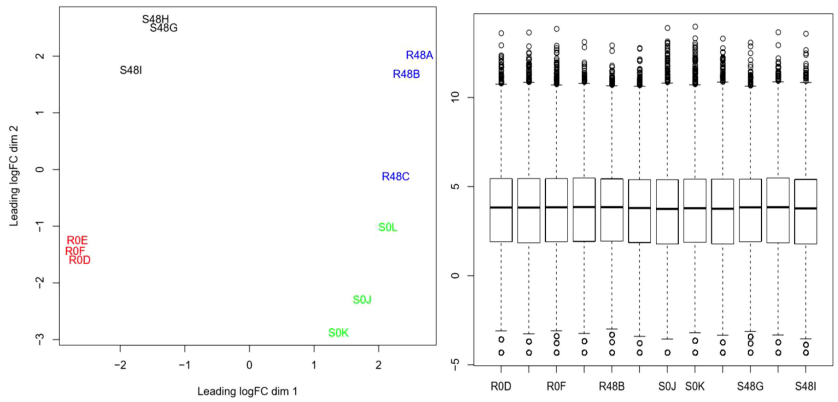


FIGURE 2
Grouping of 12 libraries (40,336 transcripts) by principal component analysis in resistant and susceptible diploid StAG cultivar accessions. The left graph shows the trend of four groups with opposite and contrasting locations. The grouping trends were in opposite locations for the samples of the resistant cultivar at 0 hours after inoculation (hai) (R0D, R0E, and R0F in red) and at 48 hai (R48A, R48B, and R48C in blue), thereby confirming the proper handling of the samples. The samples of the susceptible cultivar at 0 hai (S0J, S0K, and S0L in green) and those at 48 hai (S48G, S48H, and S48I in black) were in opposite groupings. The right graph shows the boxplot representation of the 12 libraries (19,666 genes) normalized with *edgeR* software in Rstudio.

against the susceptible cultivar at 0 hai (WP0/SP0), and another 303 differential genes noticed at 48 hai (WP48/SP48). They are included in a Venn-Euler diagram (Figure 5), which shows that 77 unique genes were differentially expressed in the resistant 'Wira Pasña', after being stimulated 48 hai by the combined accession effect and that of the pathogen (A+P).

The combined effect of accessions plus pathogen (A+P) at 48 hai stimulated 303 genes to be differentially expressed in 'Wira Pasña', of which 136 genes (Figure 6, blue set) were overexpressed ($Log_2FC > +2$, FDR ≤ 0.001). From this group, 35 genes were only upregulated (blue subset). In addition, 167 genes (green set) were underexpressed ($Log_2FC > -2$, FDR ≤ 0.001). In this last group, 42 genes were only underexpressed (green subset) in the resistant accession.

There were 400 genes (Figure 6) differentially expressed in the 'Wira Pasña' at 0 hai, of which 177 genes (yellow set) were overexpressed ($Log_2FC > +2$, FDR ≤ 0.001). From that group, only 52 genes (yellow subset) were only overexpressed due to the accession effect (A). In addition, 223 genes (red set) were underexpressed ($Log_2FC > -2$, FDR ≤ 0.001) and from this last group, 122 genes were only underexpressed (red subset) due to A.

In the resistant 'Wira Pasña', 226 genes with differential expression were quantified (subsets with 101 and 125 genes, Figure 6). They were expressed in the opposite direction because of the accession (A) and accessions plus pathogen (A+P) effects. The 101 differential genes were overexpressed due to A+P and these same ones were underexpressed due to A. The other 125 genes were overexpressed due to A and were underexpressed by stimuli of A+P at 48 hai.

The relationship of genes over and underexpressed in the transcriptome of 'Wira Pasña', considering both A+P and A, are given in Supplementary Table 4. There were 303 genes related to host plant resistance to *P. infestans* (Figure 6), of which 136 were overexpressed genes (35 + 101 genes) and 167 were underexpressed genes (42 + 125 genes) ($Log_2FC > 2$ and <-2, respectively and FDR ≤ 0.001).

Gene ontology enrichment analysis

The GO analysis with GO term and adjusted *P*-values from 303 DEGs is given Supplementary Table 6 (https://docs.google.com/spreadsheets/d/1s9WK4_gRhnq9b8bv5rrgjLg8eE1DXyrr/edit?usp=sharing&ouid=110777050509080518982&rtpof=true&sd=true). Of these, 167 underexpressed DEGs are shown in Figure 7, while the 136 overexpressed DEGs of resistant 'Wira Pasña' are given in Figure 8. The 167 underexpressed DEGs were enriched in the three main categories of the GO. Figure 7 shows the nine genes enriched for molecular function (GO.MF), 65 genes corresponding to biological process terms (GO.BP), and 20 related to cellular components (GP.CC).

The binding function (GO:0003674) had the highest significance (padj. $1.937x10^{-11)}$ in the molecular function category, followed by the functions of catalytic activity (GO:0003824) and binding of nucleic acids (GO:0003676) with pad. 1.668×10^{-5} and 1.432×10^{-4}, respectively. The other molecular functions were non-significant. For the biological process category, the metabolic process (GO:0008152) and the cellular process (GO:0009987) had the highest significance (both with padj, $1.006 \times 10^{-31)}$). The metabolic processes of organic substances (GO:0071704), primary metabolic processes (GO:0044238), cellular metabolic processes (GO:0044237), metabolic processes of nitrogenous compounds (GO:0006807), macromolecular metabolic processes (GO:0043170), and metabolic processes of cellular macromolecules (GO:0044260) had high significance (padj range: 1.547×10^{-26} to 9.836×10^{-12}). The other biological processes were not significant. Membrane components (GO: 0005623) had high

10.3389/fpls.2023.1210046

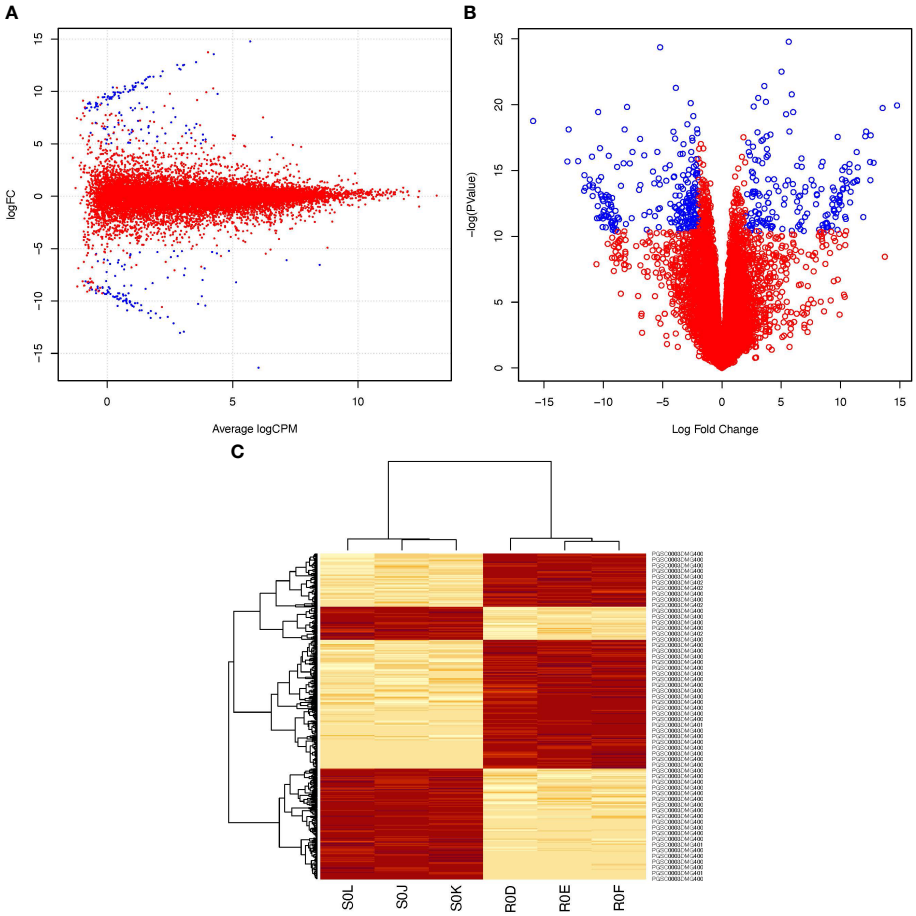


FIGURE 3
Differential expression profile of 400 genes between resistant *S. tuberosum* 'Wira Pasña' versus susceptible *S. tuberosum* 'Sumaq Perqa', without stimulation of the pathogen *P. infestans* at 0 hours after inoculation (hai). In Box **(A)** (smear plot) and Box **(B)** (volcano plot), the genes represented in blue were overexpressed (Log_2FC >+2) and underexpressed (Log_2FC < -2), respectively. In box **(C)**, four clusters are observed in the horizontal heatmap, of which the two red clusters include the overexpressed genes in 'Wira Pasña' and two other groups of different genes that were overexpressed in 'Sumaq Perqa'. The vertical heatmap shows the grouping trend of the 400 genes of the resistant 'Wira Pasña ') in two vertical clusters with different intensities (heat) of expression in the six libraries at 0 hai, of which three correspond to 'Wira Pasña' (R0D, R0E, and R0F) and another three to 'Sumaq Perqa' (S0J, S0K, and S0L). The intensity of expression of the clustered genes is related to the intensity of color. The more intense the red, the more overexpressed (> E) the gene, and the less intense, the less expressed the genes, until reaching white, when they are repressed (< E).

significance (padj. 4.119×10^{-30}) in the cellular components category, followed by the intracellular component (GO:0005622), intrinsic components of the membrane, organelles (GO:0031224), and organelle membrane-bound intracellular (GO:0043231) with padj. ranging from 2.178×10^{-26} to 2.425×10^{-22}. The other cellular components were not significant.

The 136 overexpressed DEGs were enriched in the three main categories of the GO. Figure 8 shows the 16 genes enriched for molecular function (GO.MF), of which 46 genes were related to biological process terms (GO.BP) and 15 to cellular components (GO.CC).

The binding function (GO:0005488) had the highest significance (padj. $1.937 \times 10^{-17)}$ in the molecular function category, followed by the functions of catalytic activity (GO:0003824), binding of heterocyclic components (GO:1901363), and components of organic cycles (GO:0097159) with padj. ranging from 4.131×10^{-9} to $1.047 \times 10^{-8,}$ and the hydrolase activity function (GO:0016787) with padj. of $1.011 \times 10^{-4.}$ The other molecular functions were not significant. The metabolic process (GO:0008152) and the cellular process (GO:0009987) were the most significant (padj, $1.574x10^{-39}$ and $1.202x10^{-32}$, respectively). in the biological process category. The cellular processes (GO:0009987), metabolic processes of organic substances (GO:0071704), primary metabolic processes (GO:0044238), cellular metabolic processes (GO:0044237), metabolic processes of nitrogenous compounds (GO:0006807), metabolic processes of macromolecules (GO:0043170), and biosynthesis processes (GO:0009058) had high significance with their Padj ranging from 1.202×10^{-32} to 1.587×10^{-12}. The other biological processes were not significant. Cellular components (GO:0044464) and the intracellular component (GO:0005622) had high significance (padj. 1.703×10^{-36} and 1.359×10^{-31}, respectively) in the cellular components category, followed by intracellular components (GO:0044424), membrane (GO:0016020), and membrane intrinsic components (GO:0031224) with padj. ranging from 1.703×10^{-30} and 1.359×10^{-16}. The other cellular components were not significant.

10.3389/fpls.2023.1210046

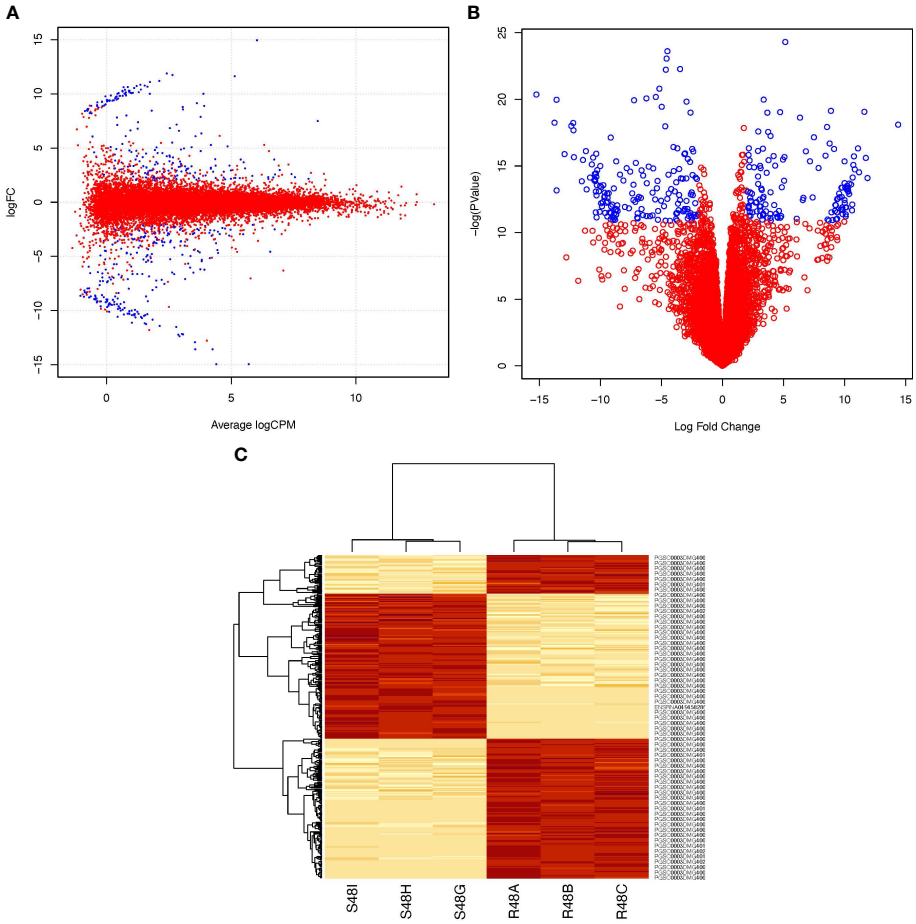


FIGURE 4
Differential expression profile of 303 genes between the resistant 'Wira Pasña', the combined effect of 'Sumaq Perqa' and 'Wira Pasña', plus the effect of *P. infestans pathogen* at 48 hours after inoculation (hai). In Box **(A)** (smear plot) and Box **(B)** (volcano plot) the genes represented in blue were overexpressed ($Log_2FC > +2$) and underexpressed ($Log_2FC > -2$), respectively. In Panel **(C)**, the horizontal heatmap shows three clusters, including a red cluster of overexpressed genes in 'Wira Pasña' and two other red groups of different genes that were overexpressed in 'Sumaq Perqa'. The vertical heatmap shows the grouping trend of the 303 differential genes in two clusters with different intensities (heat) of expression in the libraries of 'Sumaq Perqa' (S48G, S48H, and S48I) and 'Wira Pasña' (R48A, R48B, and R48C) at 48 hai. In the clusters, there was a difference between a group of genes that were overexpressed and two groups that were underexpressed in the resistant cultivar 'Wira Pasña' because of more pathogenic accessions (A+P). The intensity of expression of the clustered genes is related to the intensity of color. The more intense the red, the more overexpressed (> E) the gene, and the less intense, the less expressed the gene, until reaching white (< E).

Kyoto encyclopedia of genes and genomes metabolic pathway enrichment analysis

KEGG serves as a basic platform for the systematic analysis of gene function in terms of networks or metabolic pathways of gene products (Kanehisa, 2019) (Kanehisa and Goto, 2000) (Kanehisa et al., 2023). It was used to identify the biosynthetic pathways that are active in the resistant 'Wira Pasña' after being infected by the *P. infestans* isolate POX-067. The first analysis was for the underexpressed DEGs (Figure 9), and thereafter for the overexpressed ones (Figure 10).

The metabolic pathways (KEGG:01100) and the secondary metabolite biosynthesis pathway (KEGG:01110) had a high significance (padj. of 9.849×10^{-31} and $4.873 \times 10^{-12,}$ respectively, considering the annotation of 167 underexpressed DEGs in the KEGG database. Likewise, plant hormone transduction signal pathways (KEGG:04075) and the spliceosome pathway (KEGG:03040) were significant (padj of 2.458×10^{-3} and 2.022×10^{-2}, respectively). The other routes were not significant.

Of the 136 overexpressed DEGs that were annotated to the KEGG database, two main enriched pathways, namely, the metabolic pathway (KEGG:01100) and the secondary metabolite biosynthesis pathway (KEGG:01110), were highly significant (padj. $2{,}123 \times 10^{-31}$ and $3{,}027 \times 10^{-12}$, respectively). The plant-pathogen interaction pathway (KEGG:04626) with a padj of 3.114×10^{-1} was also found, although it was not significant, as well as the other remaining routes.

Candidate R genes for resistance of StAG accessions against *P. infestans* isolate POX-067

A total of 20 genes were found from the enrichment of the 167 underexpressed genes and the 136 overexpressed DEGs at 48 hai

10.3389/fpls.2023.1210046

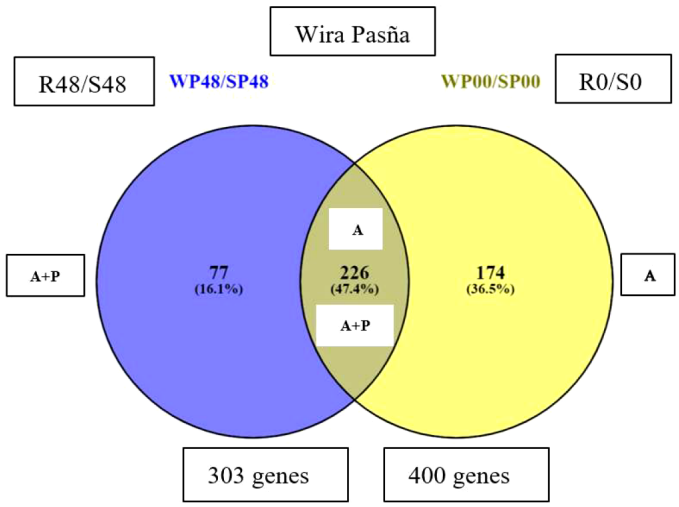


FIGURE 5
A Venn-Euler diagram showing, in the blue subset, that there were 77 genes expressed in the resistant 'Wira Pasña' by the combined effect of accessions plus treatment (A+P) at 48 hours after inoculation (hai), while in the yellow subset, there were 174 genes expressed in 'Wira Pasña' by the accessions effect (A). At the intersection, there are 226 genes that were expressed by the combined effect (A+P) in addition to the accessions effect (A) and were expressed in different directions.

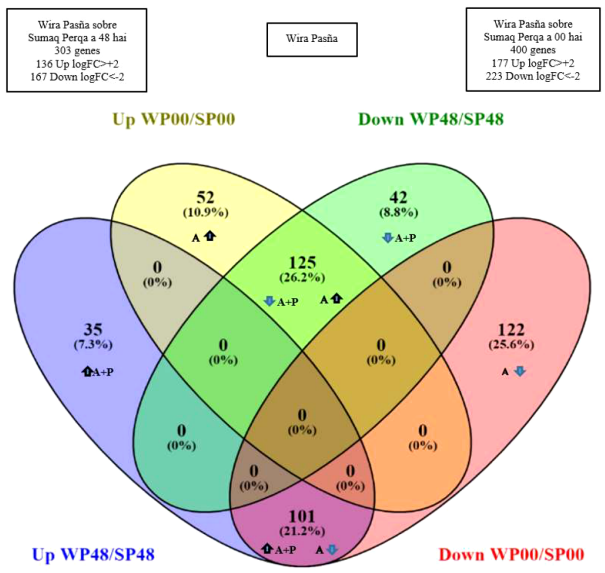


FIGURE 6
Venn-Euler diagram representing the genes of resistant cultivar 'Wira Pasña'. There were 35 and 52 genes upregulated by the combined effect (A+P) at 48 hours after inoculation (hai) and accession effect (A) at 0 hai, respectively (blue and yellow subsets respectively), of which only the subgroup of 35 genes (blue subset) are overexpressed by the combined effect (A+P). There were 42 and 122 genes downregulated by A+P at 48 hai and by A at 0 hai, respectively (green and red subsets, respectively), of which only the subgroup of 122 genes (subset red) was underexpressed by A. In addition, 226 genes were differentials (subsets intersections 101 and 125 genes) that were expressed in opposite directions (arrows) when stimulated by A+P and A.

(Table 2). They are related to resistance to pathogens in general, of which 10 of these genes were underexpressed, namely, one gene for virus resistance (*PGSC0003DMG402016602*), another gene for bacteria (*PGSC0003DMG401012062*), four genes for protein resistance NBS-LRR (*PGSC0003DMG401007871*, *PGSC0003DMG400007605*, *PGSC0003DMG400002217*, and *PGSC0003DMG400007870*), two genes (*PGSC0003DMG400016372* and *PGSC0003DMG400018462*) for protein conferring disease resistance, the *PGSC0003DMG400005471* gene of Receptors Like Kinases (RLK) type, and the *PGSC0003DMG402024222* gene for BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase. Among these 10 genes underexpressed at 0 hai, four belong to the NBS-LRR gene family, and one is an RLK type, in addition to the gene for "BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase".

The 10 genes overexpressed for host plant resistance to *P. infestans* in diploid StAG accessions (Table 2) were one gene (*PGSC0003DMG400008596*) from the NBS-LRR family of resistance proteins, two genes (*PGSC0003DMG400025547* and *PGSC0003DMG400003380*) for resistance proteins against late blight, another gene (*PGSC0003DMG400005590*) for protein disease resistance at large, a gene (*PGSC0003DMG400002427*) encoding proteina4 resistance against bacterial disease, three genes (*PGSC0003DMG400031277*, *PGSC0003DMG400036554*, and *PGSC0003DMG400031279*) for translation factors, a gene (*PGSC0003DMG400008593*) for flavonoid glucoyltransferase UGT73E2, and the *PGSC0003DMG400017234* gene for auxin-induced protein 5NG4. Among these 10 genes, the NBS-LRR gene and the two specific genes for resistance proteins against late blight stand out. They were underexpressed at 0 hai when challenged by the pathogen *P. infestans* but they were overexpressed at 48 hai.

Discussion

As noted by previous research, when determining the phenotype of an accession or genotype, the concentration of

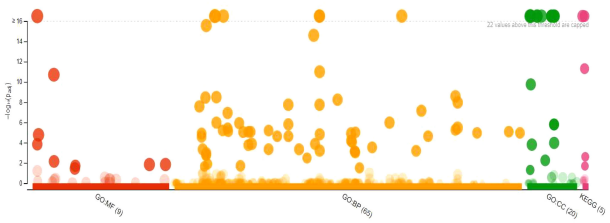


FIGURE 7
Manhattan plot of 167 underexpressed differentially expressed genes (DEGs) from diploid *Solanum tuberosum* 'Wira Pasña' (CIP-704270) after gene ontology enrichment analysis using g:Profiler.

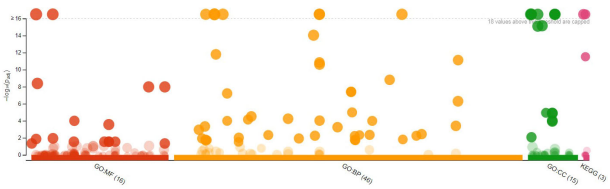


FIGURE 8
Manhattan Plot of 136 overexpressed differentially expressed genes (DEGs) from diploid *Solanum tuberosum* 'Wira Pasña' (CIP-704270) after gene ontology enrichment analysis using g:Profiler.

Term name	Term ID	p_{adj}
KEGG root term	KEGG:00000	4.941×10^{-324}
Metabolic pathways	KEGG:01100	9.849×10^{-31}
Biosynthesis of secondary metabolites	KEGG:01110	4.873×10^{-12}
Plant hormone signal transduction	KEGG:04075	2.458×10^{-3}
Spliceosome	KEGG:03040	2.022×10^{-2}
Ribosome	KEGG:03010	3.665×10^{-1}
Ribosome biogenesis in eukaryotes	KEGG:03008	5.186×10^{-1}
Glutathione metabolism	KEGG:00480	5.186×10^{-1}
Arginine and proline metabolism	KEGG:00330	5.335×10^{-1}
Terpenoid backbone biosynthesis	KEGG:00900	9.450×10^{-1}

FIGURE 9
Enrichment analysis of 167 underexpressed differentially expressed genes (DEGs) in diploid *Solanum tuberosum* 'Wira Pasña' (CIP-704270), compared using the KEGG database. g:Profiler was used with significance at padj < 0.05.

inoculum, type strain (virulence, aggressiveness), inoculation method, and environment interaction are very important for the performance of the experiment (Ali et al., 2014; Gao and Bradeen, 2016; Izarra and Lindqvist-Kreuze, 2016: Stefańczyk et al., 2017). Considering the background of the POX-067 strain, this experiment used examined and referenced concentrations, as well as known inoculation methodology (Izarra and Lindqvist-Kreuze, 2016). The time of inoculation (before flowering) and sampling (48 hai) used in this study were based on available literature (Wang et al., 2004; Evers et al., 2005; Riva Romero, 2010; Monsalve-Fonnegra et al., 2012; Ali et al., 2014; Stefańczyk et al., 2017), which indicated high differential expression after evaluating *S. phureja* inoculated with *P. infestans* between 36 and 48 hai.

Term name	Term ID	p_{adj}
KEGG root term	KEGG:00000	4.941×10^{-324}
Metabolic pathways	KEGG:01100	2.123×10^{-31}
Biosynthesis of secondary metabolites	KEGG:01110	3.027×10^{-12}
Plant-pathogen interaction	KEGG:04626	3.114×10^{-1}
Nucleotide excision repair	KEGG:03420	9.928×10^{-1}
Folate biosynthesis	KEGG:00790	9.928×10^{-1}
DNA replication	KEGG:03030	9.928×10^{-1}
RNA polymerase	KEGG:03020	9.928×10^{-1}
Zeatin biosynthesis	KEGG:00908	9.928×10^{-1}
Inositol phosphate metabolism	KEGG:00562	9.928×10^{-1}
Homologous recombination	KEGG:03440	9.928×10^{-1}
Terpenoid backbone biosynthesis	KEGG:00900	9.928×10^{-1}
Mismatch repair	KEGG:03430	9.928×10^{-1}

FIGURE 10
Enrichment analysis of 136 differentially expressed genes (DEGs) overexpressed in diploid *Solanum tuberosum* 'Wira Pasña' (CIP-704270) compared using the KEGG database. g:Profiler was used with significance at Padj < 0.05.

The interactions between the *S. tuberosum* accessions of cultivars 'Wira Pasña' and 'Sumaq Perqa' with the pathogen *P. infestans* (Figure 1) confirmed them as resistant and susceptible, respectively, as previously known (Pérez et al., 2014). *Solanum tuberosum* 'Yungay' was susceptible (Wulff et al., 2007) and is frequently used as a control for the evaluation of host plant resistance to LB (Lindqvist-Kreuze et al., 2014; Jiang et al., 2018), as well as for pathogenicity research (Lindqvist-Kreuze et al., 2020). This cultivar made it possible to validate the screening results.

In the present work, paired reads have been used to improve the accuracy of the mapping to the reference genome. The mapping alignment of unique sequences was 88%, which is higher than previously known (Gao and Bradeen, 2016). Approximately 92% of reads landing in the reference genome are DNA sequences that correspond to the species, thus indicating that the reference genome used in this study is of good quality. The remaining 8% of unmapped sequences suggests that the gene space of the reference genome used in this study does not contain this group of genes specific to diploid accessions due to species differences. The reference genome is more closely related to *S.phureja* (Xu et al., 2011).

The location and distribution of the 12 bookcases in the PCA coordinate plane were opposite (Figure 2), which indicates that the resistant and susceptible cultivars with and without *P. infestans* inoculation had an adequate variance within and between each factor. Normalization is used to correct for differences in the sizes of raw data libraries so that this does not influence the calculations and differential analyses. Some authors (Li et al., 2012) proposed a normalization method that assumes a Poisson count model and estimates the sequencing depth. However, we used *edgeR*, which models and fits a negative binomial distribution. To normalize, the *edgeR* software, such as *DESeq*, calculates the ratios between the gene counts in each sample, and the geometric mean of the gene counts in all samples, while the library size is estimated as the median of the ratios between genes (Anders and Huber, 2010). Different available results (Dillies et al., 2013; Finotello et al., 2014) indicate that the *edgeR* and *DESeq* methods are the most suitable for

10.3389/fpls.2023.1210046

TABLE 2 List of 20 candidate genes related to host plant resistance in diploid *S. tuberosum* 'Wira Pasña' (CIP-704270) to *P. infestans* isolate POX-067, at 48 hours after inoculation (hai).

Underexpressed genes				
Gene ID	Description	Log_2FC 00hai	Log_2FC 48hai	FDR
PGSC0003DMG402016602	Tospovirus resistance protein C	8.20	-11.12	1×10^{-4}
PGSC0003DMG401007871	NBS-LRR protein	2.22	-2.35	1×10^{-5}
PGSC0003DMG400007605	NBS-LRR protein	6.03	-6.95	5×10^{-5}
PGSC0003DMG400002217	NBS-LRR resistance protein	5.57	-5.67	5×10^{-5}
PGSC0003DMG400007870	NBS-LRR protein	5.28	-5.93	7×10^{-5}
PGSC0003DMG400016372	Resistance protein PSH-RGH7	6.70	-7.64	2×10^{-5}
PGSC0003DMG401012062	Bacterial spot disease resistance protein 4	8.39	-8.48	7×10^{-5}
PGSC0003DMG400018462	Disease resistance protein	3.77	-4.26	5×10^{-6}
PGSC0003DMG400005471	RLK_6-phosphoglucono-lactonase	3.71	-2.38	4×10^{-5}
PGSC0003DMG402024222	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase	3.75	-4.67	1×10^{-5}
Overexpressed genes				
Gene ID	Description	Log_2FC 00hai	Log_2FC 48hai	FDR
PGSC0003DMG400008596	Cc-nbs-lrr resistance protein	-8.29	8.47	3×10^{-5}
PGSC0003DMG400025547	Late blight resistance protein	-11.59	11.59	8×10^{-4}
PGSC0003DMG400003380	Late blight resistance protein homolog R1A-4	-3.88	4.08	1×10^{-6}
PGSC0003DMG400005590	Disease resistance protein	-7.54	7.32	2×10^{-4}
PGSC0003DMG400002427	Bacterial spot disease resistance protein 4	-4.02	3.03	1×10^{-4}
PGSC0003DMG400031277	F-box domain-containing protein	-10.0	10.01	2×10^{-4}
PGSC0003DMG400036554	F-box protein	-9.48	9.48	1×10^{-4}
PGSC0003DMG400031279	F-box domain-containing protein	-10.15	10.15	1×10^{-4}
PGSC0003DMG400008593	Flavonoid glucoyltransferase UGT73E2	-4.25	5.90	1×10^{-4}
PGSC0003DMG400017234	Auxin-induced protein 5NG4	-2.49	2.90	1×10^{-4}

It is important to highlight that of the 103 genes (of the 303 DEGs) with unknown function from 'Wira Pasña', 55 of them were underexpressed and the remaining 48 genes were overexpressed in the presence of P. infestans at 48 hai. They were annotated using g:Profiler and the PGSC_GENES database.

normalizing the size of the library (Lovén et al., 2012; Liu et al., 2015; Rapaport et al., 2015) because they use direct counts for calculating the mean and the variance. In this study the sample size was small (3 samples per condition), and the data were normalized (Figure 2) and analyzed with *edgeR*. When the sample size is small (2 to 5 samples per condition), the best results are generally obtained with *DESeq* and *edgeR*. (Soneson and Delorenzi, 2013). The correlation analyses of the expressed DEGs were based on the statistical power of edgeR, demonstrated in the heat maps and Manhattan plots, and are as valid as the weighted correlation networks analysis (WGCNA) used by Esposito et al. (2019); Esposito et al. (2021)

In this study of the transcriptome of StAG cultivar accessions stressed with *P. infestans* at 48 hai, there were significantly underexpressed genes in the presence of the pathogen, while the secondary metabolite biosynthesis pathways were highly significant for both the underexpressed and overexpressed genes. These results were similar to those observed in a transgenic *S. tuberosum* transcriptome at 0, 6, and 24 hai in the foliage (Gao and Bradeen, 2016), as well as in transcriptomic research in *Citrus* stressed with *P. parasitica* (Naveed et al., 2019). It is important to highlight that there were 103 genes with unknown functions among the 303 DEGs. They are probably genes of 'Wira Pasña', of which 55 of them were underexpressed and the other 48 were overexpressed in the presence of *P. infestans* at 48 hai.

The oomycete *P. infestans* is an extraordinarily virulent and adaptable pathogen, secreting several effector proteins that modulate the host's innate immunity (Vleeshouwers et al., 2008). All known avirulence genes belong to the RXLR effector classes, and encode secretory proteins with the RXLR motif for translocation into cells (Vleeshouwers et al., 2011). The *P. infestans POX 067* isolate used in the present study expresses the avirulence genes *Avr8* and *Avr9* (Izarra and Lindqvist-Kreuze, 2016), thereby generating effector proteins that were possibly identified in the first instance by

the overexpressed genes at 0 hai with isolate *POX 067* (Table 2, highlighting 10 overexpressed genes at 0 hai) that then activate overexpressed *R* genes at 48 hai in a second instance (Table 2, highlighting 10 genes with high $\log_2$ FC). This was demonstrated by Lin et al. (2022), with the widely conserved *Phytophthora* RXLR-WY effector, which was first recognized through the *Rpi-amr3* gene that activated the expression of resistance genes. These findings suggest a new way to redeploy *R* genes that recognize known effectors against *P. infestans*.

We found 303 genes (Supplementary Table 4), of which 136 overexpressed genes related to LB resistance. Considering the difference in expression level (Log_2FC) from 0 to 48 hai, in addition to selecting only genes that code for resistance, 20 DEGs were found to give host plant resistance to *P. infestans* in accessions of diploid cultivars of S. *tuberosum* Andigenum Group. Of these, 10 genes (Table 2) that were overexpressed at 0 hai ($\log_2$FC, range, 2.22 to 8.39), were underexpressed at 48 hai ($\log_2$FC ranging from -2.35 to -11.12) in the presence of the pathogen. In contrast, another 10 genes that were underexpressed at 0 hai ($\log_2$FC ranging from -2.49 to -11.59) were overexpressed in the presence of the pathogen at 48 hai ($\log_2$FC ranging from 2.90 to 11.59). Similar results were noted after evaluating the gene expression profile of tetraploid S. *tuberosum* SD20 for its resistance against *P. infestans* (Yang et al., 2019). Due to their differential expressions, GO analysis and KEEG enrichment reported 10 overexpressed genes closely related to LB resistance: PGSC0003DMG400008596 (Cc-nbs-lrr resistance protein), PGSC0003DMG400025547 (late blight resistance protein), PGSC0003DMG400003380 (Late blight resistance protein homolog R1A-4), PGSC0003DMG400005590 (disease resistance protein), PGSC0003DMG400002427 (bacterial spot disease resistance protein 4), PGSC0003DMG400031277 (F-box domain-containing protein), PGSC0003DMG400036554 (F-box protein), PGSC0003DMG400031279 (F-box domain-containing protein), PGSC0003DMG400008593 (flavonoid glucoyltransferase UGT73E2), and PGSC0003DMG400017234 (auxin-induced protein 5NG4).

Plants have evolved efficient defense mechanisms based on a molecular immunity system through *R* genes that allow them to resist different pathogens including bacteria, fungi, oomycetes, viruses, and nematodes. Most *R* genes (approximately 80%) encode nucleotide binding sites (NBS) and comprise three domains that allow them to be classified into TIR-NBS-LRR (TNL), CC-NBS-LRR (CNL), or RPW8-NBSLRR (RNL), of which the C-terminal LRR domain exhibits high diversity and has been associated with pathogen recognition (Qian et al., 2017). At the beginning of the infection with *P. infestans*, underexpressed genes increased, especially NBS-LRR (nucleotide binding site, leucine-rich repeats), which could indicate that at the beginning, many underexpressed genes are required to generate resistance in the late stages of infection (Yang et al., 2018). The probable reason for this is that many of the resistance genes are constitutively expressed. Thus, there were four NBS-LRR genes in 'Wira Pasña' that were underexpressed at the beginning of the infection. Indeed, in this study, the resistant cultivar overexpressed a CC-NBS-LRR gene at 48 hai, thereby highlighting its classification as a candidate gene together with two other LB protein-specific resistance genes to *P. infestans*.

The genes of the NBS-LRR domain are related to R proteins (Lozano et al., 2012). In *S. tuberosum* the highest number of NBS-LRR genes are on chromosomes 4 and 11 (15% of mapped genes), with a lower number on chromosome 3 (1%) (Jupe et al., 2013). *R* proteins are responsible for the recognition of pathogen proteins called effectors, which allow the host hypersensitivity reaction, thereby preventing infection. *PR* (pathogenesis-related proteins) genes are distributed into 17 families (Shi et al., 2019). They have been described, isolated, and characterized in *S. phureja* in relation to host plant resistance to *P. infestans* in the field (Evers et al., 2005; Evers et al., 2006). *PR-1* gene *mRNA* has also been evaluated in resistant and susceptible potato cultivars to confirm the role of *PR-1* for host plant resistance to *P. infestans* in the field. In the present study, we report six *R* genes, five from the NBS-LRR domain and one from the RLK domain, that were stimulated by *P. infestans*. Of the five, four genes from the NBS-LRR domain were underexpressed at 48 hai and one from the CC-NBS domain (*PGSC0003DMG400008596*) was overexpressed at 48 hai (Table 2). These genes seem to allow the hypersensitivity in accessions of diploid cultivars of S. *tuberosum* Andigenum Group, thus preventing infection by *P. infestans*. The *PGSC0003DMG400005471* gene of the RLK type noted in this study confirms previous results in tetraploid potato (Yang et al., 2018). Hence, RLKs are important genes for plant-pathogen recognition receptors in the early stages of the attack of *P. infestans* in potato.

It was observed in 'Wira Pasña' that an enzyme related to MAPK (a brassinosteroid associated with kinase receptors) was overexpressed at the beginning of the infection. Brassinosteroids are considered one of the most important steroid hormones with many roles in plants, e.g., adaptation of plants to different stress factors and successful stress tolerance (Khan et al., 2022). In the early stages of infection by *P. infestans*, DEGs mainly encode defense enzymes to increase their expression, thus triggering a phosphorylation cascade through protein kinase (MAPK) that directly suppresses the pathogen as the first line of defense (Yang et al., 2018). RLKs that are in the plasmatic membrane in plants, positively regulate the innate immunity of the plant (Yang et al., 2018). In this study, one RLK gene was greatly overexpressed at 48 hai, thus suggesting that these RLKs are important recognition receptors in the early stages of host plant resistance to LB.

Transcription factors play a crucial role in resistant plants by modulating the transcription of resistance-related genes for the binding of specific DNA sequences to their promoter regions (Li et al., 2017). F-box proteins constitute a family of diverse transcription factors in prokaryotes and eukaryotes and regulate the cardinal biological processes in plants including growth and development, cellular protein degradation, and response to biotic or abiotic stress (Zhao et al., 2017), in addition to being involved in the biosynthesis of anthocyanins (Zhang et al., 2020). In potato, 11 transcription factors are strongly upregulated, including WRKY1, WRKY3, and WRKY5, thereby suggesting positive regulation of these WRKYs in the resistance response to *P. infestans* infection

(Yang et al., 2018). 'Wira Pasña' had an overexpression of three genes of F-box proteins at 48 hai, which suggests that their functional role also occurs under such biotic stress. This result strongly supports the selection of these genes for host plant resistance to *P. infestans*. In addition, further functional research on the F-box genes will be necessary to explore their roles in the response to *P. infestans*-induced stress in potato.

UDP-glycose flavonoid glycosyl-transferase (UFGT) catalyzes metabolites closely related to the biosynthesis of flavonoids and anthocyanins, which are stimulated by stressors such as low temperatures (10°C or less) (Yu et al., 2017). In this study, there was an overexpression at 48 hai of the flavonoid *UGT73E2* gene, which agrees with a previous finding where the temperature for the progress of *P. infestans* infection was low (Yu et al., 2017). It has also been found that an increased expression of DEGs at 24 hai encodes defense enzymes and disease-related proteins, including flavonoid-3-hydroxylase (Yang et al., 2018). Similarly, in 'Wira Pasña', there was increased expression of the flavonoid *UGT73E2* gene at 48 hai due to the inductive effect of the pathogen, thereby triggering phosphorylation cascades, regulating the expression of genes downstream related to resistance proteins against the pathogen.

To date, more than 70 *Rpi* genes have been identified and mapped in 32 *Solanum* species. High-level host plant resistance has been found in several diploid Mexican species, including *Solanum bulbocastanum* Dunal. and *S. pinnatisectum* Dunal., most of which have been derived from wild and tuber-bearing species, four from Peru, but none from the diploid *S. tuberosum* Andigenum Group (Paluchowska et al., 2022). This work is the first report highlighting 20 major *R* genes, including highly expressed *P. infestans* effector detectors at 0 hai and other resistance candidate *R* genes at 48 hai and more than 300 minor genes (Supplementary Table 6) related to horizontal and durable resistance. The combined use of *Rpi* genes that recognize the essential and conservative effectors of *P. infestans* and pyramiding *Rpi* genes may assist in achieving long-lasting, broad-spectrum resistance to late blight.

The goal to develop resistance in potato crops to the pathogen may not be easy to accomplish because of the evolution of pathogens to overcome, with time, any resistance conferred by *R* genes (Majeed et al., 2022), thereby suggesting that more systematic approaches are consistently required. They may use wild *Solanum* germplasm sources for resistance genes (Nelson et al., 2018). Although difficult to be accomplished, quantitative host plant resistance conferred by several minor genes seems to be a better strategy for achieving durable host plant resistance to *P. infestans* (Du et al., 2015; Nelson et al., 2018). In this investigation, we found 10 major *R* genes (high Log_2FC) at 48 hai and more than 300 minor genes (Supplementary Table 6) related to quantitative trait loci for late blight resistance in the *S. tuberosum* Andigenum Group transcriptome, thus strengthening our knowledge of resistance induction in potato against the pathogen *P. infestans*. The identification of new sources of genes with durable resistance to *P. infestans* in native potatoes such as the diploid *S. tuberosum* Andigenum Group would contribute to the improvement of this tuber crop.

The *PGSC0003DMG400025547* and *PGSC0003DMG400003380* genes encode host plant resistance to *P. infestans* in diploid StAG accessions. The first gene encodes LB resistance proteins and the other an LB resistance protein homologous to *R1a-4*. These two genes were underexpressed (log_2FC -11.59 and -3.88, respectively) when diploid StAG accession was not inoculated at 0 hai, but they were overexpressed (log_2FC 11.59 and 4.08, respectively) at 48 hai by *P. infestans* stimulation. These diploid StAG accession genes could be different from the *Rpi-vnt1 genes* previously known (*PGSC0003DMG400024364* and *PGSC0003DMG400024363*), (van der Vossen et al., 2003), e.g., *Rpi* ancestral genes of the wild potato *S. bulbocastanum* that confer broad resistance to *P. infestans*. In the context of the Peruvian Andes, where the EC-1 clonal lineage of *P. infestans* predominates (Perez et al., 2001), the potential use of the *Rpi-vnt1, Avr-blb1 y Avr-blb2* genes and of the two specific of diploid StAG accessions (PGSC0003DMG400025547 and PGSC0003DMG400003380) facilitates breeding for host plant resistance to *P. infestans* along with the genes for the NBS-LRR proteins found in this study (FDR, range 1×10^{-4} to 1×10^{-6}).

Conclusion

This leaf transcriptome research was able to identify candidate genes that provide host plant resistance to LB in StAG accessions. There were 303 genes differentially expressed by the pathogen stimulus, of which 10 overexpressed genes were defined as *R* genes for the defense response: PGSC0003DMG400008596, PGSC0003DMG400025547, PGSC0003DMG400003380, PGSC0003DMG400005590, PGSC0003DMG400002427, PGSC0003DMG400031277, PGSC0003DMG400036554, PGSC0003DMG400031279, PGSC0003DMG400008593y, and PGSC0003DMG400017234. Of these, four genes (*PGSC0003DMG400008596*, *PGSC0003DMG400025547*, *PGSC0003DMG400003380*, and *PGSC0003DMG400005590*) encode for NBS-LRR proteins. These results provide further evidence that diploid cultivated potatoes are a source of host plant resistance genes to the pathogen *P. infestans*.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://www.ncbi.nlm.nih.gov/, Bioproject PRJNA861684 and PRJNA861685.

Author contributions

GC, conception and design-execution of all stages of the project. RB, project design. WP, design-execution of phenotyping and editing manuscript drafts. EN, Design-execution RNA extraction/preservation, sequencing, and being co-corresponding author. RO, supervision, writing manuscript and editing drafts, and being co-corresponding author. All co-authors read and agreed on the text of

the article. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Camisea Socioeconomic Development Fund (Law No. 28451-MEF-PERU) of the Universidad Nacional de San Cristóbal de Huamanga (UNSCH), Ayacucho, Peru.

Acknowledgments

The authors thank Dr. Gisella Orjeda for providing us with logistical support in her laboratory at Universidad Peruana Cayetano Heredia [UPCH], Lima, Perú. To M.Sc. Lourdes Tapia (Universidad Nacional Agraria La Molina [UNALM], Lima, Perú), for the support *in vitro* multiplication of accessions. They acknowledge KEGG copyright permission 221648 for using reproduced database images in this article. The work was carried out through inter-institutional collaboration with the International Potato Center (CIP, Lima, Perú), UNALM, UPCH, UNSCH, and the Swedish University of Agricultural Sciences.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1210046/full#supplementary-material

References

Ali, A., Alexandersson, E., Sandin, M., Resjö, S., Lenman, M., Hedley, P., et al. (2014). Quantitative proteomics and transcriptomics of potato in response to *Phytophthora infestans* in compatible and incompatible interactions. *BMC Genomics* 15, 497. doi: 10.1186/1471-2164-15-497

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11, R106. doi: 10.1186/gb-2010-11-10-r106

Avrova, A. O., Venter, E., Birch, P. R. J., and Whisson, S. C. (2003). Profiling and quantifying differential gene transcription in *Phytophthora infestans* prior to and during the early stages of potato infection. *Fungal Genet. Biol.* 40, 4–14. doi: 10.1016/s1087-1845(03)00063-x

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Carreño, H. (2017) *Identificación de genes regulados diferencialmente en respuesta a la infección por Globodera pallida en una variedad de papa resistente y otra susceptible al nemátodo* (Lima, Perú: Universidad Peruana Cayetano Heredia). Available at: http://repositorio.upch.edu.pe/handle/upch/752 (Accessed December 9, 2019).

Cevallos Hidalgo, P. A. (2015). *Tamizaje molecular de genotipos nativos de papa (Solanum spp.) en busca de QTLs de resistencia a Phytophthora infestans. Ing. Agr. Thesis*. Universidad Central del Ecuador, Quito, Ecuador. Available at: http://www.dspace.uce.edu.ec/handle/25000/6735 [Accessed December 1, 2019].

Chang-Kee, A. J. (2016) *Caracterización de seis poblaciones de Solanum pimpinellifolium de las regiones Piura y Lima respecto de su resistencia a Phytophthora infestans* (Lima, Perú: Universidad Peruana Cayetano Heredia). Available at: http://repositorio.upch.edu.pe/handle/upch/327 (Accessed December 9, 2019).

De la Cruz, G., MIranda, T. Y., Blas, R. H., Neyra, E., and Orjeda, G. (2020). Simple sequence repeat-based genetic diversity and analysis of molecular variance among on-farm native potato landraces from the influence zone of Camisea Gas Project, northern Ayacucho, Peru. *Amer J. Potato Res.* 97, 143–161. doi: 10.1007/s12230-020-09763-7

Dillies, M.-A., Rau, A., Aubert, J., Hennequet-Antier, C., Jeanmougin, M., Servant, N., et al. (2013). A comprehensive evaluation of norMalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief Bioinf.* 14, 671–683. doi: 10.1093/bib/bbs046

Dobin, A., and Gingeras, T. R. (2015). Mapping RNA-seq reads with STAR. *Curr. Protoc. Bioinf.* 51, 11.14.1–11.14.19. doi: 10.1002/0471250953.bi1114s51

Du, J., Verzaux, E., Chaparro-Garcia, A., Bijsterbosch, B., Keizer, L., Zhou., J., et al. (2015). Elicitin recognition confers enhanced resistance to *Phytophthora infestans* in potato. *Nat. Plants* 1. doi: 10.1038/nplants.2015.34

Enciso-Maldonado, G. A., Lozoya-Saldaña, H., Colinas-Leon, M. T., Cuevas-Sanchez, J. A., Sanabria-Velázquez, A. D., Bamberg, J., et al. (2022). Assessment of wild *Solanum* species for resistance to *Phytophthora infestans* (Mont.) de Bary in the Toluca Valley, Mexico. *Amer J. Potato Res.* 99, 25–39. doi: 10.1007/s12230-021-09856-x

Eshraghi, L., Anderson, J., Aryamanesh, N., Shearer, B., McComb, J., Hardy, G. E., et al. (2011). Phosphite primed defence responses and enhanced expression of defence genes in *Arabidopsis thaliana* infected with *Phytophthora cinnamomi*. *Plant Pathol.* 60, 1086–1095. doi: 10.1111/j.1365-3059.2011.02471.x

Espinoza, M.-L. R. (2017) *Análisis transcriptómico de la respuesta a heladas en papas nativas* (Lima, Perú: Universidad Peruana Cayetano Heredia). Available at: http://repositorio.upch.edu.pe/handle/upch/1037 (Accessed December 9, 2019).

Espinoza, N., Lizarraga, R., Silva Rodriguez, D., Buitron, F., Bryan, J., and Dodds, J. H. (1989)Tissue culture, micropropagation, conservation and export of potato germplasm. In: *CIP research guide*. (Lima, Perú: International Potato Center). Available at: https://scholar.google.com/scholar_lookup?title=Tissue+culture%2C+micropropagation%2C+conservation+and+export+of+potato+germplasm&author=Espinoza%2C+N.&publication_year=1989 (Accessed April 12, 2023).

Esposito, S., Aversano, R., Bradeen, J., D'Amelia, V., Villano, C., and Carputo, D. (2021). Coexpression gene network analysis of cold-tolerant Solanum commersonii reveals new insights in response to low temperatures. *Crop Sci.* 61, 3538–3550. doi: 10.1002/csc2.20473

Esposito, S., Barteri, F., Casacuberta, J., Mirouze, M., Carputo, D., and Aversano, R. (2019). LTR-TEs abundance, timing and mobility in *Solanum commersonii* and *S. tuberosum* genomes following cold stress conditions. *Planta* 250, 1781–1787. doi: 10.1007/s00425-019-03283-3

Evers, D., Ghislain, M., Hoffmann, L., Hausman, J. F., and Dommes, J. (2006). A late blight resistant potato plant overexpresses a gene coding for α-galactosidase upon infection by *Phytophthora infestans*. *Biol. Plant* 50, 265–271. doi: 10.1007/s10535-006-0017-1

Evers, D., Schweitzer, C., Nicot, N., Gigliotti, S., Herrera, M., Hausman, J. F., et al. (2005). Two *PR-1* loci detected in the native cultivated potato *Solanum phureja* appear differentially expressed upon challenge by late blight. *Physiol. Mol. Plant Pathol.* 67, 155–163. doi: 10.1016/j.pmpp.2005.12.003

Finotello, F., Lavezzo, E., Bianco, L., Barzon, L., Mazzon, P., Fontana, P., et al. (2014). Reducing bias in RNA sequencing data: a novel approach to compute counts. *BMC Bioinf.* 15, S7. doi: 10.1186/1471-2105-15-S1-S7

Forbes, G., Pérez, W., and Andrade, J. (2014) *Field assessment of resistance in potato to Phytophthora infestans: International cooperators guide* (Lima, Perú: International Potato Center). Available at: https://cgspace.cgiar.org/handle/10568/65180 (Accessed December 10, 2019).

Fry, W. E., Birch, P. R. J., Judelson, H. S., Grünwald, N. J., Danies, G., Everts, K. L., et al. (2015). Five reasons to consider *Phytophthora infestans* a reemerging pathogen. *Phytopathology* 105, 966–981. doi: 10.1094/PHYTO-01-15-0005-FI

Gao, L., and Bradeen, J. M. (2016). Contrasting potato foliage and tuber defense mechanisms against the late blight pathogen *Phytophthora infestans*. *PloS One* 11, e0159969. doi: 10.1371/journal.pone.0159969

Gollery, M. (2006). An assessment of the current state of bioinformatics education. *Bioinformation* 1, 247. doi: 10.6026/97320630001247

Haverkort, A., Boonekamp, P., Hutten, R., Jacobsen, E., Lotz, B., Kessel, G., et al. (2016). Durable late blight resistance in potato through dynamic varieties obtained by cisgenesis: Scientific and societal advances in the DuRPh Project. *Potato Res.* 59, 35–66. doi: 10.1007/s11540-015-9312-6

Izarra, M., and Lindqvist-Kreuze, H. (2016). Expression of RXLR effectors in the EC-1 clonal lineage of *Phytophthora infestans* in Peru. *Rev. Peruana Biol.* 23, 293–299. doi: 10.15381/rpb.v23i3.12864

Jiang, R., Li, J., Tian, Z., Du, J., Armstrong, M., Baker, K., et al. (2018). Potato late blight field resistance from QTL dPI09c is conferred by the NB-LRR gene R8. *J. Exp. Bot.* 69, 1545–1555. doi: 10.1093/jxb/ery021

Jupe, F., Witek, K., Verweij, W., Sliwka, J., Pritchard, L., Etherington, G. J., et al. (2013). Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. *Plant J.* 76, 530–544. doi: 10.1111/tpj.12307

Kanehisa, M. (2019). Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* 28, 1947–1951. doi: 10.1002/pro.3715

Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M., and Ishiguro-Watanabe, M. (2023). KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* 51, D587–D592. doi: 10.1093/nar/gkac963

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27

Khan, M. T. A., Yusuf, M., Qazi, F., and Ahmad, A. (2022). *Brassinosteroids signalling: intervention with phytohormones and their relationship in plant adaptation to abiotic stresses* (Singapore: Springer). doi: 10.1007/978-981-16-5743-6

Law, C. W., Alhamdoosh, M., Su, S., Dong, X., Tian, L., Smyth, G. K., et al. (2018). RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. *F1000Res* 5, ISCB Comm J–1408. doi: 10.12688/f1000research.9005.3

Leggett, R. M., Ramirez-Gonzalez, R. H., Clavijo, B., Waite, D., and Davey, R. P. (2013). Sequencing quality assessment tools to enable data-driven informatics for high throughput genomics. *Front. Genet.* 4. doi: 10.3389/fgene.2013.00288

Lehsten, V., Wiik, L., Hannukkala, A., Andreasson, E., Chen, D., Ou, T., et al. (2017). Earlier occurrence and increased explanatory power of climate for the first incidence of potato late blight caused by *Phytophthora infestans* in Fennoscandia. *PloS One* 12, e0177580. doi: 10.1371/journal.pone.0177580

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, J., Lindqvist-Kreuze, H., Tian, Z., Liu, J., Song, B., Landeo, J., et al. (2012). Conditional QTL underlying resistance to late blight in a diploid potato population. *Theor. Appl. Genet.* 124, 1339–1350. doi: 10.1007/s00122-012-1791-0

Li, W., Zhu, Z., Chern, M., Yin, J., Yang, C., Ran, L., et al. (2017). A natural allele of a transcription factor in rice confers broad-spectrum blast resistance. *Cell* 170, 114–126.e15. doi: 10.1016/j.cell.2017.06.008

Lin, X., Olave-Achury, A., Heal, R., Pais, M., Witek, K., Ahn, H.-K., et al. (2022). A potato late blight resistance gene protects against multiple Phytophthora species by recognizing a broadly conserved RXLR-WY effector. *Mol. Plant* 15, 1457–1469. doi: 10.1016/j.molp.2022.07.012

Lindqvist-Kreuze, H., Gamboa, S., Izarra, M., Pérez, W., Correa, M. Y., Astete, A., et al. (2020). Population structure and host range of the potato late blight pathogen *Phytophthora infestans* in Peru spanning two decades. *Plant Pathol.* 69, 334–346. doi: 10.1111/ppa.13125

Lindqvist-Kreuze, H., Gastelo, M., Perez, W., Forbes, G., De Koeyer, D., and Bonierbale, M. (2014). Phenotypic stability and genome-wide association study of late blight resistance in potato genotypes adapted to the tropical highlands. *Phytopathology* 104, 624–633. doi: 10.1094/PHYTO-10-13-0270-R

Liu, Y., Lin-Wang, K., Deng, C., Warran, B., Wang, L., Yu, B., et al. (2015). Comparative transcriptome analysis of white and purple potato to identify genes involved in anthocyanin biosynthesis. *PloS One* 10, e0129148. doi: 10.1371/journal.pone.0129148

Lovén, J., Orlando, D. A., Sigova, A. A., Lin, C. Y., Rahl, P. B., Burge, C. B., et al. (2012). Revisiting global gene expression analysis. *Cell* 151, 476–482. doi: 10.1016/j.cell.2012.10.012

Lozano, R., Ponce, O., Ramirez, M., Mostajo, N., and Orjeda, G. (2012). Genome-wide identification and mapping of NBS-encoding resistance genes in *Solanum tuberosum* Group Phureja. *PloS One* 7, e34775. doi: 10.1371/journal.pone.0034775

Lurwanu, Y., Wang, Y.-P., Wu, E.-J., He, D.-C., Waheed, A., Nkurikiyimfura, O., et al. (2021). Increasing temperature elevates the variation and spatial differentiation of pesticide tolerance in a plant pathogen. *Evol. Appl.* 14, 1274–1285. doi: 10.1111/eva.13197

Magarey, R. D., Chappell, T. M., Trexler, C. M., Pallipparambil, G. R., and Hain, E. F. (2019). Social ecological system tools for improving crop pest management. *J. Integr. Pest Manag* 10, 2. doi: 10.1093/jipm/pmz004

Majeed, A., Siyar, S., and Sami, S. (2022). Late blight of potato: From the great Irish potato famine to the genomic era – An overview. *Hell Plant Prot J.* 15, 1–9. doi: 10.2478/hppj-2022-0001

Massa, A., Childs, K., and Buell, C. (2013). Abiotic and biotic stress responses in Group Phureja DM1-3 516 R44 as measured through whole transcriptome sequencing. *Plant Genome* 6, 14. doi: 10.3835/plantgenome2013.05.0014

Monsalve-Fonnegra, Z. I., Monsalve-Restrepo, M. I., Urrea-Trujillo, A. I., and Zapata, J. L. (2012). Expresión diferencial durante la interacción *Solanum tuberosum –Phytophthora infestans*. *Rev. Colombiana Biotecnología* 14, 77–92.

Naveed, Z. A., Huguet-Tapia, J. C., and Ali, G. S. (2019). Transcriptome profile of Carrizo citrange roots in response to *Phytophthora parasitica* infection. *J. Plant Interact.* 14, 187–204. doi: 10.1080/17429145.2019.1609106

Nelson, R., Wiesner-Hanks, T., Wisser, R., and Balint-Kurti, P. (2018). Navigating complexity to breed disease-resistant crops. *Nat. Rev. Genet.* 19, 21–33. doi: 10.1038/nrg.2017.82

Oliveros, J. C. (2016) *Venny 2.1.0*. Available at: https://bioinfogp.cnb.csic.es/tools/venny/ (Accessed December 9, 2019).

Ovchinnikova, A., Krylova, E., Gavrilenko, T., Smekalova, T., Zhuk, M., Knapp, S., et al. (2011). Taxonomy of cultivated potatoes (*Solanum* section Petota: Solanaceae). *Bot. J. Linn Soc.* 165, 107–155. doi: 10.1111/j.1095-8339.2010.01107.x

Paluchowska, P., Śliwka, J., and Yin, Z. (2022). Late blight resistance genes in potato breeding. *Planta* 255, 127. doi: 10.1007/s00425-022-03910-6

Perez, W. G., Gamboa, J. S., Falcon, Y. V., Coca, M., Raymundo, R. M., and Nelson, R. J. (2001). Genetic structure of Peruvian populations of *Phytophthora infestans*. *Phytopathology* 91, 956–965. doi: 10.1094/PHYTO.2001.91.10.956

Pérez, W., Ñahui, M., Ellis, D., and Forbes, G. A. (2014). Wide phenotypic diversity for resistance to *Phytophthora infestans* found in potato landraces from Peru. *Plant Dis.* 98, 1530–1533. doi: 10.1094/PDIS-03-14-0306-RE

Pradel, W., Hareau, G., Quintanilla, L., and Suárez, V. (2017). *Adopción e impacto de variedades mejoradas de papa en el Peru: Resultado de una encuesta a nivel naciona)* (Lima, Perú: Centro Internacional de la Papa). doi: 10.4160/9789290602118

Qian, L.-H., Zhou, G.-C., Sun, X.-Q., Lei, Z., Zhang, Y.-M., Xue, J.-Y., et al. (2017). Distinct patterns of gene gain and loss: Diverse evolutionary modes of NBS-encoding genes in three Solanaceae crop species. *G3* 7, 1577–1585. doi: 10.1534/g3.117.040485

Rapaport, F., Khanin, R., Liang, Y., Pirun, M., Krek, A., Zumbo, P., et al. (2015). Erratum to: Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biol.* 16, 261. doi: 10.1186/s13059-015-0813-z

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., et al. (2019). g: Profiler: a web server for functional enrichment analysis and conversions of gene list update). *Nucleic Acids Res.* 47, W191–W198. doi: 10.1093/nar/gkz369

Riva Romero, R. (2010) *Análisis de la expresión del Gen PR-1 meidante la técnica de PCR en tiempo real (RT -PCK) en tomate (Solanum lycopersicum) infectado con Phytophtora infestans* (Sangolqui, Ecuador: Escuela Politécnica del Ejército, Facultad de Ingeniería en Biotecnología). Available at: http://repositorio.espe.edu.ec/jspui/handle/21000/2640 (Accessed December 4, 2019).

Shi, J., Zhao, L., Yan, B., Zhu, Y., Ma, H., Chen, W., et al. (2019). Comparative transcriptome analysis reveals the transcriptional alterations in growth- and development-related genes in sweet potato plants infected and non-infected by SPFMV, SPV2, and SPVG. *IJMS* 20, 1012. doi: 10.3390/ijms20051012

Soneson, C., and Delorenzi, M. (2013). A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinf.* 14, 91. doi: 10.1186/1471-2105-14-91

Stefańczyk, E., Sobkowiak, S., Brylińska, M., and Śliwka, J. (2017). Expression of the potato late blight resistance gene *Rpi-phu1* and *Phytophthora infestans* effectors in the compatible and incompatible interactions in potato. *Phytopathology* 107, 740–748. doi: 10.1094/PHYTO-09-16-0328-R

UCLA (2017) *UCLA | Bioinformatics*. Available at: https://bioinformatics.ucla.edu/ (Accessed December 4, 2019).

van der Vossen, E., Sikkema, A., Hekkert, B., te, L., Gros, J., Stevens, P., et al. (2003). An ancient *R* gene from the wild potato species *Solanum bulbocastanum* confers broad-spectrum resistance to *Phytophthora infestans* in cultivated potato and tomato. *Plant J.* 36, 867–882. doi: 10.1046/j.1365-313x.2003.01934.x

Vleeshouwers, V. G. A. A., Raffaele, S., Vossen, J. H., Champouret, N., Oliva, R., Segretin, M. E., et al. (2011). Understanding and exploiting late blight resistance in the age of effectors. *Annu. Rev. Phytopathol.* 49, 507–531. doi: 10.1146/annurev-phyto-072910-095326

Vleeshouwers, V. G. A. A., Rietman, H., Krenek, P., Champouret, N., Young, C., Oh, S.-K., et al. (2008). Effector genomics accelerates discovery and functional profiling of potato disease resistance and *Phytophthora infestans* avirulence genes. *PloS One* 3, e2875. doi: 10.1371/journal.pone.0002875

Vleeshouwers, V. G. A. A., van Dooijeweert, W., Paul Keizer, L. C., Sijpkes, L., Govers, F., and Colon, L. T. (1999). A laboratory assay for *Phytophthora infestans* resistance in various *Solanum* species reflects the field situation. *Eur. J. Plant Pathol.* 105, 241–250. doi: 10.1023/A:1008710700363

Wang, X., El Hadrami, A., Adam, L., and Daayf, F. (2005). Genes encoding pathogenesis-related proteins PR-2, PR-3 and PR-9, are differentially regulated in potato leaves inoculated with isolates from US-1 and US-8 genotypes of *Phytophthora infestans* (Mont.) de Bary. *Physiol. Mol. Plant Pathol.* 67, 49–56. doi: 10.1016/j.pmpp.2005.09.009

Wang, X., Hadrami, A. E., Adam, L., and Daayf, F. (2004). US-1 and US-8 genotypes of *Phytophthora infestans* differentially affect local, proximal and distal gene expression of phenylalanine ammonia-lyase and 3-hydroxy, 3-methylglutaryl CoA reductase in potato leaves. *Physiol. Mol. Plant Pathol.* 65, 157–167. doi: 10.1016/j.pmpp.2005.01.003

Wu, E.-J., Wang, Y.-P., Shen, L.-L., Yahuza, L., Tian, J.-C., Yang, L.-N., et al. (2019). Strategies of *Phytophthora infestans* adaptation to local UV radiation conditions. *Evol. Appl.* 12, 415–424. doi: 10.1111/eva.12722

Wu, E.-J., Wang, Y.-P., Yahuza, L., He, M.-H., Sun, D.-L., Huang, Y.-M., et al. (2020). Rapid adaptation of the Irish potato famine pathogen Phytophthora infestans to changing temperature. *Evolutionary Appl.* 13, 768–780. doi: 10.1111/eva.12899

Wulff, E. G., Pérez, W., Nelson, R. J., Bonierbale, M., Landeo, J. A., and Forbes, G. A. (2007). Identification of stable resistance to *Phytophthora infestans* in potato genotypes evaluated in field experiments in Peru. *Exp. Agric.* 43, 353–363. doi: 10.1017/S0014479707004991

Xu, X., Pan, S., Cheng, S., Zhang, B., Mu, D., Ni, P., et al. (2011). Genome sequence and analysis of the tuber crop potato. *Nature* 475, 189–195. doi: 10.1038/nature10158

Yang, X., Guo, X., Yang, Y., Ye, P., Xiong, X., Liu, J., et al. (2018). Gene profiling in late blight resistance in potato genotype SD20. *Int. J. Mol. Sci.* 19, 1728. doi: 10.3390/ijms19061728

Yang, X., Liu, J., Xu, J., Duan, S., Wang, Q., Li, G., et al. (2019). Transcriptome profiling reveals effects of drought stress on gene expression in diploid potato genotype P3-198. *Int. J. Mol. Sci.* 20, 852. doi: 10.3390/ijms20040852

Yu, X., Zhu, Y., Fan, J., Wang, D., Gong, X., and Ouyang, Z. (2017). Accumulation of flavonoid glycosides and UFGT gene expression in mulberry leaves (*Morus alba* L.) before and after frost. *Chem. Biodivers* 14, e1600496. doi: 10.1002/cbdv.201600496

Zhang, L., Wang, S., Chen, Y., Dong, M., Fang, Y., Zhang, X., et al. (2020). Genome-wide identification of the F-box gene family and expression analysis under drought and salt stress in barley. *Phyton* 89, 229. doi: 10.32604/phyton.2020.10022

Zhao, Z., Zhang, G., Zhou, S., Ren, Y., and Wang, W. (2017). The improvement of salt tolerance in transgenic tobacco by overexpression of wheat F-box gene *TaFBA1*. *Plant Sci.* 259, 71–85. doi: 10.1016/j.plantsci.2017.03.010

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 18 September 2023
DOI 10.3389/fpls.2023.1159848

Check for updates

OPEN ACCESS

EDITED BY
Manuela Nagel,
Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Germany

REVIEWED BY
Jinzhong Yang,
Qingdao Agricultural University, China
Richard Dormatey,
Gansu Agricultural University, China

*CORRESPONDENCE
Xianping Li
lxianping@hotmail.com

†These authors have contributed equally to this work

RECEIVED 06 February 2023
ACCEPTED 24 August 2023
PUBLISHED 18 September 2023

CITATION
Liang S, Jiang W, Yang Y, Lu L, Whitworth JL, Novy RG, Bao L, Yin Z, Li Z, He P, Xu Y and Li X (2023) An evaluation based on the analytic hierarchy process and GGEbiplot on French fry potato genotypes in Yunnan, China.
Front. Plant Sci. 14:1159848.
doi: 10.3389/fpls.2023.1159848

COPYRIGHT
© 2023 Liang, Jiang, Yang, Lu, Whitworth, Novy, Bao, Yin, Li, He, Xu and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

An evaluation based on the analytic hierarchy process and GGEbiplot on French fry potato genotypes in Yunnan, China

Shumin Liang[1†], Wei Jiang[1†], Yan Yang[1†], Lili Lu[1], Jonathan L. Whitworth[2], Richard G. Novy[2], Lixian Bao[1], Ziyou Yin[1], Zhou Li[3], Pinggen He[4], Youxian Xu[5] and Xianping Li[1*]

[1]Industrial Crops Research Institute, Potato Engineering Technology Research Center of Yunnan, Yunnan Academy of Agricultural Sciences, Kunming, Yunnan, China, [2]Aberdeen Research and Extension Center, United State Department of Agriculture-Agricultural Research Service, Aberdeen, ID, United States, [3]Zhaotong Academy of Agricultural Sciences, Zhaotong, Yunnan, China, [4]Lijiang Academy of Agricultural Sciences, Lijiang, Yunnan, China, [5]Agricultural Technology Extension Center of Xuanwei, Xuanwei, China

A total of 33 potato (*Solanum tuberosum* L.) cultivars and breeding clones imported from the United States and two local cultivars (Yunshu 401 and Cooperation 88, CK) were planted and evaluated. To determine their suitability for processing into French fries at five locations (e1-e5) in Yunnan Province, China, we developed a comprehensive evaluation system using the analytical hierarchy process (AHP). Eleven evaluation indicators for French fry quality, yield, and agronomic characteristics with a relative importance (weight coefficients) of 0.483, 0.301 and 0.216, respectively, were used to analyze the 35 potato genotypes (designated g1-g35).The genotypes were ranked and the results revealed that (1) on the average, the 33 potato genotypes imported from the United States showed a lower performance compared to the local cultivars. Compared with the CK, they were classified as not vigorous (Mean 5.11 vs CK 7.75), matured earlier (Mean 5.79 vs CK 1.70), and had a low resistance to late blight (Mean 3735.59 vs CK 1418.55), requiring the use of fungicides to control the disease at the five trial locations. (2) The US cultivar 'Defender' (g3) ranked in the top six at all five test locations because it had higher yield (29.56 t h m^{-2}), better fry quality (4.64), higher dry matter content (20.41%), better tuber length/width ratio (1.99), longer tubers (13.57cm), stronger plant vigor (7.17) and higher resistance to late blight (AUDPC: 3134.2). (3) By using GGEbiplot analysis, superior genotypes with high and stable yields were g3 and 'Echo Russet' (g33). 'Yunshu 401' (g34) and 'Yukon Gem' (g4) had high but not stable yields. The ideal test environments and hence experimental locations were Luquan (LQ, e2) and Lijiang (LJ, e4) which resulted in the best discrimination between genotypes among the five experimental locations in Yunnan. Overall, the developed evaluation system based on AHP and GGEbiplot analysis including 11 evaluation indicators for French fry quality, yield and agricultural traits can be a model for evaluation and promotion of new French fry cultivars, and evaluating and selecting the test location.

KEYWORDS

potato, French fries quality, evaluation system, genotype-environment interactions, the analytic hierarchy process, Yunnan-China

1 Introduction

China is the largest producer of potatoes (*Solanum tuberosum* L.) in the world with a production of 94.4 million tons in 2021, which accounted for 25.1% of the world's total yield that year (FAO, 2023). Owing to increased urbanization and a preference for fast food, French fry is one of the primary modes of consumption of potato in China (Liu et al., 2017; Kaur et al., 2021). Frozen French fries are popular all over the world and sold in many supermarkets. In addition, the French fries in fast food restaurants are popular with children and adults.

Currently, only a few cultivars are suitable for processing into French fries in China due to the lack of potato germplasm resources, narrow genetic backgrounds and close genetic relationships (Xu and Jin, 2017). 'Russet Burbank' from the US and 'Shepody' from Canada are still the primary cultivars of raw potatoes to produce French fries in China. The import French fry potato genotypes from the US, particularly the Pacific Northwest, can greatly enrich the germplasm resources of potato in China (Jansky et al., 2009). However, the regions of the Pacific Northwest are generally arid, irrigated, and in high latitudes that are completely different from those in the low latitude plateau of Yunnan Province with adequate rainfall. It is unclear whether those imported potato cultivars are adapted to the growing environments in Yunnan.

Analytic hierarchy process (AHP) and GGE biplot can support the selection of suitable cultivars for Yunnan. AHP can evaluate germplasm resources (Wang et al., 2021; Zhao et al., 2022), screening new crop cultivars (Buciumeanu et al., 2021), and, therefore, be included in the decision-making of cultivars selection (Rahman and Saha, 2008; Ramírez-García et al., 2015). It uses weighting factors as input data in terms of their importance, ranks them and estimates the overall weight of individual criteria or parameter (Morandi et al., 2020; Singha et al., 2020; Yao et al., 2022). The GGE biplot combines the genotype effect (G) with genotype by environmental interaction effect (GE) and is widely used to integrate analysis of the various stability parameters to select stable and high-yielding germplasm resources in multiple environments (Yan et al., 2000; Ruswandi et al., 2022). Therefore, GGE biplot may overcome limitations of single indexes or tuber quality of French fries alone and indicate the overall advantages and disadvantages of potato genotypes (Abong et al., 2009; Waxman et al., 2018; Sawicka et al., 2021). However, there is no definition of the importance and the weight of evaluation indicators including lower and upper limits. Furthermore, there are also no standard field agronomic traits used to select potato cultivars for French fries processing. Thus it is urgent to select appropriate criteria and indicators for field agronomic traits and the quality of French fries.

The aim of this study was to construct a comprehensive evaluation system of potato genotypes to screen 33 imported and 2 local genotypes for French fry quality, yield and agricultural traits in order to agree on cultivars planting decisions in Yunnan. Furthermore, these traits were combined, and quality, yield and agronomic traits of the potato genotypes were evaluated by using AHP and GGEbiplot and results used to identify suitable cultivars for the Yunnan Province.

2 Materials and methods

2.1 Experimental genotypes

A total of 33 genotypes were imported from the US, and they were numbered in sequence as g1 to g33. In addition, 'Yunshu 401' (g34) and 'Cooperation 88' (g35,CK) were used as the local controls (Supplementary Table 1).

2.2 Experimental sites

The field experiments were conducted over a two-year period from May to October 2019 at three potato experimental stations: e1: Huize (HZ), e2: Luquan (LQ), and e3-1: Zhaotong (ZT), and from April to November 2020 at five potato experimental stations e1, e2, ZT (e3-2); and e4: Lijiang (LJ); and e5: Xuanwei (XW) in Yunnan Province, south western China (Supplementary Figure 1). The stations were typical low latitude, monsoon, and had vertical climatic (based on variable altitude) zones. The altitude of the spring potato planting area is generally above 1,900 m. The rainfall is generally between 250-650 mm, and the average temperature is generally approximately 15°C during the growth period. The geographic and meteorological data of the experimental sites were obtained from weather stations at the study sites (Table 1).

2.3 Experimental design

In vitro pathogen-free plantlets of the US cultivars and breeding clones were provided by the University of Idaho, Seed Potato Germplasm Program (Moscow, ID, USA), and each genotype was subsequently propagated to 800 plants in tissue culture at the Yunnan Academy of Agricultural Sciences (YAAS) Potato Center (Kunming,China) in May 2019. The plants derived from tissue culture were transplanted to the field at the three experimental stations (e1; e2; and e3-1) after the seedlings had been harden off for two weeks in greenhouse, the day temperature was 22°C and the night temperature was 18°C. After transplanting, the plants derived from tissue culture were irrigated by sprinklers every day, and irrigated 10 days. There were no any other treatments and plant protection applied at seedling stage. The plants were spaced 0.3 m, and the rows were spaced 0.7 m. A randomized complete block (RCB) design was used with three replicates in 2019-2020. Each plot was planted with 80 plants derived from tissue culture of each genotype with three replicates at the three experimental stations in 2019. The other plants derived from tissue culture were transplanted to sterile substrate in a greenhouse in Song Ming, Yunnan, and the mini-tubers that were harvested from greenhouse in the 2019 were used as seed potatoes for the experiment in 2020. Each plot was planted with 60 plants of every genotype with three replicates at the five potato experimental stations in 2020. Late blight (*Phytophthora infestans*) was controlled in all the locations for all genotypes, but late blight was controlled by chemicals in XW, the late blight still broke out. After the plants died completely

TABLE 1 Brief description of the geographic, metrological parameters of five experimental sites in Yunnan Province, southwest China.

Location	East longitude	North latitude	Altitude (m)	Average temperature in growing season	Cumulative precipitation during the growing season (mm)	annual precipitation(mm)	Frost-free period per year (days)	Sowing date	Harvesting date
e1(HZ)	103°19'12.0''	25°53'24''	2668	14.81	616.2	723	262	2019.5.8 2020.4.7	2019.9.24 2020.10.14
e2(LQ)	102°34'4''	26°5'42''	2794	–	–	–	–	2019.5.10 2020.4.1	2019.9.18 2020.10.23
e3-1(ZT)	103°38'42''	27°16'35''	1930	20.12	257	511.2	347	2019.5.10	2019.8.23
e3-2(ZT)	103°23'42''	27°14'24''	2480	14.39	621.6	753.2	292	2020.4.1	2020.9.14
e4 (LJ)	100°5'	26°47'	2780	14.68	636.2	664	211	2020.4.10	2020.11.3
e5 (XW)	102°6'33.24''	26° 13'13.39''	2210	–	–	–	–	2020.4.8	2020.10.25

e1: Huize (HZ), e2: Luquan (LQ), e3-1: Zhaotong (ZT), e3-2: Zhaotong (ZT), e4: Lijiang (LJ), e5: Xuanwei (XW).

(because of late blight), the tubers were harvested by hand in both years.

2.4 Data collection

2.4.1 Data of quality traits

2.4.1.1 Evaluation of French fry quality

Ten of the most uniform and least blemished potato tubers were selected from each of the three replicated plots after the harvest at maturity. The soil and debris were removed from the tubers, which were then washed. They were hand-peeled starting at the stem-end (basal) side of all 10 potato tubers using stainless steel knives, and the peeled tubers were cut along the long axis with an automatic French fry cutter machine into 7 mm × 7 mm rectangular strips; the central four strips of each tuber were collected (Waxman et al., 2018). The 40 central strips were then soaked, blanched, dried, and par-fried (Supplementary Figure 2). Palm oil was used to fry the potato strips. The French fries of all genotypes were handled and processed with the same method (Supplementary Figure 2). The crispness, exterior shell, mealiness, moistness, texture variation and internal appearance of French fries of each genotype were evaluated by three panelists from the Yunnan Academy of Agricultural Sciences. The panelists were asked to score the six parameters using a standard rating scale from 1 to 9 as previously described. For example, for Crispness, 1 means less crisp and 9 means severely undesirable crunchy; for exterior shell: 1 means weak, 9 means tough; for all six parameters, 5 indicates the most optimal quality (Supplementary Table 6) (Waxman et al., 2018; Liang et al., 2020).

2.4.1.2 The dry matter content

More than 5 kg tubers of each plot was sampled to determine the content of dry matter during the harvest. The air and water weight of each material were measured using an electronic portable scale (1 g), and the specific gravity was calculated using Equation (1). The dry matter content was calculated using Equation (2), which is the regression equation between the dry matter content and the specific gravity as found in the Mepkep table (Heltoft et al., 2017).

$$T = (A - 30)/[(A - 30) - B] \ldots\ldots \quad (1)$$

$$D = 213.9 \star T - 211.3 \ldots\ldots \quad (2)$$

Where T is the specific gravity; A is the weight in air, and B is the weight in water. D is the dry matter content, and the weight is provided in g.

2.4.1.3 Tuber length, width and the length/width

The longest and widest positions of 20 potato tubers were measured with a digital Vernier caliper, and the measured data were automatically recorded in a computer. The length/width = the tuber length/tuber width.

2.4.2 Yield and its components

Total yield: At harvest, all the tubers in each plot were collected. The yield, tuber count, and grading were performed manually. The

tubers were harvested by hand and graded on weight (large tubers > 100 g, medium tubers 50–100 g, and small tubers< 50 g). The tubers > 100 g were considered to be commercial-grade tubers. The number of single tubers in each grade was counted and weighed to determine the total yield. The ratio of large tubers was determined as the percentage of tubers that weighed >100 g relative to the total yield of fresh tubers. The weight per marketable tubers was determined as the weight of tubers that weighed > 100 g/the number of the tubers that weighed > 100 g. The ratio of small tuber was determined as the percentage of tubers that weighed< 50 g relative to the total yield of fresh tubers.

2.4.3 Agronomic properties

The plant vigor was determined by examination after emergence (emergence rate > 80%) and again after 30 days (Supplementary Table 2).

The plant maturation was investigated from middle to late August (Supplementary Table 3).

Resistance to late blight was assessed by the area under the disease progress curve (AUDPC). From the first susceptible plant was found, then the incidence of late blight was evaluated in every 7 days for three to eight times. The AUDPC was calculated based on the incidence of late blight. The method of International Potato Center (Forbes et al., 2014) was used as the standard to investigate the rate of incidence and calculate the AUDPC.

2.5 Statistical analysis

2.5.1 Data analysis

The data of the experiments were collected in 2019 and 2020, respectively. And the data used for analysis by AHP and GGE was the average of two years at e1 and e2, the average of e3-1 in 2019 and e3-2 in 2020, and the data at e4 and e5 in 2020. Because *in vitro* plantlets and the microtubers were virus-free, and there was no difference in growth, disease resistance, length/width of tubers (Supplementary Figure 3) in 2 years, the performance of the 33 cultivars and breeding clones at 3 experimental locations in two years was consistent. Therefore, the differences between 2019 and 2020 and the differences between 3 and 5 experimental locations were not analyzed.

2.5.2 The data dimensionless according to the positive and negative effects of a single factor

The raw data, such as the "crispness value" obtained from the panelists, ranged between 1and 9. However, since 5 were the best score of six parameters, and 9 and 1 were the worst scores, it was impossible to directly evaluate the parameters and make strategic decision. In addition, other data such as the value of "small tuber ratio (< 50 g)" and the late blight AUDPC have negative effects on the yield and income of growers. Therefore, the raw data were calculated using an "S-type" curve model, an inverse "anti-S-type" curve model or the parabolic model (ladder type), and the formulas used were Function (1) to Function (7) (Table 2). In formula (1): f (x): membership function; x: measured data; a: low limit; b: high limit; b1 and b2: the best score corresponding to the measured data/ the highest limit. This was also used to describe functions (2) to (7). The scores of the 11 indicators (see 2.5.3, sub-criteria P1-P11) were added as the score of genotypes at the five sites.

Membership function (1):

$$f(x)=\begin{cases}10, x \le a\\ 0.1+0.9*(\frac{x-a}{b1-1}), a<x<b1\\ 100, x=b1, b1=b2\\ 1-0.9*(\frac{x-b2}{b-b2}), b2<x>b\end{cases}$$

Membership function (2):

$$f(x)=\begin{cases}10, x \le 19.5\\ \frac{x}{25}, 19.5<x<25\\ 100, x=25\\ 1-(\frac{x-25}{25}), 25<x<100\end{cases}$$

Membership function (3):

$$f(x)=\begin{cases}0, x \le a\\ \frac{x}{\max(x)}, x>a\end{cases}$$

Membership function (4):

$$f(x)=\begin{cases}0, x \le 0\\ \frac{x}{149.3}, 0<x<149.3\\ 100, 149.3<x<294.1\end{cases}$$

Membership function (5):

$$f(x)=\begin{cases}100, x \le a\\ 1-\frac{x}{12}, 0<x<12\\ 10, x=12\\ 0.1*\frac{\max(x)-x}{\max(x)}, 12<x<100\end{cases}$$

Membership function (6):

$$f(x)=\begin{cases}10, x \le 1\\ 1-\frac{x}{9}, 0<x<9\end{cases}$$

Membership function (7):

$$f(x)=\begin{cases}10, x=\max(x)\\ 1-(\frac{\frac{x}{\max(x)}-0.2}{1-0.2}), x>0\end{cases}$$

2.5.3 Establishment of a comprehensive evaluation system by AHP

According to the interrelationship and subordinate levels of the factors that are directly related to the selection of high-yield and high-quality potato cultivars imported from the US, different levels of aggregation and combination were conducted to build a hierarchical framework model (Table 3).

Goal hierarchy (A) was a comprehensive evaluation of the imported potato processing cultivars from the US in Yunnan, China. The first hierarchy (B) was composed of three primary

TABLE 2 S-type and anti-S-type membership functions, parameters, and weights.

Index The x of the Function(1-7)		Membership function	Membership function parameter (Critical value)				Weight	
			a(Lower limit)	b1	b2	b(Up limit)		
Quality of french fries	Crispness	Function(1)	1	5	5	9	0.149	The weight of quality traits 0.483
	The hardness of exterior shell		1	5	5	9		
	Taste fineness		1	5	5	9		
	Moistness		1	5	5	9		
	Texture variation		1	5	5	9		
	Internal appearance		1	5	5	9		
Dry matter content (%)			19.5	25	25	100	0.123	
Length/width of tubers		Function(2)	1.50	1.75	1.75	2.5	0.123	
Length of tubers (cm)		Function(3)	5				0.088	
The ratio of big tuber			0			100	0.107	The weight of yield traits 0.301
The weight per marketable tubers (g/one tuber)		Function(4)	0	149.3	294.1		0.077	
Small tuber ratio (width<5cm)		Function(5)	0	12	12	100	0.064	
Yield (Kg/hm^2)		Function(3)	0			100	0.053	
Maturity		Function(6)	1			9	0.089	The weight of agronomic traits 0.216
Plant vigor		Function(1)	1	5	5	9	0.074	
Late blight resistance (AUDPC)		Function(7)	0			100	0.053	

Effect of index for tuber economic and quality benefit. Quality of french fries: Parobolic,5 is the highest score,100; Length/width of tubers:Parobolic, the value=1.75 is the highest score, 100, if the value is less than 1.5, the score is 10; Dry matter content (%): The value is less than 19.5%, the score is10, the value is 25%,the score is100, the value is more than 25%, the score decreases; Length of tubers (cm): The value<5cm,the score is 10, the value >5cm, the score is the value/the max value of all cultivars; The ratio of big tuber: The max value of all cultivars,the score is100, the score of others is the value/the max value of all cultivars; The weight per marketable tubers (g/one tuber): The value is in the 0-149.3, the score is the value/149.3, the value is in the 149.3-294.1, the score is 100; Small tuber ratio (width<5cm): The value is 0, the score is 100, the value is 12%, the score is 10, the value is in the 12-100%, the score is 10-0; Yield (Kg/hm^2): The max value of all cultivars,the score is100, the score of others is the value/the max value of all cultivars; Maturity: The value is in the 1-9, the score is 100-10; Plant vigor: Parobolic,5 is the highest score,100; Late blight resistance (AUDPC): The max value of all cultivars, the score is10, the score of others is calculated by Function(7).

criteria (the quality traits [B1], yield traits [B2] and agronomic traits [B3]). The second hierarchy (P) was composed of 11 sub-criteria (Quality of French fries [P1], Dry matter content [P2], Length/width of tubers [P3], Length of tubers [P4], The ratio of large tubers [P5], The weight per marketable tuber [P6], Small tuber ratio [P7], Yield [P8], Maturity [P9], Plant vigor [P10], and Late blight resistance [P11]). The scores of the comprehensive evaluation system were used to select the French fry processing genotypes that were high-quality, high-yielding, vigorous and moderately resistant to late blight.

The importance of each criterion and the ranking of criteria were based on the scoring of experts in potato breeding. Further, the method used an underlying proportion criterion recording scale to rate the relative preference on a one-to-one basis of each criterion (Table 3). In the first hierarchy, the quality traits (B1) were significantly more important than the yield traits (B2), and B2 was extremely important compared with the growth traits (B3). The second hierarchy in the B1, the quality of French fries (P1), was significantly more important than the dry matter content (P2) and further importance ranking was P2=P3>P4. In the second hierarchy

TABLE 3 Research framework for potato genotype suitability modeling in Yunnan Province, China.

Goal hierarchy(A)	Criteria: First hierarchy (B)	Sub−criteria: Second hierarchy (P)
Suitability of potato genotypes in Yunnan, China	The quality traits (B1)	Quality of French fries (P1), Dry matter content (P2), Length/width of tubers (P3), Length of tubers (P4)
	The yield traits(B2)	The ratio of large tuber(P5), The average weight per marketable tuber (P6), Small tuber ratio (width<5cm) (P7), Yield(P8)
	The growth traits(B3)	Maturity (P9), Plant vigor (P10), Late blight resistance (P11)

in the B2, importance of the ratio of large tubers was P5>P6>P7>P8. In the second hierarchy in the B3, importance of maturity was P9> P10 >P11.

In the construction of the judgment matrix, the weights of factors were calculated from the judgment matrix, the maximum eigenvalue of matrix (λ max), and the corresponding eigenvectors were obtained by calculation and checked for consistency. The judgment of consistency could be checked by the CR (Consistency ratio) of CI (Consistency index) with the appropriate value (Supplementary Tables 4, 5), and the four CR were 0.031, 0.008, 0.008 and 0.015, respectively (Table 4), and the fact that the CR values were all< 0.1 is a reasonable level of consistency (Malczewski, 1999; Kamble and Raut, 2019). So, the weight coefficients based on the relative importance, and proportion of various evaluation indicators of a certain tested object were 0.149(P1), 0.123(P2), 0.123(P3), 0.088(P4), 0.107(P5), 0.077(P6), 0.064(P7), 0.053(P8), 0.089(P9), 0.074(P10), 0.053(P11) respectively. Therefore, the weight coefficients of quality traits was 0.483(P1, P2, P3 and P4), the weight coefficients of yield traits was 0.301(P5, P6, P7 and P8), and the weight coefficients of agronomic traits was 0.216(P9, P10 and P11).

2.6 GGEbiplot was used to analyze the interaction between genotypes and the environment

To determine the influence degree of GE (Genotype × Environment interaction) on stability and adaptability of exotic potato genotypes, the 33 US genotypes and breeding clones were planted in five different ecological zones in Yunnan Province, China. The ecological zones or environments were evaluated using the repeatable part in GE to determine which zones were suitable for genotype screenings. Genotype evaluation, test environment evaluation and discriminating ability of the environments were analyzed using a GGEbiplot. The comparison and histogram of all 11 indicators were analyzed using R software (Vienna, Austria). All the raw data were sorted and calculated by Microsoft Excel 2010 (Redmond, WA, USA).

3 Results

3.1 Quality traits, yield traits and growth traits of the studied potato genotypes

3.1.1 Quality traits

Quality traits were used to evaluate the fitness of tubers for French fries processing from 5 experimental stations. The average value of the quality of French fries was 5.0-5.5 at e5 > e4 > e2. The average value of quality of French fries was closest to 5.0 at e2, and the average value of quality of French fries was 4.0-5.0 at e3 > e1. Simultaneously, e3 had the lowest recorded score between 2.5 and 3.3 (Figure 1A). The highest frequency histogram of the average value of quality of French fries was 4.5-5.5 (Figure 1B). The average value of dry matter content was 16-22% (Figure 1C). The potatoes grown at e1 had the largest average dry matter content, and the average dry matter content was ordered as e1 > e2 > e3 > e4 > e5. The lowest average dry matter content was at e5. The largest dry matter content > 30% and was recorded at e1 and e4, while the lowest dry matter content was recorded at e3. The highest frequency histogram of the dry matter content was between 18% and 22% (Figure 1D).The average value of the length/width of the tubers was 1.5-1.8 (Figure 1E), and the largest average length/width of the tubers was at e3 (Figure 1E). The average length/width of the tubers was ordered as e3 > e1 > e2 >e4 > e5. The tubers had the lowest

TABLE 4 Judgement matrix and the results of a consistency check.

A-B					B1-Pi					
A	B1	B2	B3	W	B1	P1	P2	P3	P4	W
B1	1	1.4	2.52	0.484	P1	1	1.4	1.4	1.68	0.307
B2	1/1.4	1	1.8	0.301	P2	1/1.4	1	1	1.2	0.255
B3	1/2.52	1/1.8	1	0.215	P3	1/1.4	1	1	1.2	0.255
λmax=3.032 CI=0.016 RI=0.52					P4	1/1.68	1/1.2	1/1.2	1	0.182
CR=0.031<0.1					λmax=4.024 CI=0.008 RI=0.89 CR=0.009<0.1					
B3-Pi					B2-Pi					
B3	P9	P10	P11	W	B2	P5	P6	P7	P8	W
P9	1	1.4	1.68	0.412	P5	1	1.2	1.44	2.016	0.356
P10	1/1.4	1	1.2	0.343	P6	1/1.2	1	1.2	1.68	0.255
P11	1/1.68	1/1.2	1	0.245	P7	1/1.44	1/1.2	1	1.4	0.212
λmax=3.016 CI=0.008 RI=0.52					P8	1/2.016	1/1.68	1/1.4	1	0.177
CR=0.015<0.1					λmax=4.024 CI=0.008 RI=0.89 CR=0.009<0.1					

The abbreviation of CR, CI and RI mean consistency ratio, consistency index and random index.

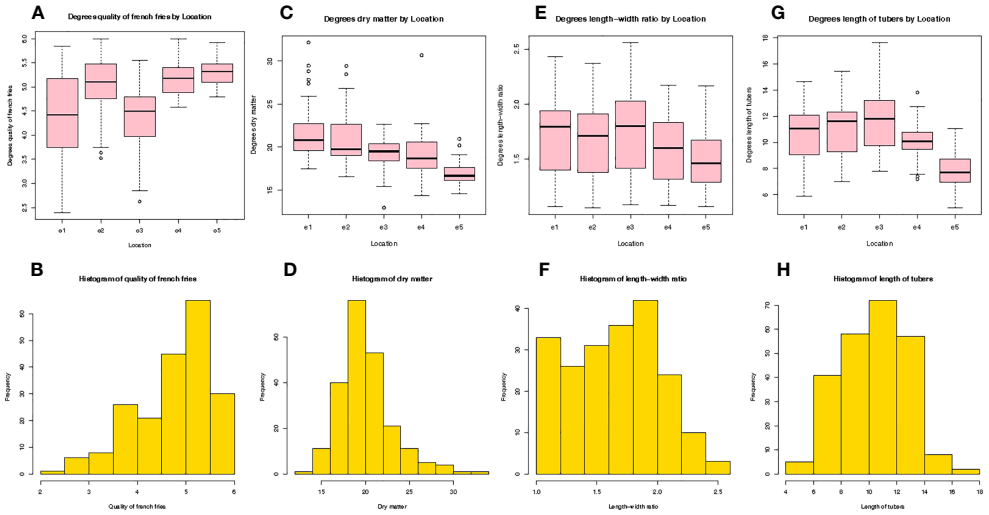


FIGURE 1
A comparison and histogram of the quality of French fries **(A, B)**, dry matter **(C, D)**, length to width ratio **(E, F)**, and length of tubers **(G, H)** averaged across five sites in the spring in Yunnan Province, China. (HZ: e1; LQ: e2; ZT: Average of e3-1and e3-2; LJ: e4; XW: e5).

length/width at e5, which was nearly 1.5, and the largest length/width of tubers > 2.5 at e3. The highest frequency histogram of the length/width of tubers was 1.5-2.0 (Figure 1F). The average value of length of tubers was 7-12 cm (Figure 1G), and the largest average length of tubers was at e1, which was nearly 12 cm, and the average length of tubers was ordered as e3 > e2 > e1 > e4 > e5. The tubers at e5 were the shortest at nearly 8 cm. The highest frequency histogram of the length of tubers was 8-14 cm (Figure 1H). In summary, the performance of the tuber quality traits were not consistent at the five experimental stations, and the average value of quality of French fries was highest in e5, the highest average dry matter content was in e1,and the highest average value of the length/width of the tubers and length of tubers were in e3.

3.1.2 Yield traits

The yield components such as the ratio of large tubers and the small tuber ratio directly affect the yield of potatoes and their economic income per unit area. Our results showed that the highest average value of the ratio of large tubers was found at e3 (Figure 2A), and the average ratio of large tubers was ordered as e3 > e2 > e1 > e4 > e5. The highest frequency histogram of the ratio of large tubers was between 50% and 88% except for e5 (Figure 2B). The highest average value of the weight per marketable tubers was at e3 (Figure 2C), and the average weight per marketable tubers was ordered as e3 >e4 > e2 > e1 > e5. The highest frequency histogram of the weight per marketable tubers was between 120 g and 200 g (Figure 2D). The highest average value of the small tuber ratio (44.8%) was observed at e5 (Figure 2E), and the average value of small tuber ratio in the other four sites was approximately 10%. The average small tuber ratio was ordered as e5 (44.8%) > e3 (15.1%) > e1 (11.0%) > e4 (9.3%) > e2 (8.8%). The highest frequency histogram of small tuber ratio was between 0 and 20% (Figure 2F). The highest average value of yield (kg/hm^2) was at e3 (Figure 2G), and the average yield was ordered as e3 > e2 > e4 > e1 > e5. The highest frequency histogram of the weight per marketable tuber was 15,000-30,000 (kg/hm^2) (Figure 2H). A yield higher than 90,000 (kg/hm^2) was recorded at e1, and the yield recorded at e2 was more than 60,000 (kg/hm^2), and there was a lowest yield record at e3. In summary, the ratio of large tubers and the weight per marketable tubers were higher, the yield was higher, but the lower small tuber ratio the higher yield such as in e3.

3.1.3 Agronomic traits

Agronomic traits (maturity, plant vigor and AUDPC) were always used to analyze and evaluate the disease resistance and genetic diversity of imported cultivars. Our results showed that: the average value of maturity was close to 7 at e1 (Figure 3A), while it was close to 6 at e3 and 5 at e4. The highest frequency histogram of maturity was 4-6 (Figure 3B). However, the value of maturity data of e2 and e5 were missing. The average value of plant vigor was close to 7 at e3 (Figure 3C), 6 at e4, close to 5 at e2 and close to 4.5 at e1. The highest frequency histogram of plant vigor was 4-5 (Figure 3D). However, the value of plant vigor data of e5 was missing. The highest average value of the AUDPC was at e4 (Figure 3E), and the average AUDPC was ordered as e4 > e2 > e1 > e3. The highest frequency histogram of AUDPC was 4,000-5,000 (Figure 3F). However, the value of AUDPC data of e5 was missing. In summary, the good performance of agronomic traits indicates that these cultivars were with strong adaptability at e3.

3.2 Evaluation of each genotype in five sites using AHP

We calculated the selected 11 indicators based on the constructed judgment matrix in AHP, and obtained the scores of the cultivars at each location. By constructing a pairwise comparison judgment matrix to determine the important of

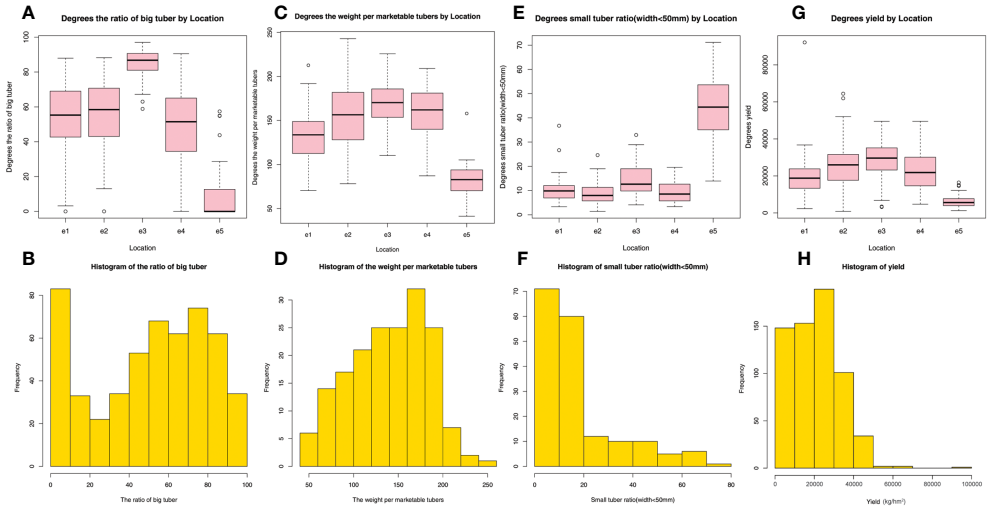


FIGURE 2
A comparison and histogram of the ratio of large tubers **(A, B)**, the weight per marketable tuber **(C, D)**, ratio of small tubers (width< 5 cm) **(E, F)** and yield **(G, H)** at five sites in the spring in Yunnan Province, China.

different factors on screening excellent potato cultivars, errors caused by accidental factors were avoided. The cultivars with the highest score in five sites were as follows: g1 ('Premier Russet') in e2 and e3, g31 ('AO03123-2') in e1, g34 ('Yunshu401') in e4 and e5, respectively (Supplementary Tables 7–11). The second ranked cultivars in the five sites were as follows: g34 in e1, g3 ('Defender') in e4, g7 ('Alpine Russet') in e2, g2 ('Palisade Russet') in e3 and g10 ('Castle Russet') in e5. The third ranked genotypes in the five sites were as follows: g3 in e1, g2 in e4, g32 ('AO96305-3') in e2, g21 ('A03921-2') in e3 and g3 in e5. In summary, g3 was in the top six genotypes in all five locations indicating that it appeared the most suitable for Yunnan, China, of the US entries.

3.3 GGEbiplot evaluation results from 33 US potato genotypes and control cultivars Yunshu 401 and Cooperation 88

3.3.1 Selection of the most desirable genotype

GGE Biplot is used to select cultivars with good comprehensive characters, strong adaptability, stability and high yield. The

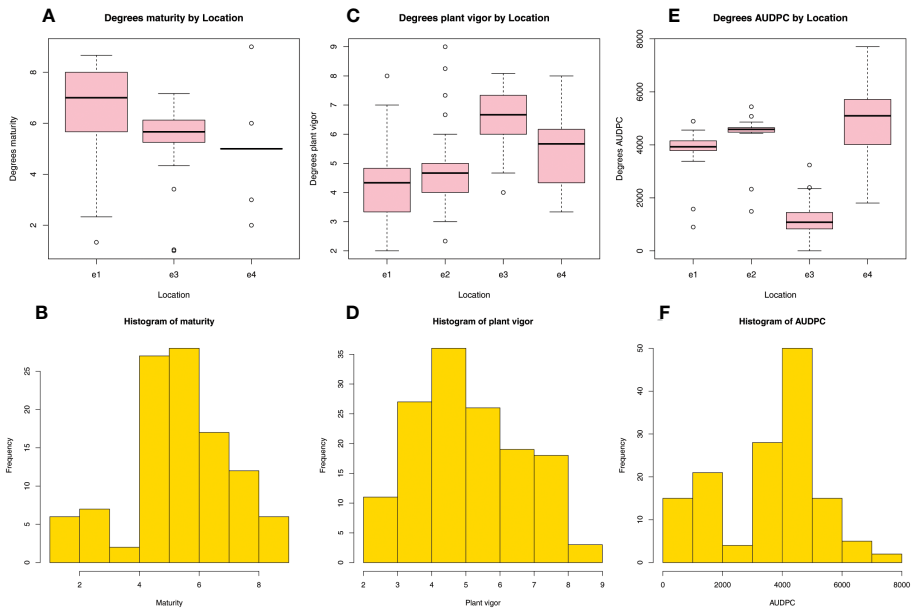


FIGURE 3
A comparison and histogram of the maturity **(A, B)**, plant vigor **(C, D)** and AUDPC **(E, F)** of five sites in the spring in Yunnan Province, China. AUDPC, area under the disease progress curve.

variation of 78.14% of the G and the interaction between GE can be effectively explained using a GGEbiplot through analyzing our 5 experimental locations data (Figure 4). The four test locations of e1, e2, e4 and e5 were considered as one group, and e3 was another group. The genotype g34 ('Yunshu401') which located at the top corner of the polygon was determined to be the most suitable genotype with the highest yield at these four test sites, followed by g3 ('Defender'). Since g34 was the control genotype, g3 was technically the superior genotype with the highest yield at these four sites. In addition, g4 ('Yukon Gem') located at the top corner of the polygon in another group was also the superior genotype with the highest yield at e3 (Figure 4). the genotypes g2, g23, g25 and g14 were located at the other top corner of the polygon but the traits were not desirable. Over all, g3 and g4 were the most adaptable genotypes in Yunnan.

3.3.2 The relationships among environment and the ability to discriminate their degree of representation

The cosine of angle between the two environmental vectors approximates the genetic correlation coefficient between them. If the angle less than 90° indicates a positive correlation, and the angle greater than 90° indicates a negative correlation, and the angle close to 90° indicates no correlation. Positive correlation indicates that the two environments have similar ranking of cultivars; negative correlation indicates that the two environments have a different ranking of cultivars, and the included angles of these five test sites were all< 90° (Figure 5). This showed that there was a close positive correlation between these five test sites. The order of genotypes was similar, and the close positive correlation indicates that some test sites may be repeatedly established. The length of the environment vector is the degree to which the experimental location is able to distinguish varieties. The environmental vectors of e2, e4 and e3 were longer than those of the e1 and e5 test sites, and the e2, e4 and e3 test sites could clearly discriminate between the genotypes

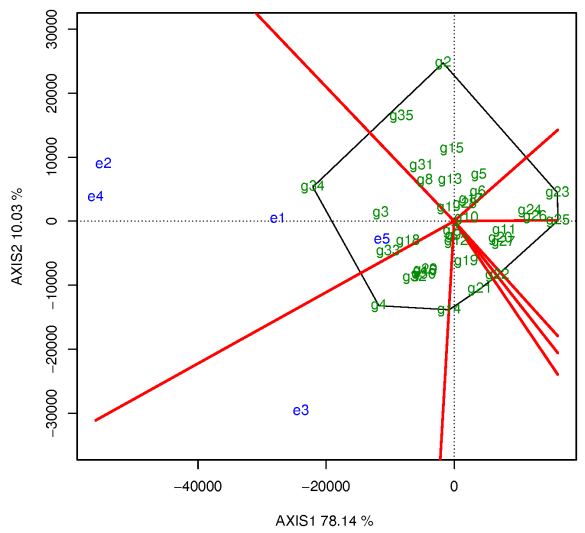


FIGURE 4
Superior genotypes based on the evaluation.

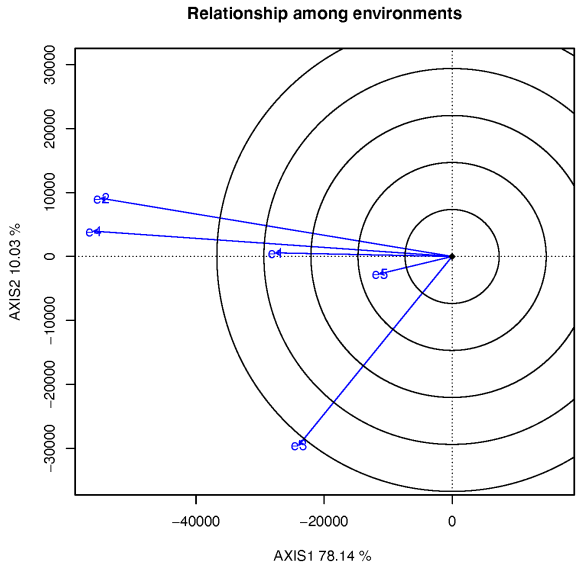


FIGURE 5
The relationship among environment.

(Figure 5). Therefore, the e1 and e5 test sites can be removed to reduce the expense of these tests without affecting the evaluation.

The angle between the test site and the average environmental axis vector is a measurement of the ability of target environment to be representative. A smaller angle indicates stronger representation. In contrast, a large angle indicates that the representation is weak. The test sites e1 and e5 were ineffective because they could not discriminate genotypes. So, e3 (the angle is bigger than the other four locations) could discriminate but was not representative and could only be used to eliminate unstable genotypes. However, it could not be used to select excellent genotypes (Figure 6). Overall, e2 and e4 were most suitable to select genotypes owing to their comprehensive determination of high and stable yields (Figure 6).

3.3.3 The high and stable yield genotypes

The line with a single arrow is the average environmental axis. The direction marked by the straight line indicates an increase in production in that direction. The average yield of g34 was the highest, followed by g4, g3, g33 ('Echo Russet Russet'), and g18 ('A02267-1Y') and so on. (Figure 6). g34 and g4 were the high-yielding cultivars, but their output varied across environments and was poor in some of them. The average yield of cultivars g1, g7, g9, g10, g12, g13 was close to the total experimental average. A straight line with double arrows perpendicular to the average environmental axis and passing through the origin represents the tendency of each variety to interact with each environment. The length of the short dashed line on the vertical environmental axis represents stable of yield, and the shorter the line, the more stable yield, the longer the line, the more unstable yield. The genotypes g23 ('A05182-7Y') had the lowest yields, followed by g25 ('A06336-2Y'), g26 ('A07008-4TE'), and g24 ('A06084-1TE') (Figure 6). Since these genotypes have stable low yields, they could be eliminated from the group of genotypes under study. The genotypes g3 and g33 were identified as the cultivars with the highest and most stable yields.

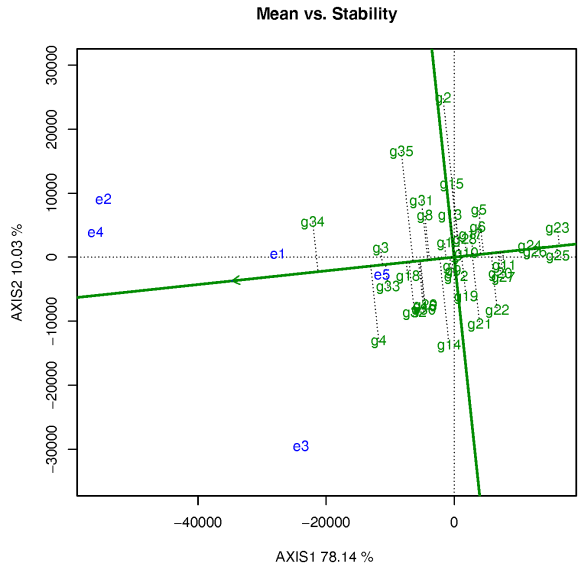


FIGURE 6
The high and stable yield genotypes.

4 Discussion

4.1 Development of an evaluation scheme

Currently, there are very few evaluations of comprehensive agronomic traits and qualities of the potatoes used to manufacture French fries. In the past, the genotypes were measured by the experience of researchers and potato breeders for many years and the specific performance of cultivars in the field. There were many non-quantitative factors and subjective factors during the evaluation process. Other studies used a weight coefficients (relative importance, and proportion of various evaluation indicators of a certain tested object) of yield traits as 0.73(Wang, 2017). However, it was 0.301 in this study. This difference is owing to the improvement in yield traits through cultivation techniques in addition to the potential of the variety itself. However, the quality traits of French fries are generally determined by the characteristics of variety. If the French fries do not have a high quality, a high yield is meaningless. Therefore, the weight of yield traits was reduced, and the weight of quality traits was increased to 0.483. AHP is a most widely used technique to assess factors, such as potatoes that are the primary raw materials in a potato chips processing company (Kamble and Raut, 2019), the suitability of land for potato (Iliquín Trigoso et al., 2020), and the soil fertility index suitable for potato production (Bagherzadeh et al., 2018). It consists of estimating the importance of each criterion according to experts' opinions (Iliquín Trigoso et al., 2020). The AHP technique, which is used to resolve hierarchically structured decision-making problems at different levels, seeks to estimate the relative weight of criteria (Saaty, 1988; Srdjevic, 2005) that could accurately and objectively reflect the quality of materials. Stability and high yield are important indicators to evaluate the genotypes, which are generally measured by the interactions between genotype and environment (Li et al., 2016). These factors play a key role in the selection and evaluation of genotypes that are widely adaptable. Thus, not only the field agronomic traits should be evaluated, but the tuber quality traits should also be evaluated (Liu et al., 2017). It is necessary to use agronomic traits combined with potato processing quality indicators for comprehensive evaluations to identify genotypes that are suitable for processing into French fries in Yunnan. The GGEbiplot has been widely used to evaluate other crops, such as 10 genotypes of winter wheat (Triticum aestivum) in 10 test sites, and comprehensively evaluated the primary genotypic effect G and its interaction effect with the environment GE, which could obtain a reasonable evaluation of the representativeness and discrimination of the test site (Yan et al., 2000).Thus, this study used the two methods AHP and GGEbiplot to evaluate the potato genotypes imported from the US and grown at five sites in Yunnan, China. In addition, the AHP was used to comprehensively evaluate the potato genotypes for their quality, yield, and growth traits, and the GGEbiplot was used to genotype the main effect G and its GE.

4.2 The evaluation indicators and the restrictive criteria of the indicators screen

4.2.1 Quality traits

The source of raw potato of different location tubers and genotypes has significant effect on the quality of French fries, this is consistent with the results of Agblor and Scanlonin, 2002. The aim of this study was to screen evaluation indicators and to construct a comprehensive evaluation system of potato genotypes, including the potato cultivars for their suitability for processing into French fries. To achieve commercial success, genotypes for French fries have to fulfill stringent requirements concerning tuber quality traits, such as the color of French fries etc (Bahati-Kajunju, 2021), the length of tuber, the length/width of tuber, appropriate shape, dry matter content, texture changes after frying, and the hardness of French fries after frying (Agblor and Scanlon, 2002; Cunningham et al., 2008). In the past few years, potato genotypes used for French fries were evaluated, but the raw potato tubers all originated from the same site (Abong et al., 2009; Sharara, 2015; Kaur et al., 2021). In these studies, multiple indicators were used, but the indicator was only a simple description, such as a desirable oblong shape, higher dry matter (dry matter of one potato genotype was 21.61%) and a lower content of reducing sugars (reducing sugars content of one potato genotype was 0.06%). However, this method is not suitable to evaluate multiple potato genotypes at numerous sites. The selection of indicators and the critical values are very important for the evaluation system. In one study, the color, texture, taste and smell, general appearance, moisture, fat content, and dark ends as features were used to assess the quality of the French fries processed from 14 potato cultivars (Sawicka et al., 2021). In another study, the crispness, exterior shell, mealiness, moistness, textural variation, and internal appearance were used to assess the quality of the French fries of three processing potato cultivars, including 'Russet Burbank', 'Alpine Russet', and 'Clearwater Russet' (Waxman et al., 2018). And many studies vary in the selection of the critical values

of evaluation indicators. For example, tubers for the French fry market are longer, and the best tuber length/width (L/W) ratio of the tubers is 1.75. This is the gold standard for the evaluation of French fries that are processed for McDonald's in the US, and the L/W ratio should exceed 1.8 (Neilson et al., 2021). The L/W ratio standard for the French fries was 1.5 in a study by Liu et al. (2017). In this study, the critical values of L/W ratio were 1.5, 1.75 and 2.5. If the L/W ratio is small than 1.5, the score was 10. If the L/W ratio< 1.75, the score was 100, and if the L/W ratio > 2.5, the score was 10, A higher overall score was given to genotypes with the desired L/W ratio, based on the L/W ratio is in 1.5~1.0, the tuber is near round and too much flesh is wasted after cutting French fries, and if the L/W ratio > 2.5, French fries is easy to break in the middle. Kaur et al. (2021) noted that the length of tubers > 5 cm was acceptable, and the tuber length that was the most suitable for French fries was 7.9-10.2 cm. In the UK, incremental bonus points are paid on long cultivars when a percentage of tubers > 7.5 cm long (Kirkman, 2007). The width of tubers > 5 cm was acceptable, but penalties will be incurred if the width of tubers is< 5 cm. The dry matter content that is the most suitable for French fries is 19.1-21.6%, and the best dry matter content is ≥ 20% (Kaur et al., 2021). A dry matter content ≤ 19.5% is not acceptable for French fries. Upper limits do not apply, although penalties may be incurred for >25% dry matter during the manufacturing of French fries (Kirkman, 2007). However, others consider that a dry matter content ≥ 20% is acceptable (Abong et al., 2009). A dry matter content > 25% would indicate that a variety could be used to process French fries (Liu et al., 2017). The best dry matter content is 20.7%-24.0%, which is the gold standard for the evaluation of processing McDonald's fries in the US. In this study, we found if the value of dry matter content< 19.5%, the French fries color of this cultivars was dark, and if the value of dry matter content >25%, the French fries color of this cultivars was slight dark, and the internal appearance of this cultivars was additional hollowing. So when the value was< 19.5%; the score was 10. A value of 25% resulted in a score of 100, and a value > 25% resulted in a decreased score. In conclusion, the L/W ratio is near 1.75, dry matter content near 25% and the tuber length is more than 5cm is more suitable for French fries.

4.2.2 Yield traits

Yield traits offer important information about potato cultivars, such as the higher the ratio of large tubers, the higher the yield. Yield components are also important factors for growers to obtain monetary returns. In this study, the highest value of the ratio of large tubers and yield in each location resulted in a score of 100, and the yield and the ratio of large tuber data of other genotypes were standardized. In addition, the uniformity of tubers is very important for processing French fries to ensure the uniformity of the manufacturing process. One UK manufacturer established the indicators, such as tuber count per 10 kg, and the acceptable tuber count range is 34-67 per 10 kg – an indicator of the distribution of tuber sizes (Kirkman, 2007), which corresponds to the weight per marketable tuber of 294-149 g per tuber. In this study, the weight per marketable tuber was used as one evaluating indicator. Since the ratio of large tubers determines the market price, the tuber consistency and small tuber ratio are the restrictive standards (minimum or maximum standards) for the French fry companies to purchase raw potato tubers. If the proportion of tubers< 5 cm wide exceeds 12%, the lots could be rejected (Kirkman, 2007). In this study, a small tuber ratio was used as a limiting standard. If the small tuber ratio exceeded 12%, the score was 10 out of a total of 100. In summary, the ratio of large tubers, tuber consistency and a small tuber ratio was more important or had a higher priority than the yield factor in our study, but in the past, the small tuber ratio was not selected as one indicator. So the selected of yield traits indicators and their limiting standard were more reasonable.

4.2.3 Agronomic traits

The important agronomic traits include a wide adaptability for characteristics, such as maturity, plant vigor and late blight resistance (AUDPC) (Jansky et al., 2009; Fajardo et al., 2013; Jansky et al., 2013), that contribute to the overall performance of cultivars. In addition, many cultivars differ in plant maturity, vigor and resistance to late blight (AUDPC) under different environments (Johansen et al., 1967). Collins et al., 1999 reported that the plant maturity and vigor strongly correlate with late blight resistance. In our study, late blight is the first major disease to affect the spring potatoes in Yunnan Province, China. Plants that have stronger resistance to late blight, mature later and are more vigorous. However, the potato genotypes imported from the US were less resistance to late blight comparing with the two control cultivars. In addition, the potato genotypes imported from the US are sensitive to day length and not as vigorous as the cultivars adapted to day length of Yunnan Province, China. Thus, the best plant vigor value was 9, their score was 100, and the best plant maturity value was 1. The resistance to late blight of potato was calculated by AUDPC, and a lower AUDPC indicates a stronger resistance to late blight (Forbes et al., 2014). Furthermore, the late blight of all the potato genotypes in this study was chemically controlled, and the incidence of late blight was investigated after chemical control. However, since late blight was not effectively controlled by chemicals in XW, the disease developed rapidly, and the seedlings rapidly died, resulting in low yields.

In our study, the results of AHP and GGEbiplot are highly consistent. The potato genotypes with high comprehensive scores in the AHP are highly adaptable, stable, and high yielding in the GGEbiplot, which is consistent with the performance of the potato genotypes in the field. In addition, the quality of the French Fries is also excellent, such as the g3 ranked top, g23 and g26 ranked bottom by using AHP in 5 sites, and g3 also was showed higher and stable yields, g23 and g26 were also showed lower and unstable yield by using the GGEbiplot. This showed that the results of a comprehensive analysis of genotypes using AHP and GGEbiplot analyses are reasonable and effective, and the established potato genotypes evaluation system could be used as the basis for future evaluation and the identification of cultivars.

5 Conclusions

Eleven indicators for quality traits, yield traits and agronomic traits were selected as the specific evaluation indicators in this study. The weights coefficients (relative importance, and proportion of various evaluation indicators of a certain tested object) of quality traits, yield traits and agronomic traits were 0.483, 0.301 and 0.216, respectively, and discriminated significantly from the schemes applied so far. The 35 potato genotypes were evaluated by using AHP and GGEbiplot analysis in Yunnan, China. The cultivars with high and stable yield, good quality of French fries, suitable tuber length/width ratio, and longer tubers were selected for more suitable growing and high quality of French fries in Yunnan. However, as no results exceed the standard cultivar g3, further cultivars will be evaluated to test their suitable for planting and French fries in Yunnan. Furthermore, reasonable suggestions for the exclusion of test sites without affecting the evaluation of genotypes were provided and help to save costs.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

SL: Investigation, Data collation and analysis, Writing – Original Draft. WJ: Investigation the field data. YY: Improvement of French fries frying method and evaluation method of French fries quality. LL: Tissue cultured plant expanded. JW and RN: Provided the *In vitro* pathogen-free plantlets and revised the manuscript. LB: The Pre-Elite potato seed produced and the experiment in e5 was completed. ZY: The experiment in e2 was completed. ZL: The experiment in e3 was completed. PH: The experiment in e4 was completed. YX: The experiment in e5 was completed. XL: Project administration. All authors contributed to the article and approved the submitted version.

Funding

The authors declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by the National Key R&D Program of China, grant number 2017YFE0104700.

Acknowledgments

The authors thank the reviewers and all of the editors for their helpful comments and suggestions on this paper.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1159848/full#supplementary-material

SUPPLEMENTARY FIGURE 1
The disribution of experimental sites in Yunnan Province, China. e1: Huize (HZ), e2: Luquan (LQ), e3-1: Zhaotong (ZT), e3-2: Zhaotong (ZT), e4: Lijiang (LJ), e5: Xuanwei (XW);The same as bellow.

SUPPLEMENTARY FIGURE 2
Process flow chart of the French fries.

SUPPLEMENTARY FIGURE 3
The tubers length/width of the different genotypes at e1, e2, e3-1, e3-2 in 2019 and 2020.

References

Abong, G. O., Okoth, M. W., Karuri, E. G., Kabira, J. N., and Mathooko, F. M. (2009). Evaluation of selected Kenyan potato cultivars for processing into french fries. *J. Anim. Plant Sci.* 2 (3), 141–147. Available at: http://ir.mksu.ac.ke/handle/123456780/4267.

Agblor, A., and Scanlon, M. G. (2002). Effect of storage period, cultivar and two growing locations on the processing quality of french fried potatoes. *Am. J. Potato Res.* 79 (3), 167–172. doi: 10.1007/BF02871932

Bagherzadeh, A., Gholizadeh, A., and Keshavarzi, A. (2018). Assessment of soil fertility index for potato production using integrated Fuzzy and AHP approaches, Northeast of Iran. *Eurasian J. Soil Sci* 7 (3), 203–212. doi: 10.18393/ejss.399775

Bahati-Kajunju, N. H. (2021). "Processing traits of potato varieties grown in Uganda for potential production of French fries," in *Masters dissertation* (Makerere University).

Buciumeanu, E. C., Vizitiu, D. E., DincĂ, L., and Radomir, A. M. (2021). Use of analytic hierarchy process on the choice of grapevine varieties suitable for a sustainable viticulture and market demands in the viticultural center stefanesti. *Horticulture* 65 (1), 299–304. Available at: https://www.horticulturejournal.usamv.ro/pdf/2021/issue_1/Art42.pdf.

Collins, A., Milbourne, D., Ramsay., L., Meyer, R., Chatot-Balandras, C., Oberhagemann, P., et al. (1999). QTL for field resistance to late blight in potato are

strongly correlated with maturity and vigour. *Mol. Breeding.* 5 (5), 387–398. doi: 10.1023/A:1009601427062

Cunningham, S. E., Mcminn, W. A. M., Magee, T. R. A., and Richardson, P. S. (2008). Effect of processing conditions on the water absorption and texture kinetics of potato. *J. Food Engineering.* 84 (2), 214–223. doi: 10.1016/j.jfoodeng.2007.05.007

Fajardo, D., Haynes, K. G., and Jansky, S. (2013). Starch characteristics of modern and heirloom potat cultivars. *Am. J. Potato Res.* 90 (5), 460–469. doi: 10.1007/s12230-013-9320-5

FAO. (2023). Available at: https://www.fao.org/faostat/en/#home.

Forbes, G., Pérez, W., and Andrade Piedra, J. (2014). *Field assessment of resistance in potato to Phytophthora infestans* (Lima (Peru: International Potato Center (CIP), 35.

Heltoft, P., Wold, A. B., and Molteberg, E. L. (2017). Maturity indicators for prediction of potato (*Solanum tuberosum L.*) quality during storage. *Postharvest Biol. Technol* 129, 97–106. doi: 10.1016/j.postharvbio.2017.03.011

Iliquín Trigoso, D., Salas López, R., Rojas Briceño, N. B., Silva López, J. O., Gómez Fernández, D., Oliva, M., et al. (2020). Land suitability analysis for potato crop in the Jucusbamba and Tincas Microwatersheds (Amazonas, NW Peru): AHP and RS–GIS approach. *Agronomy* 10 (12), 1898. doi: 10.3390/agronomy10121898

Jansky, S. H., Dempewolf, H., Camadro, E. L., Simon, R., Zimnoch-Guzowska, E., Bisognin, D. A., et al. (2013). A case for crop wild relative preservation and use in potato. *Crop Sci.* 53 (3), 746–754. doi: 10.2135/cropsci2012.11.0627

Jansky, S. H., Jin, L. P., Xie, K. Y., Xie, C. H., and Spooner, D. M. (2009). Potato production and breeding in China. *Potato Res.* 52 (1), 57–65. doi: 10.1007/s11540-008-9121-2

Johansen, R. H., Miller, J. C., Newsom, D. W., and Fontenot, J. F. (1967). The influence of environment on the specific gravity, plant maturity and vigor of potato progenies. *Am. Potato J.* 44, 107–122. doi: 10.1007/BF02862909

Kamble, S. S., and Raut, R. D. (2019). Evaluating the factors considered for procurement of raw material in food supply chain using Delphi-AHP methodology-a case study of potato chips processing company in India. *Int. J. Productivity Qual. Manag* 26 (2), 176–189. doi: 10.1504/IJPQM.2019.097765

Kaur, S., Aggarwal, P., and Sharma, S. (2021). Evaluation of potato genotypes for french fry quality. *Agric. Res. J.* 58, 275–279. doi: 10.5958/2395-146X.2021.00040.5

Kirkman, M. A. (2007). *Global markets for processed potato products//Potato Biology and Biotechnology* (Elsevier Science BV), 28–51.

Li, Y. J., Li, D. M., Fan, S. J., He, J. Q., Liu, Y., Zhang, J. L., et al. (2016). Analysis of variety adaptability and yield stability of potato for evaluating GGE-biplot application. *J. Lanzhou University: Natural Sci.* 52 (5), 617–622. doi: 10.13885/j.issn.0455-2059.2016.05.008

Liang, S. M., Yang, Y., Shi, Y. J., Jiang, W., Lu, L. L., Yin, Z. Y., et al. (2020). Evaluation on processing potato varieties introduced from USA in Yunnan Province. *Food Nutr. China.* 26 (5), 18–22. doi: 10.19870/j.cnki.11-3716/ts.2020.05.005

Liu, J., Liang, Y. C., Yu, B., Li, C., Wang, Y. P., Cheng, L. X., et al. (2017). Screening for French Fries processing potato lines according to colour qualities and texture properties. *Scientia Agricultura Sinica.* 50 (22), 4247–4265. Available at: https://kns.cnki.net/kcms/detail/11.1328.S.20171127.1100.004.html.

Malczewski, J. (1999). *GIS and multicriteria decision analysis* (New York: John Wiley & Sons Inc), 392.

Morandi, D. T., de Jesus França, L. C., Menezes, E. S., Machado, E. L. M., da Silva, M. D., and Mucida, D. P. (2020). Delimitation of ecological corridors between conservation units in the Brazilian Cerrado using a GIS and AHP approach. *Ecol. Indic.* 115, 106440. doi: 10.1016/j.ecolind.2020.106440

Neilson, J. A. D., Smith, A. M., Mesina, L., Vivian, R., Smienk, S., and De Koyer, D. (2021). Potato Tuber Shape Phenotyping Using RGB Imaging. *Agronomy* 11 (9), 1781. doi: 10.3390/agronomy11091781

Rahman, R., and Saha, S. K. (2008). Remote sensing, spatial multi criteria evaluation (SMCE) and analytical hierarchy process (AHP) in optimal cropping pattern planning for a flood prone area. *J. Spatial Sci* 53 (2), 161–177. doi: 10.1080/14498596.2008.9635156

Ramírez-García, J., Carrillo, J. M., Ruiz, M., Alonso-Ayuso, M., and Quemada, M. (2015). Multicriteria decision analysis applied to cover crop species and cultivars selection. *Field Crops Res.* 175, 106–115. doi: 10.1016/j.fcr.2015.02.008

Ruswandi, D., Syafii, M., Wicaksana, N., Maulana, H., Ariyanti, M., Indriani, N. P., et al. (2022). Evaluation of high yielding maize hybrids based on combined stability analysis, Sustainability Index, and GGE Biplot. *BioMed. Res. Int.* 1-12. doi: 10.1155/2022/3963850

Saaty, T. L. (1988). *What is the Analytic Hierarchy Process? In Mathematical Models for Decision Support* (Berlin/Heidelberg, Germany: Springer), 109–121.

Sawicka, B., Pszczółkowski, P., Kiełtyka-Dadasiewicz, A., Barbaś, P., Ćwintal, M., and Krochmal -Marczak, B. (2021). The effect of effective microorganisms on the quality of potato chips and French Fries. *Appl. Sci.* 11 (4), 1415. doi: 10.3390/app11041415

Sharara, M. (2015). Evaluation of five potato varieties for producing French fries. *Alexandria J. Food Sci. Technol* 12 (2), 1–10. doi: 10.12816/0025394

Singha, C., Swain, K. C., and Swain, S. K. (2020). Best crop rotation selection with GIS-AHP technique using soil nutrient variability. *Agriculture* 10, 213. doi: 10.3390/agriculture10060213

Srdjevic, B. (2005). Combining different prioritization methods in the analytic hierarchy process synthesis. *Comput. Operations Res.* 32 (7), 1897–1919. doi: 10.1016/j.cor.2003.12.005

Wang, X. B. (2017). *Comprehensive evaluation of introduction potato germplasm resources* (Lanzhou: Gansu Agricultural University). dissertation/master's thesis.

Wang, Z., Yang, P., Peng, H., Li, C., Yue, C., Li, W., et al. (2021). Comprehensive evaluation of 47 tea [Camellia sinensis (L.) O. Kuntze] germplasm based on entropy weight method and grey relational degree. *Genet. Resour. Crop Evol* 68 (8), 3257–3270. doi: 10.1007/s10722-021-01184-7

Waxman, A., Stark, J., Thornton, M. K., Olsen, N., Guenthner, J., and Novy, R. G. (2018). The Effect of harvest timing on French Fry textural quality of three processing potato varieties: Russet Burbank, Alpine Russet, and Clearwater Russet. *Am. J. Potato Res.* 96 (1), 33–47. doi: 10.1007/s12230-018-9686-5

Xu, J. F., and Jin, L. P. (2017). Advances and perspectives in research of potato genetics and breeding. *Scientia Agricultura Sinica.* 50 (6), 990–1015. Available at: https://kns.cnki.net/kcms/detail/11.1328.S.20170320.1252.006.html

Yan, W., Hunt, L. A., Sheng, Q., and Szlavnics, Z. (2000). Cultivar evaluation and mega-environment investigation based on the GGE biplot. *Crop Sci* 40 (3), 597–605. doi: 10.2135/cropsci2000.403597x

Yao, L. L., Zhou, H. F., Yan, Y. J., and Yuan, S. (2022). Projection of suitability for the typical agro-ecological types in Central Asia under four SSP-RCP scenarios. *Eur. J. Agron* 140, 126599. doi: 10.1016/j.eja.2022.126599

Zhao, Y., Huo, B., Lin, S., Zhang, S., Mao, C., Jiang, J., et al. (2022). Germplasm innovation and establishment of comprehensive evaluation system for Hedgerow Garden Chrysanthemum. *Agronomy* 12 (8), 1736. doi: 10.3390/agronomy12081736

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 01 July 2024
DOI 10.3389/fpls.2024.1341788

Check for updates

OPEN ACCESS

EDITED BY
Mathieu Rouard,
Alliance Bioversity International and CIAT, France

REVIEWED BY
Christopher Richards,
Agricultural Research Service (USDA), United States
Hana Chair,
Centre de Coopération Internationale en Recherche Agronomique pour le Développement (Montpellier), France

*CORRESPONDENCE
Noelle L. Anglin
noelle.anglin@usda.gov

RECEIVED 20 November 2023
ACCEPTED 29 April 2024
PUBLISHED 01 July 2024

CITATION
Anglin NL, Chavez O, Soto - Torres J, Gomez R, Panta A, Vollmer R, Durand M, Meza C, Azevedo V, Manrique - Carpintero NC, Kauth P, Coombs JJ, Douches DS and Ellis D (2024) Promiscuous potato: elucidating genetic identity and the complex genetic relationships of a cultivated potato germplasm collection.
Front. Plant Sci. 15:1341788.
doi: 10.3389/fpls.2024.1341788

COPYRIGHT
© 2024 Anglin, Chavez, Soto - Torres, Gomez, Panta, Vollmer, Durand, Meza, Azevedo, Manrique - Carpintero, Kauth, Coombs, Douches and Ellis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Promiscuous potato: elucidating genetic identity and the complex genetic relationships of a cultivated potato germplasm collection

Noelle L. Anglin[1,2]*, Oswaldo Chavez[1], Julian Soto - Torres[1], Rene Gomez[1], Ana Panta[1], Rainer Vollmer[1], Marisol Durand[1], Charo Meza[1], Vania Azevedo[1], Norma C. Manrique - Carpintero[1], Philip Kauth[3,4], Joesph J. Coombs[5], David S. Douches[5] and David Ellis[1]

[1]International Potato Center (CIP), Lima, Peru, [2]Seed Savers - Preservation Department, United States Department of Agriculture Agriculture Research Service (USDA ARS) Small Grains and Potato Germplasm Research, Aberdeen, ID, United States, [3]Seed Savers Exchange, Decorah, IA, United States, [4]REAP Food Group, Madison, WI, United States, [5]Department of Plant Soil and Microbial Sciences, Michigan State University (MSU), East Lansing, MI, United States

A total of 3,860 accessions from the global *in trust* clonal potato germplasm collection w3ere genotyped with the Illumina Infinium SolCAP V2 12K potato SNP array to evaluate genetic diversity and population structure within the potato germplasm collection. Diploid, triploid, tetraploid, and pentaploid accessions were included representing the cultivated potato taxa. Heterozygosity ranged from 9.7% to 66.6% increasing with ploidy level with an average heterozygosity of 33.5%. Identity, relatedness, and ancestry were evaluated using hierarchal clustering and model-based Bayesian admixture analyses. Errors in genetic identity were revealed in a side-by-side comparison of *in vitro* clonal material with the original mother plants revealing mistakes putatively occurring during decades of processing and handling. A phylogeny was constructed to evaluate inter- and intraspecific relationships which together with a STRUCTURE analysis supported both commonly used treatments of potato taxonomy. Accessions generally clustered based on taxonomic and ploidy classifications with some exceptions but did not consistently cluster by geographic origin. STRUCTURE analysis identified putative hybrids and suggested six genetic clusters in the cultivated potato collection with extensive gene flow occurring among the potato populations, implying most populations readily shared alleles and that introgression is common in potato. *Solanum tuberosum subsp. andigena* (ADG) and *S. curtilobum* (CUR) displayed significant admixture. ADG likely has extensive admixture due to its broad geographic distribution. *Solanum phureja* (PHU), *Solanum chaucha* (CHA)/*Solanum stenotomum subsp. stenotomum* (STN), and *Solanum tuberosum subsp. tuberosum* (TBR) populations had less admixture from an accession/population perspective relative to the species evaluated. A core and mini core subset from the genebank material was also constructed. SNP genotyping was also carried out on 745 accessions from the Seed Savers potato collection which confirmed no genetic duplication between the two potato collections, suggesting that the collections hold very different genetic resources

of potato. The Infinium SNP Potato Array is a powerful tool that can provide diversity assessments, fingerprint genebank accessions for quality management programs, use in research and breeding, and provide insights into the complex genetic structure and hybrid origin of the diversity present in potato genetic resource collections.

KEYWORDS

SolCAP SNP array, *in vitro*, diversity analysis, introgression, *Solanaceae*, genetic analysis, genetic resources

Introduction

Potato was domesticated approximately 10,000 years ago in the Andes and landraces farmed today still have a wide variety of shapes, skin, and tuber colors that are often not seen in modern varieties (Ovchinnikova et al., 2011). Indigenous potato farmers in the Andes often plant 10–30 different landraces of multiple ploidy levels in the same field (Huaman and Spooner, 2002) to ensure production of at least some tubers annually due to seasonal variation in productivity of individual landraces. This risk mitigation strategy helps protect small holder farmers from new and existing biotic and abiotic stresses in potato production in the Andes. Most commercial cultivars in the USA and Europe are autotetraploids (2n = 4x = 48); however, diploids, triploids, and pentaploids are also commonly found among the cultivated native potato landraces regularly farmed throughout the Andes. Some of the challenges in improving cultivated potato results from a high level of heterozygosity, polyploid genetics, adaptation of native landraces to a short-day photoperiod, complex polysomic inheritance, inbreeding depression, narrow genetic base, and biotic and abiotic factors (Simko et al., 2006; Hamilton et al., 2011; Hirsch et al., 2013).

Markers have been used in many crop plants to assess genetic diversity, determine population structure, discover, and track quantitative trait loci (QTL's), produce genetic linkage maps, assist in selection for particular traits, understand the influence of genotypes on phenotypes and more - all to understand or improve genetics and key traits. Many different types of molecular markers have been employed since the 1980's, but single nucleotide polymorphisms (SNPs) are increasingly used due to recent advances in genome sequencing technology and the abundance of SNPs in most crop plants. The affordable cost and high throughput nature of SNP markers have made them powerful tools for genetic analysis of plant species and highly useful in breeding (Bertioli et al., 2014). Discovery of SNPs in simple genomes is relatively easy, requiring collection and evaluation of sequence data; although, in complex genomes such as potato, SNP detection is more difficult due to repetitive segments of the genome and multiple ploidy levels (Mammadov et al., 2012). Genome complexity reduction methods have been developed to aid in the discovery of novel SNPs; nevertheless, it is often challenging to identify SNP markers in polyploids such as potato, tobacco, cotton, canola, and wheat (Mammadov et al., 2012; Bertioli et al., 2014; Logan-Young et al., 2015) due to separating allelic versus homoeologous SNPs which increase the rate of false positives (Clevenger and Ozias-Akins, 2015).

The Illumina Infinium SolCAP 8,303 V1 Potato Array SNPs were originally selected from 69,011 high quality SNPs derived from six commercial potato cultivars 'Atlantic,' 'Premier Russet,' 'Snowden,' 'Bintje,' 'Kennebec,' and 'Shepody' (Hamilton et al., 2011). The Illumina Infinium SolCAP Potato Array V2 contains 12,720 SNPs, including the SNPs from the original V1 (8,303 SNPs) Potato Array with additional markers derived from the Infinium High Confidence SNPs Array (69K, Hamilton et al., 2011), which were selected for genome coverage, candidate genes, and regions with resistance genes. Both potato SNP arrays (V1 and V2) have been used in numerous studies as a genomic tool to improve cultivated potato or gain insight on genetic attributes. These SNP markers were used to measure linkage disequilibrium for genome wide association (GWA) mapping and population structure in European diploid and tetraploid germplasm (Stich et al., 2013). Genotyping a diversity panel of 250 lines of wild species, genetic stocks, and cultivated potato revealed that changes in heterozygosity and allele dosage has not occurred in over 150 years of breeding, but clear selection for alleles in biosynthetic pathways has occurred (Hirsch et al., 2013). The Illumina Infinium SolCAP Potato V1 Array has been used to develop linkage maps (Felcher et al., 2012), genotype populations for QTL analysis (Douches et al., 2014) and assess variation in glycoalkaloid biosynthesis (Manrique-Carpintero et al., 2013, 2014). In other studies, relationships deduced from the SNP markers on the SNP array were generally complementary to existing taxonomic classifications for 74 *Solanum* lines representing 25 wild taxa and were also effective in resolving complex taxa boundaries among germplasm with close genetic relationships (Hardigan et al., 2015).

Conservation of genetic diversity in genebanks is a critical activity, not only for studies of crop diversity, but more importantly for crop improvement in breeding and research. Because breeding efforts and other research using these genetic resources is expensive and time consuming, it is critical that the users of genebank materials receive the genotypic and morphologically correct material they expect. Recent reports have

confirmed errors across the plant research and breeding communities of germplasm in transgenic lines, T-DNA lines, cell cultures, and genetic stocks (Bergelson et al., 2016). Genetic contamination or mixing between accessions can occur in any genebank or program handling large numbers of plant material, especially when phenotypic variation is subtle (Bergelson et al., 2016). In the medical field, it is estimated that up to one third of all cell lines may be contaminated or misidentified (Hughes et al., 2007) and although estimates of error in plant genebanks are not thought to be so high, SNP genotyping revealed approximately 5% error in *Arabidopsis* accessions (Anastasio et al., 2011). Girma et al. (2012) evaluated 3,156 accessions of paired yam accessions and found a 20.6% error rate (not true to type) employing 53 morphological descriptors to assess genetic uniformity among paired samples. The entire cultivated sweetpotato collection at CIP was evaluated for genetic identity using 20 SSR markers along with evaluating morphological characterization in the field of paired samples which demonstrated a total of 19.4% error rate in genetic fidelity among paired samples (*in vitro* compared to original mother plants) [Anglin et al., 2021]. Another study evaluated errors in the sweetpotato breeding program at CIP and found a 27.7% error rate which was suggested to have occurred when moving germplasm from *in vitro*, to screenhouse, and then the field (Gemenet et al., 2020). Unfortunately, identifying and sorting out these errors can be extremely difficult and costly, yet it is necessary for genebank operations and the users of a germplasm collection.

The International Potato Center (CIP) in Lima, Peru houses the world's largest collections of potato. The cultivated potatoes are maintained as clones *in vitro*, in cryopreservation, and distributed worldwide for research, breeding, and education in accordance with the International Treaty for Plant Genetic Resources for Food and Agriculture (ITPGRFA). Internal CIP reports have suggested that identity errors in the *in vitro* germplasm collection have occurred in the past 40+ years of *in vitro* maintenance and regeneration (Perazzo et al., 1999–2000); though, the true extent of these errors within the collection has never been genetically determined. A small subset of the potato collection was previously genotyped using original voucher samples or mother plants and their *in vitro* counterpart and determined an overall error rate of 4.4% in 250 potato accessions (Ellis et al., 2018); however, the entire collection had not been assessed systematically. Further, genetic diversity and population structure of this collection has never been evaluated nor have core or mini core collections been constructed to enhance the utility and access of the genebank collection for users. Therefore, the objectives of this study were to: (i) fingerprint the potato landrace collection in the CIP genebank; (ii) determine if *in vitro* and original clonal mother plants of the same accessions in the collection have identical SNP fingerprints; (iii) evaluate genetic diversity among accessions using SNP markers and assess phylogenetic relatedness; (iv) confirm ploidy levels of material in the collection; (v) reveal population structure among the cultivated taxa; (vi) develop and release a core and mini core collection to represent the majority of the unique genetic diversity in the germplasm collection; (vii) fingerprint potato clones from the Seed Savers Exchange potato collection and compare it to the germplasm collection at CIP to assess overlap of accessions, and (viii) assess if these genetic analyses help to understand the taxonomic challenges existing in the cultivated species in *Solanum* section *Petota*.

Materials and methods

Plant material and DNA extractions

All plant material (Supplementary Table 1) was obtained from the CIP genebank in Lima, Peru. The accessions in this study consisted of landraces and cultivated improved potato accessions from the genebank collection. Based on existing information and curator knowledge of the species, the accessions that were part of this study included 588 diploids, 191 triploids, 2994 tetraploids, 8 pentaploids, and 79 accessions of unknown ploidy level at the time of this study. The unknown material was acquired from the highlands of Peru and had yet to be classified properly. The number of accessions per taxa (based on Hawkes, 1990, the taxonomic treatment used for potato classifications at CIP) included: 2693 (69.8%) accessions of *Solanum tuberosum* (C. Linneo) subsp. *andigena* (Juz. & Bukasov) [ADG], 120 (3.1%) accessions of *S. chaucha* (Juz & Bukasov) [CHA], 94 (2.4%) accessions of *S. stenotomum* subsp. *goniocalyx* (Juz. & Bukasov) [GON], 27 (0.7%) accessions of *S. juzepezukii* (Bukasov) [JUZ], 190 (4.9%) accessions of *S. phureja* (Juz. & Bukasov) [PHU], 270 (7.0%) accessions of *S. stenotomum* subsp. *stenotomum* (Juz. & Bukasov) [STN], 157 (4.1%) accessions of *S. tuberosum* (C. Linneo) subsp. *tuberosum* [TBR], 14 (0.4%) accessions of *S. x ajanhuiri* (Juz & Bukasov) [AJH], eight (0.2%) accessions of *S. curtilobum* (Juz. & Bukasov) [CUR] and 287 (7.4%) accessions of classified as *Solanum spp* [SOL]. The accessions originated though collecting or donations from Argentina (167), Bangladesh (18), Bolivia (428), Bhutan (5), Chile (126), Colombia (218), Costa Rica (1), Ecuador (305), Guatemala (28), India (1), Mexico (30), New Zealand (4), Peru (2379), Philippines (3), Russia (12), Spain (1), Sweden (8), Venezuela (29), and 97 with no country designation (Figure 1). The majority of the collection (97.1%) from this study is derived from collections in South America.

For each accession derived from the CIP genebank collection, DNA was extracted from four *in vitro* plantlets contained in one single test tube and the corresponding original mother plant maintained in the field to evaluate genetic identity among putative clones and to compare identity of the field-maintained samples and *in vitro* samples. The original mother plants have been clonally maintained through the annual regeneration of tubers in the field in the Andean highlands at the CIP field station at Huancayo, Peru (12°01'43"S 75°14'37"W, 3.206 m.a.s.l.). In total, 3860 accessions were used in this study that were determined to be true to type. The accessions found not to be true to type were genotyped but were eliminated from the data set for diversity assessment, phylogeny construction, and STRUCTURE analysis. DNA extractions were performed following a modified CTAB protocol (Doyle and Doyle, 1990). Freeze dried leaf material from 745 potato accessions deposited at the Seed Savers Exchange in Decorah, IA were sent to Michigan State University for DNA extraction. All samples for this study were quantified with a Nanodrop spectrophotometer ND-1000 (Thermo Fisher Scientific

FIGURE 1
Geographic distribution of 3308 accessions of 3860 that had GIS data from the CIP genebank included in this study. (Not all accessions contain this information in the database and if collected a number of years ago is very difficult to ever acquire). The majority of the accessions were collected or donated from regions in the Andean Mountain range down the West Coast of South America. ADG had the broadest geographical distribution.

Waltham, MA) and run on agarose gels to check quality and quantity of each extract. All DNA samples were diluted to 30 ng/µL for SNP genotyping. SNP genotyping was carried out by Neogene (Lincoln, NE) using the Illumina Infinium SolCAP V2 12K potato SNP array.

Flow cytometry

Determination of ploidy levels was done by using SNP data to predict ploidy levels (as described in Ellis et al, 2018) and confirmed by flow cytometer as follows; approximately 50–60 mg of young leaf tissue was chopped with a razor blade in a petri dish containing 250 µL of LB01 buffer (15 mM Tris, 2 mM Na_2EDTA, 0.5 mM spermine.4HCl, 80 mM KCl, 20 mM NaCl, 0.1 % (v/v) Triton X-100, 0.1% β-mercaptoethanol, adjusted to pH 8), [Doležel et al., 1989] to release nuclei (Galbraith et al., 1983). Another 250 µL of LB01 buffer was added and the suspension was incubated for 2 minutes at room temperature. The nuclei suspension was recovered by filtering the cell suspension through a 50 µm CellTrics filter to remove cell fragments and other large debris. Nuclei were stained with 50 µg/mL propidium iodide (PI) and 50 µg/mL RNase was added to the nuclear suspension to prevent staining of double-stranded RNA. Samples were incubated in the dark for two minutes and then analyzed with a BD Accuri™ C6 (BD biosciences, San Jose, CA) Flow Cytometer. The following parameters were used for each sample: Medium Run speed (Fluidics), 1,000 events for the threshold of FL2-H and 10,000 of FSC-H (forward scatter height). The run settings lasted two minutes with at least 400 events for a G0/G1 peak, relative fluorescence intensity of PI-stained nuclei (FL2A and FL3A) and less than 5% of coefficient of variation (CV). Standardization and interpretation of the flow cytometry results were performed using native potato reference standards of known ploidy levels (2x, 3x, and 4x).

SNP genotyping

Samples were assayed using the Illumina Infinium SolCAP 12K V2 Potato Array on the Illumina iScan (Illumina Inc, San Diego, CA). Samples with greater than 20% missing data were genotyped a second time in an attempt to reduce the amount of missing data. Samples that still showed 20% or more missing data were removed from the data set. The SNP data was analyzed with Illumina's GenomeStudio (v 1.0) software. Samples were genotyped as diploids (three cluster calling) and as tetraploids (five cluster calling) to generate the specific ploidy phylogenetic trees (i.e. -phylogeny of all diploids); however, analyses of the entire collection was mainly carried out with five cluster calling genotype data since the majority of the accessions are tetraploids. The samples were genotyped as diploids (AA, AB, BB) using the GenomeStudio software auto-clustering feature and the SolCAP custom three cluster calling file (Felcher et al., 2012). The samples were also genotyped with five cluster calling developed for tetraploids (nulliplex=AAAA, simplex=AAAB, duplex=AABB, triplex=ABBB, and quadruplex=BBBB). A total of 5,031 SNPs were obtained from five cluster calling and 8,045 with diploid calling. SNPs that did not produce a clear signal in ≥10% of the individuals or could not be clustered, were removed along with SNPs noted in previous studies to be poor or questionable (http://solcap.msu.edu/potato_infinium.shtml). Genotyping data was deposited in CIP's dataverse repository and can be found here https://data.cipotato.org/dataset.xhtml?persistentId=doi:10.21223/LBCFCF.

Data analysis

The sample data set was subsequently filtered to only include putative unique accessions for the evaluation of inter- and intraspecific relationships and to perform STRUCTURE analysis. Principal Components Analysis (PCA) was also employed using the R package 'adegenet'. The phylogenetic trees were constructed using hierarchical cluster analysis using the unweighted pair group method with arithmetic mean (UPGMA) to calculate the distance matrix in R version 3.2.2 (R Core Team, 2015). The color labels and palettes for the species tree were created using the Dendextend version 1.1.2 and RColorBrewer version 1.1–2 packages in R. iTOL was utilized to annotate and color segments of the tree (Letunic and Bork, 2016).

Structure

Population structure was estimated using the program STRUCTURE version 2.3.4 (Pritchard et al., 2000; Falush et al.,

2003) by assigning the 3,860 accessions to populations or multiple populations based on genotypes produced from 8045 SNP markers. This program infers population structure using a Bayesian approach which identifies clusters and assigns individuals to specific clusters based on Hardy-Weinberg equilibrium and linkage equilibrium. Lambda was initially determined to be 0.5433 for this data set by setting K=1 (the number of populations) and was subsequently set to 0.5433 for the remaining runs. Multiple runs of STRUCTURE were performed by setting K from 1–10. Because STRUCTURE is computationally intensive, the length of burn in and number of MCMC reps was set fairly low to explore the data at different K values and then subsequently increased when an optimal K was discovered. Therefore, the burn-in length was initially 10,000 and MCMC replications were 20,000 for each run (K=2–10). The runs were replicated three times for the initial evaluation. The models employed were admixture and correlated allele frequencies. STRUCTURE Harvester (http://taylor0.biology.ucla.edu/structureHarvester/) was used to determine the appropriate K value (Dent and vonHoldt, 2012) from the results produced. A K=6 was determined to be the optimum. Once the most appropriate K value was determined using a low number of burn-in and replications, the runs were repeated at K =5–7 with an increased burn-in length of 500,000 with MCMC replications of 500,000 to improve the robustness of the Bayesian analysis. Each K value was run in triplicate.

Predictive taxonomic classification system using STRUCTURE data

The six genetic clusters delineated by STRUCTURE were color coded and the percent of each color was expressed on an accession basis using the color scales rule with conditional formatting in EXCEL. From this, a classification system was derived to describe the frequency of STUCTURE genetic clusters for each species (Table 1A). This predictive classification system was tested with known and unknown accessions to determine its utility in the prediction of unclassified accessions into species This classification system was then applied to each accession to see how well it could blindly predict the species determination based solely on morphological data.

TABLE 1 Derivation of a classification system for the prediction of cultivated potato species using the relative % frequency of the six identified STUCTURE populations (represented by six different colors) for each species using the color scales rule with conditional formatting in EXCEL where the relative % frequencies in the cells are color-coded from red (low frequency) to green (high frequency).

A)

Species	STRUCTURE population color					
	yellow	red	blue	pink	teal	green
ADG	>10%					>50%
CHA		50–100%			<10%	<30%
GON		>50%			10–40%	NO
AJH		10–50%	NO	20–100%		
JUZ		10–50%	10%	20–100%		NO
PHU	NO		NO		>80%	NO
STN	<20%	>50%	NO	<60%	<50%	<50%
TBR	<10%		>50%	<40%	<5%	<30%
CUR		10–50%	<10%	20–100%		<25%

– low to no value of this STRUCTURE population.
NO – zero value for this population.

B)

Acc #	yellow	red	blue	pink	teal	green
706032	0	0.565	0	0	0.434	0
706214	0.001	0.704	0	0	0.295	0
706669	0	0.672	0	0	0.327	0

(Continued)

Continued

B)						
Acc #	yellow	red	blue	pink	teal	green
706747	0	0.847	0	0	0.152	0
706831	0	0	0	0	1	0
705211	0.001	0	0.02	0.218	0.003	0.758
705260	0.021	0.002	0.441	0.011	0.004	0.522
705315	0.103	0.002	0	0.007	0.003	0.885
705328	0.068	0.115	0	0.022	0.075	0.72
707848	0.089	0.151	0.071	0.506	0.001	0.182
709027	0.002	0.006	0.792	0.101	0.003	0.096
709028	0.002	0.02	0.802	0.002	0.001	0.173
709029	0.002	0.019	0.801	0.003	0.003	0.172
709031	0.002	0.018	0.807	0.001	0.003	0.168

C)						
Acc #	yellow	red	blue	pink	teal	green
706032	0	0.565	0	0	0.434	0
706214	0.001	0.704	0	0	0.295	0
706669	0	0.672	0	0	0.327	0
706747	0	0.847	0	0	0.152	0
706831	0	0	0	0	1	0
705211	0.001	0	0.02	0.218	0.003	0.758
705260	0.021	0.002	0.441	0.011	0.004	0.522
705315	0.103	0.002	0	0.007	0.003	0.885
705328	0.068	0.115	0	0.022	0.075	0.72
707848	0.089	0.151	0.071	0.506	0.001	0.182
709027	0.002	0.006	0.792	0.101	0.003	0.096
709028	0.002	0.02	0.802	0.002	0.001	0.173
709029	0.002	0.019	0.801	0.003	0.003	0.172
709031	0.002	0.018	0.807	0.001	0.003	0.168

D)							
Acc #	yellow	red	blue	pink	teal	green	spp prediction
706032	0	0.565	0	0	0.434	0	gon
706214	0.001	0.704	0	0	0.295	0	gon
706669	0	0.672	0	0	0.327	0	gon
706747	0	0.847	0	0	0.152	0	gon

(Continued)

Continued

D)							
Acc #	yellow	red	blue	pink	teal	green	spp prediction
706831	0	0	0	0	1	0	phu
705211	0.001	0	0.02	0.218	0.003	0.758	adg
705260	0.021	0.002	0.441	0.011	0.004	0.522	adg
705315	0.103	0.002	0	0.007	0.003	0.885	adg
705328	0.068	0.115	0	0.022	0.075	0.72	adg
707848	0.089	0.151	0.071	0.506	0.001	0.182	cur
709027	0.002	0.006	0.792	0.101	0.003	0.096	tbr
709028	0.002	0.02	0.802	0.002	0.001	0.173	tbr
709029	0.002	0.019	0.801	0.003	0.003	0.172	tbr
709031	0.002	0.018	0.807	0.001	0.003	0.168	tbr

A) Classification system to predict the species of cultivated potato based on the frequency of different STRUCTURE populations from accessions with confirmed species designations. B) Example of the raw data from taxonomically undetermined species (X) with relative % frequency of the six STUCTURE populations in EXCEL. C) Example of data from B) expressed with conditional formatting in EXCEL. D) Data from C) with prediction of species based on STRUCTURE populations frequency.

Core and mini core designation

The software program Core Hunter ver. 3.0.1 (De Beukelaer et al., 2018) in R was utilized to designate a core collection from the SNP data. This program can construct a core collection based on genetic distances calculated from marker data or phenotypic traits. Specific algorithms are employed to optimize selection of entries based on metrics specific of the purposes of a core collection (represent whole diversity, extreme genotypes, or a distribution pattern). In this study, a core collection was built to represent the genetic diversity of the 3860 accessions in the CIP collection using the accessions to nearest entry function (ANE). Five cluster calling from the SNP data (nulliplex=AAAA, simplex=AAAB, duplex=AABB, triplex=ABBB, and quadruplex=BBBB) and species designations were utilized for this analysis. A selection of 11.68% of the total accessions were chosen for representation in a core collection. After Core Hunter produced results, the selections were evaluated, and minor modifications were made based on knowledge of agronomic values, traditional knowledge, and trait information. These minor modifications made by the potato curator were only performed for closely related accessions within a grouping. An ultimate core collection of 451 accessions was produced. The same process was repeated for the selection of the mini core collection resulting in a mini core of 45 accessions.

Results and discussion

Genetic parameters and ploidy level

The Illumina Infinium SolCAP V2 12K potato SNP array was employed to evaluate genetic identity, diversity, and population structure of the global *in trust* potato germplasm collection at CIP. The potato germplasm collection at CIP is composed of landraces and native potato material mostly originating from the Andes region of South America. The species classification used in this study is based on the taxonomy of Hawkes (1990). Although Spooner taxonomy (Spooner et al., 2007) is widely used in potato classifications, CIP and other institutions worldwide have historically used the taxonomic treatment of Hawkes's (Hawkes, 1990) and still employ it to classify potato due to familiarity, history of Hawkes working at CIP, and its utility for classifying genebank materials. A total of 3,308 of these accessions have GIS data in their passport data. A map of the geographic distribution of accessions and species that had GIS data can be seen in Figure 1 with ADG having the broadest geographic distribution of any of the potato species. The majority of the collection (~95%) that was part of this study was derived from South America which is typically very unique (mostly all short day adapted), compared to North America and European germplasm (long day adapted).

For each accession in this study, DNA was extracted from both a bulk of four *in vitro* plantlets contained in one single test tube and a corresponding original mother plant to evaluate genetic identity among putative clones maintained of the same accession. Genetic identity was determined through genotyping with the Illumina Infinium SolCAP Potato Array (V2) and morphological comparisons of the mother plant with an *in vitro* plantlet of the same accession was done in the field to ensure basic plant traits (descriptors) were equivalent among the pair. Only material determined to be true-to-type (TTT) by SNP fingerprinting (*in vitro* and mother plant matched in genotype) were used for analysis in this study as there were reports of identity errors in the CIP collection (Perazzo et al., 1999–2000; Ellis et al., 2018; Anglin et al., 2021) as well as errors being reported in other plant collections (Anastasio et al., 2011; Girma et al., 2012; Gemenet et al., 2020). The

genotyping data demonstrated an error rate of 19.9% between the matching pairs of clones. In total, 3,860 accessions had matching fingerprints for a mother plant with their *in vitro* counterparts and matched morphologically, and thus, these accessions were further used in this study to evaluate genetic diversity and population structure. The remaining accessions that contained errors (did not have matching genotypes between mother plant and *in vitro* pairs) were not included in this study and will be evaluated further to correct the errors in the collection prior to being distributed out to users or utilized for research purposes.

The five cluster SNP data was applied to predict ploidy levels of all the accessions. Previously in Ellis et al. (2018), SNP data was utilized to predict ploidy and subsequently confirmed with flow cytometry, and this method was also repeated here, since species classification in potato is not always an adequate indicator of ploidy level (Ghislain et al., 2006; Ellis et al., 2018). In both studies, SNP data was found to be an accurate and fast indicator of ploidy levels. Many of the accessions in the collection were originally assumed to be a particular ploidy solely based on their species classification or morphological characterization and never evaluated further. Thus, the SNP data were utilized to confirm ploidy level of each accession since this is an important trait for breeding and management of the collection. Overall, 92.5% of the potato accessions had ploidy data that matched historical data based mostly on morphological and species designation and 7.45% (288 accessions) needed adjustments. Of those, seventy-seven accessions (of 288) were not yet classified to a species level, and thus, had no ploidy data collected until the SNP data was applied. These data were confirmed with flow cytometry and the predicted ploidy based on SNP data matched the flow cytometry data. Where ploidy levels did not match the original ploidy determinations, flow cytometry was used to confirm the ploidy levels and, in all cases, the predicted ploidy based on SNP was confirmed with the flow cytometry data. Importantly, based on these data, most potato species had a mix of various ploidy levels with no species being purely one single ploidy. For example, AJH was mostly diploid (79%), yet some AJH accessions were determined to be either triploid (14%) or tetraploid (7%). PHU was similar in that most of the PHU accessions were diploid (87%), but a few accessions were triploid (6%) and tetraploid (7%). STN was mostly diploid (74%); however, some triploids (8%) and tetraploids (18%) were revealed. All of the species had some minor variation of ploidy levels within the species; however, some were only variable for one accession or very few accessions. Some of these observations could be due to incorrect species classification of an accession, from mixed ploidy within a species, or both classification and mixed ploidy (Supplementary Table 1).

The SNP data were also utilized to calculate heterozygosity. The heterozygosity in the entire potato collection ranged from 9.7% to 66.6% with an average heterozygosity of 33.5% (Figure 2). The TBR accessions were the most heterozygous species in the collection which was also seen in a previous study evaluating a subset of the CIP potato collection (Ellis et al., 2018). TBR was more heterozygous than even the pentaploid accessions (CUR). There are, however, very few CUR accessions (eight) within the germplasm collection compared to 157 accessions of TBR which could have lowered the heterozygosity in CUR. The ADG accessions were the most variable species with a wide range of heterozygosity (10.8 – 62.8%) along with the accessions only classified as *Solanum* (SOL) which included diploid, triploid, and tetraploid accessions. Overall, heterozygosity tended to increase with ploidy level where species generally thought to be diploids had lower heterozygosity than species with higher ploidy levels (3x, 4x, and 5x) with heterozygosity maximized with some of the tetraploids.

PCA

Because of the size of this data set, the 3,860 true to type (TTT) accessions were analyzed by Principal Components Analysis (PCA) to help reduce the dimensionality of the data, yet still preserve much of the variability (Figure 3). The first principal component contained 11.28% of the genetic variance and the second component contained 4.71% of the variance. Many of the species in the PCA were overlapping suggesting they are highly correlated. This suggests significant introgressions or hybridization has occurred among these species and accessions or the genus is over described. The ADG species shown in green, split into two distinct groupings, one of which overlapped slightly with TBR. This was also observed in a separate study evaluating the Colombian Central Collection of potatoes (Manrique-Carpintero et al., 2023). The other portion of ADG overlaps with most of the other cultivated potato species. The STN, GON, and CHA species also appear to be highly correlated with PHU and some of the accessions only classified as *Solanum* species.

Phylogeny

Phylogenetic trees were constructed for the entire set of true-to-type cultivated genebank accessions (3,860 accs) to look at inter- and intraspecific relationships on a collection wide basis (Figure 4). Generally, species and accessions in the entire collection (3,860 accs) of different ploidy levels clustered as in a previous study (Ellis et al., 2018) which assessed genetic variability of a limited number of material but included representative accessions of the cultivated potato species from the CIP genebank. The *S. x ajanhuiri* (AJH), *S. curtilobum* (CUR), *S. x juzepeczukii* (JUZ) accessions, which are considered bitter potatoes, grouped together even though they consist of diploids, pentaploids, and triploids, respectively. *S. tuberosum* subsp. *andigena* accessions (ADG) which represented 69.8% of the true-to-type cultivated collection, split into different clades, one of which was more similar to bitter potato accessions (AJH, CUR, JUZ) and TBR, with the other clade clustering on its own. *S. stenotomum* subsp. *goniocalyx* (GON) and *S. stenotomum* subsp. *stenotomum* (STN) were intermixed as was seen previously (Ellis et al., 2018) and did not cluster monophyletically. Hawkes (1990) classified these as subspecies suggesting that STN and GON did not have significant morphological differences and are highly related. Spooner et al. (2007) previously collapsed these subspecies

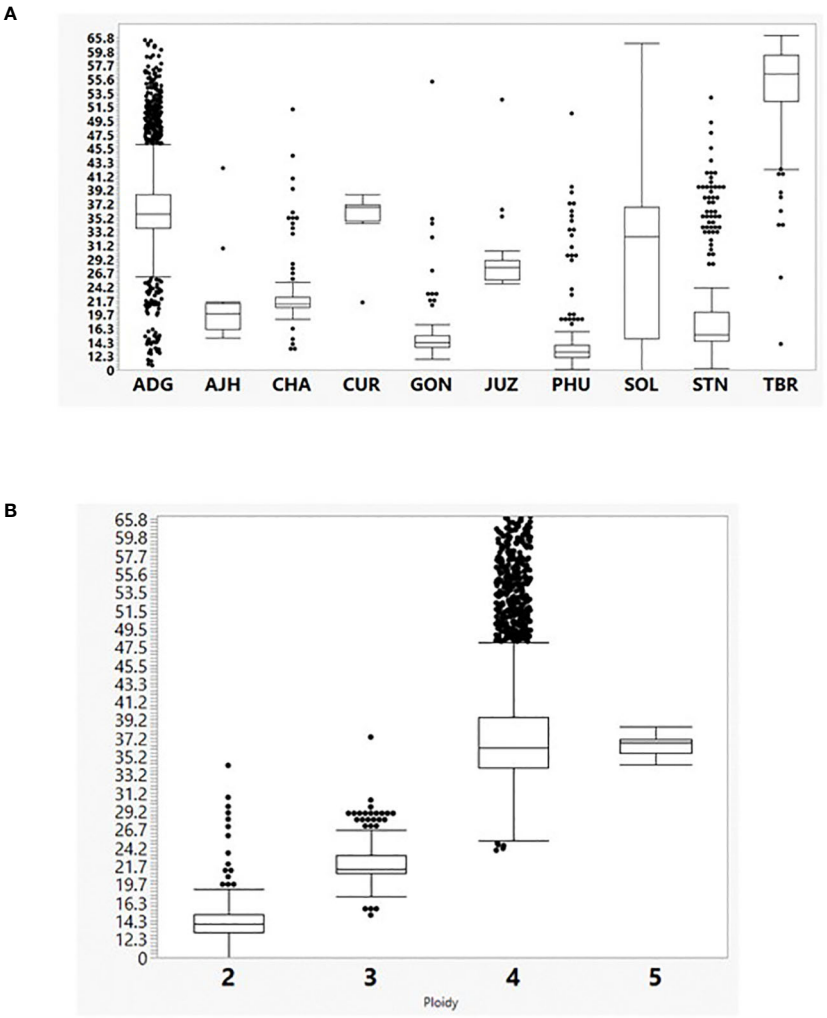


FIGURE 2
(A) The percent heterozygosity (y axis) of the 3860 cultivated potato CIP genebank accessions used in this study grouped by species classification (x axis). Hawkes, 1990: *Solanum tuberosum* (C. Linneo) *subsp. andigena* (Juz. & Bukasov) [ADG], *S. chaucha* (Juz & Bukasov) [CHA], *S. stenotomum* subsp. *goniocalyx* (Juz. & Bukasov) [GON], *S. juzepezukii* (Bukasov) [JUZ], *S. phureja* (Juz. & Bukasov) [PHU], *S. stenotomum* subsp. *stenotomum* (Juz. & Bukasov) [STN], *S. tuberosum* (C. Linneo) subsp. *Tuberosum* [TBR], *S. x ajanhuiri* (Juz & Bukasov) [AJH], *S. curtilobum* (Juz. & Bukasov) [CUR] and unclassified accessions as *Solanum spp* [SOL]. **(B)** Percent heterozygosity (y-axis) grouped by ploidy level based on SNP data.

(STN and GON) along with other potato species (PHU, ADG, and CHA) into a single group: *S. tuberosum* Andigenum group. The results here support Spooner taxonomy with the lumping of STN and GON into a single taxonomic group.

Previous studies have shown that species designation is not a good indicator of ploidy level in potato (Ghislain et al., 2006; Spooner et al., 2007; Ovchinnikova et al., 2011; Ellis et al., 2018). Therefore, all accessions that had a profile consistent with a diploid pattern from the SNP data were grouped together and a phylogeny was constructed (Figure 5) to assess interspecific relationships among diploid potatoes. In general, this included accessions from *S. phureja* (PHU), *S. stenotomum* subsp. *goniocalyx* (GON), *S. stenotomum* subsp. *stenotomum* (STN), *S. ajanhuiri* (AJH), accessions only classified as *Solanum*, and a few miscellaneous accessions from other taxa that appeared to be diploid. The majority of *S. phureja* (PHU) accessions clustered together. A few exceptions are noted, and it is possible these are either hybrids or the species designation of these accessions are misclassified. Accessions classified as *S. ajanhuiri* (AJH) also clustered together. Both Spooner et al. (2007) and Hawkes (1990) considered AJH as a distinct potato species. Additionally, a few genetic redundancies were apparent in the tree among these diploid accessions.

A phylogeny of the triploids and pentaploids was also constructed (Figure 6) to examine intra- and interspecific relationships and none of the taxa formed a distinct monophyletic clade. Overall, the majority of the triploid accessions (JUZ and CHA) displayed little genetic variability with short branch lengths between the various accessions suggesting high genetic similarity. Since triploids are generally an evolutionary dead end, it is reasonable to believe that their variability would be low. There were eight pentaploids included in the analysis of which four accessions were identical based on the SNP data suggesting they are likely genetic duplicates. The majority of the accessions of bitter potatoes (AJH, CUR, and JUZ) with one exception (CIP 707124

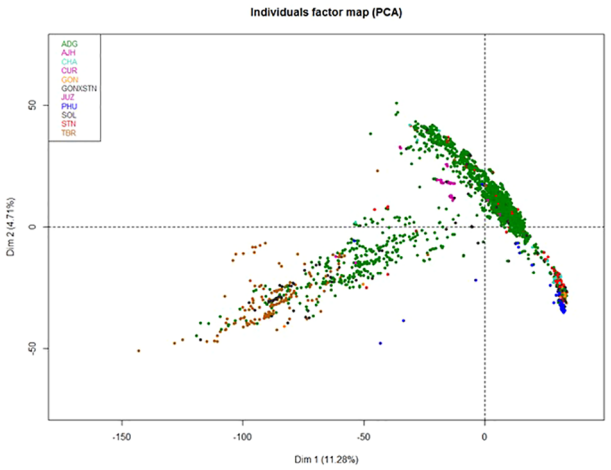


FIGURE 3
Principle Component Analysis (PCA) of the 3860 potato accessions from the genebank genotyped by the SolCap SNP Array.

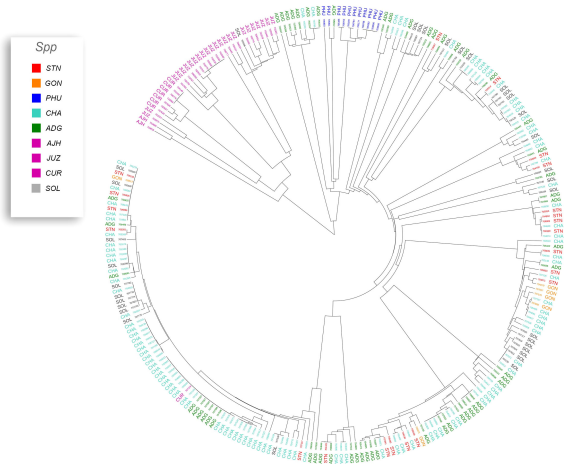


FIGURE 5
Dendrogram of all diploids in the cultivated potato collection.

CUR) all clustered together. All the species of bitter potatoes were considered as unique by both Hawkes (1990) and Spooner et al. (2007). *S. chaucha* (CHA) also had many genetic similarities among accessions shown by short branch lengths between individuals which supports the suggestion above that triploids have low genetic variability. Some of the unclassified *Solanum species* accessions were genetically similar, based on the SNP analysis to other classified accessions suggesting they are redundant. Further, some potato accessions classified as other species also grouped within some of these non-variable regions in the phylogeny. This data suggests either that SNPs from the array may not be sufficiently dense enough in the genome to genetically distinguish these accessions as different species or there is very little genetic variability in the accessions that are triploid and pentaploid. It is also possible that ascertainment bias from the SNP array is the cause of this due to the lack of rare variants. Another possibility is that these accessions were duplicates or closely related which was not apparent when they were originally introduced into the collection. One potential drawback of utilization of SNP arrays on diverse material is that SNP arrays can lead to ascertainment bias especially when the markers on an array were discovered from a small number of samples or samples that do not represent the broader population (Caruana et al., 2019). Arrays can lack a significant proportion of rare variants and be biased towards variants in the populations used to develop the respective array (Geibel et al., 2021).

Structure

Population structure analysis allows for understanding of gene flow, admixture, and inference of demographic histories of populations (Stift et al., 2019). Simulated data has demonstrated that the program STRUCTURE is more robust than other methods in

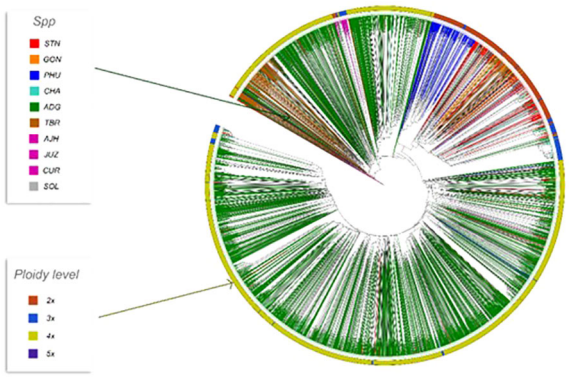


FIGURE 4
Dendrogram based on the data from the Illumina Infinium SolCAP V2 12K potato SNP array of the 3860 accessions of cultivated potato germplasm from the genebank used in this study. The outside ring is ploidy level. The branches are colored by species designation in the CIP database.

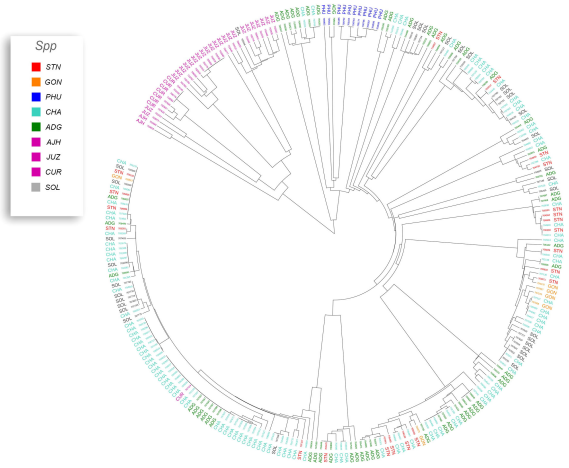


FIGURE 6
Dendrogram of triploids and pentaploids in the cultivated potato collection.

handling biases due to mixed ploidy levels (Stift et al., 2019) and because this data had a mixture of diploid, triploid, tetraploid, and pentaploid accessions, STRUCTURE was employed to provide a robust genetic structure analysis and understand gene flow and admixture of a large pool of potato landrace germplasm. The genotyping data was used in the program STRUCTURE to determine the number of genetic clusters in the genebank collection. A total of six genetic clusters were observed (Figure 7). One striking feature is the high level of admixture and substantial gene flow observed among the majority of accessions genotyped suggesting and supporting the notion of high heterogeneity of potato. The extensive gene flow was especially apparent in the ADG accessions which appeared to pick up alleles from every lineage and contained significant introgressions within most accessions. Hoopes et al. (2022) using whole genome sequence (WGS) data recently reported significant allelic diversity in cultivated tetraploid material from Europe and North America demonstrating ancestral introgressions from wild species that predated breeding efforts. All the cultivated species from South America were included in this study and many of these could be hybrids, with either wild species or rare cultivated material not in the genebank collection in their background as some of these introgressions do not appear to be derived from the cultivated species included in this work. More genotyping of wild potatoes would be required to determine the origin of putative wild alleles as they do not appear to be completely derived from the cultivated collection.

The phylogeny of the ADG accessions split into two major groups and the STRUCTURE data suggests that one of these genetic groupings had more alleles shared with the bitter potatoes represented in pink (AJH, CUR, and JUZ) which explains why a subset of the ADG accessions clustered in the phylogenetic tree with the bitter potatoes and TBR, which also contains this introgression. After closer morphological examination, these ADG accessions with this introgression, have characteristics more similar to the wild potato species and tend to be more bitter to the taste than the other ADG accessions, suggesting higher glycoalkaloid level in these tubers than the remaining ADG accessions which do not contain the putative bitter alleles (pink). Spooner et al. (2007) and Hawkes (1990) both recognized AJH, CUR, and JUZ as distinct species. Yet, it is interesting that the bitter potatoes despite being different ploidies, and different species according to both Spooner and Hawkes, cluster together in the phylogeny (Figure 4) and share fairly similar introgressions\population structure giving the appearance they are genetically similar or related. Hawkes (1990) proposed that CUR arose as a hybridization from JUZ and ADG. The STRUCTURE data supports this showing CUR having shared alleles with the bitter potatoes (pink) and ADG (green and yellow). Hawkes (1990) also proposed that AJH arose from a hybridization among STN x *S. megistacrolobum* (wild species) and that JUZ was a hybrid of STN x *S. acuale* (wild). While there were no wild species included in this dataset, both AJH and JUZ show signatures of introgression from the CHA/STN population (red) partially supporting Hawkes's hypothesis. More data would be required from the wild species to evaluate these hybridization theories of Hawkes; however, the cultivated species proposed by Hawkes are supported by these data.

The STRUCTURE data suggests that GON is a hybrid of PHU and the CHA/STN lineage with a larger contribution of the alleles

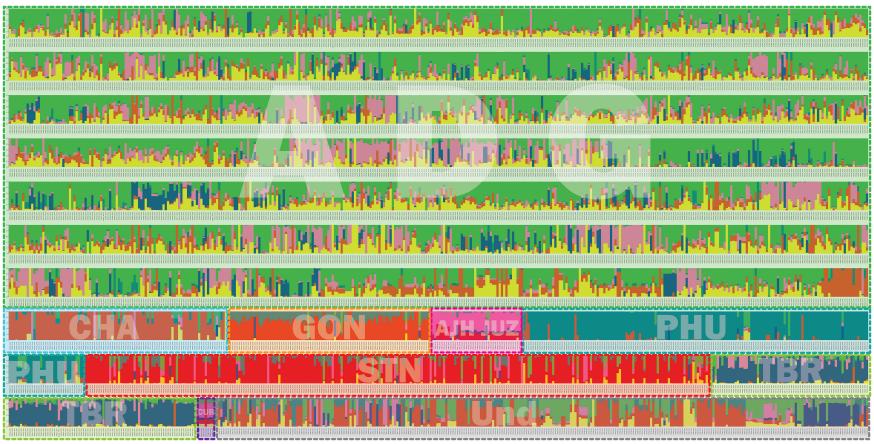


FIGURE 7
Structure data for 3860 cultivated potato accessions from the CIP genebank used in this study. The species are represented by 3 letter codes with "ND" signifying not determined as they were new introductions and not fully classified at the time of this work. The accessions are classified per taxa based on Hawkes, 1990 which has long been the only taxonomic system for potato classifications at long been the only taxonomic system for potato classifications at CIP and included: *Solanum tuberosum* (C. Linneo) *subsp. andigena* (Juz. & Bukasov) [ADG], *S. chaucha* (Juz & Bukasov) [CHA], *S. stenotomum* subsp. *goniocalyx* (Juz. & Bukasov) [GON], *S. juzepezukii* (Bukasov) [JUZ], *S. phureja* (Juz. & Bukasov) [PHU], *S. stenotomum* subsp. *stenotomum* (Juz. & Bukasov) [STN], *S. tuberosum* (C. Linneo) subsp. *tuberosum* [TBR], *S. x ajanhuiri* (Juz & Bukasov) [AJH], *S. curtilobum* (Juz. & Bukasov) [CUR] and unclassified accessions as *Solanum spp* [ND]. The six populations are represented by colors = green, pink, yellow, red, dark blue, and turquoise.

from CHA/STN lineage than the PHU lineage. These results support Spooner taxonomy (2007) which eliminated GON as a separate species/subspecies and lumped it together within the *S. tuberosum* Andigenum group. However, Spooner also lumped CHA and STN into the *S. tuberosum* Andigenum group, whereas Hawkes kept these as separate species or subspecies. Hawkes proposed GON as a species that diverged from STN, and these data show the divergence is likely due to an introgression from the PHU population. PHU and GON do have an overlapping geographical distribution with PHU ranging from Venezuela to central Bolivia and GON found from Northern Peru to central Bolivia (Figure 1). The STRUCTURE and phylogenetic data partially support Spooner taxonomy in that CHA and STN appear to be derived from a single population and are not unique. In contrast, PHU does appear to be derived from a unique population; however, PHU was lumped by Spooner into the *S. tuberosum* Andigenum group. Furthermore, CHA and STN appear to share alleles from the same population also partially supporting Spooner taxonomy of lumping these together into the *S. tuberosum* Andigenum group.

Three of the six genetic clusters identified in STRUCTURE (PHU, CHA/STN, and TBR) had lower introgressions overall, with the remaining genetic clusters arising through extensive introgression/hybridization, which appears to be a continuous process in potato. These results suggest and confirm the promiscuity of potato in picking up and exchanging alleles whenever and wherever possible. Several of the accessions classified as TBR, the most cultivated potato species in USA and Europe, appear to have accessions displaying significant introgressions from the bitter potatoes represented with the pink color (CUR/JUZ/AJH) and some introgression from ADG. Both Spooner and Hawkes recognized TBR as a distinct species which is supported by this analysis as well. Several of the ADG accessions have some introgressions from the TBR population as well, along with other introgressions which may be derived from wild species. The cultivated and wild potato curators noted wild characteristics and more bitterness in tubers of the ADG accessions with the pink coloring in the STRUCTURE (Figure 7) supporting the hypothesis that these may be derived from wild species or have significant wild species introgressions. Further, ADG and CUR accessions were the most admixed species in this data set. This is likely due to its wide geographical distribution of ADG relative to the other potato species, which allowed ADG to gain access to other potatoes (possibly wilds) and shared alleles freely producing all the introgressions.

Utilizing the data from the STRUCTURE analysis, a classification system was derived using the percent relative frequency of the six STUCTURE populations for each species (Table 1) and then applied to the accessions which were classified as unknown (SOL) or of questionable species determination and for the putative hybrids along with confirmation of species designation performed independently in the field. The accuracy of the classification system varied with species yet proved highly predictable (>75% correct) for species designation of unknowns of ADG (75% correct), GON (76% correct), PHU (100% correct), STN (88% correct) and TBR (100% correct). In contrast, the classification system derived for AJH, CUR, and JUZ, which remain as recognized species by Spooner et al. (2007) were not predictive for the putative species likely due to the low number of accessions tested of these species was 5, 1, 1, respectively. Interestingly, CHA alone also had very low predictability (6% correct); however, all incorrect calls were classified morphologically as STN (50%) or STN hybrids, with which CHA groups in the dendrogram (Table 1). Overall, the percent of correct calls confirmed by morphological analysis was 76% with individual pure species calls however, if putative hybrids were included, the predictive values of this classification system increased to 87% if a positive species call was extended to one of the suspected hybrid parents.

These data support Hawkes (1990) taxonomy in that the use of genetic data, SNP data in this case, had predictive value in identifying and differentiating STN, PHU, GON, and TBR. These data also correctly predict ADG although the STRUCTURE data shows a very high level of admixture in ADG, and thus, the derived classification system only relies on two of the six STRUCTURE genetic clusters. Prediction of species in genebank accessions can help fill in major gaps in databases allowing users to better select individual accessions for their research needs. Further, it will help the genebank delineate material as many of the accessions in this study were only classified as *Solanum species*.

This work further demonstrates the utility of complimentary methods phylogenetics, STRUCTURE, PCA, and a classification system produced from SNP genotyping data for reticulate phylogenetic relationships, allowing a new perspective on the genebank collection that was not possible to understand previously from morphological data alone. This is especially applicable to many germplasm collections in which growing out the entire collection and evaluating it in a single year is not feasible. Each of these analyses have different algorithms \ approaches but taken as a whole lead to a robust analysis and insights on the species, hybrid origin of individuals or species, and the overall diversity of the collection. These data revealed accessions that are genetically similar or duplicates which, once verified, can now be archived or eliminated for cost efficiencies, and reduce the high cost of maintaining clonal collections. The data also revealed accessions that were misidentified taxonomically which is important to correct for end users to be able to target the species or accessions they desire. The STRUCTURE data was a helpful tool to identify misclassified accessions in the genebank collection and visualize the extent of admixture within and between accessions of the same genetic lineage. For example, several of the accessions labeled as CHA do not appear to be classified properly as their STRUCTURE data does not fit the pattern of other CHA accessions nor do they fit the predictive classification system derived from the data. CIP 707124 which was originally classified as CUR appears derived 100% from the CHA/STN population in STRUCTURE which explains why this accession clustered with CHA in the phylogeny and is predicted as CHA from the classification system. Another example, CIP 703882 which was classified as GON and thus assumed to be a diploid, yet the SNP data suggested it as a tetraploid, the STRUCTURE data did not appear like other accessions of GON with admixture more similar to ADG, and this accession did not cluster with other GON accessions and is

likely ADG. CIP 703545 originally classified as PHU also did not produce a pattern in STRUCTURE similar to the other PHU accessions, nor did it cluster with the PHU accessions and clustered with CIP 703668 STN suggesting it is misclassified as a PHU. The classification system offered two possible designations, either GON or STN, which can be verified in the field. The collection also had 287 accessions that were only identified as SOL and without a potato taxonomist to evaluate them, they have remained unclassified. The data collected here helps parse those accessions into different ploidies and into a species designation which can be easily predicted based on their STRUCTURE admixture along with their clustering in a phylogenetic tree. Indeed, the derivation of a classification system from the STRUCTURE data identified these formally unknowns to species with >75% accuracy based on field morphological determinations. These types of analyses can aid genebanks in determining and classifying taxa and benefit the overall management of the collection.

Seed Savers

Seed Savers is an organization that maintains heirloom and historical varieties. Their mission is to collect, regenerate, store, and distribute varieties (https://www.seedsavers.org/mission) and have been operating since 1975. Their mission is analogous to the mission of any genebank. In order to understand the potential overlap in the Seed Savers potato germplasm collection and that of CIP, Seed Savers was contacted, and a portion of their potato collection was also fingerprinted. A total of 745 potato accessions from the Seed Savers collection were genotyped with the Illumnia Infinium SolCap V2 potato SNP Array and a dendrogram was produced showing the variability of the potatoes in the Seed Savers collection (Figure 8). Overall, much of the Seed Saver collection consists of unique accessions with divergent fingerprint patterns. Relatively few accessions seemed to be genetically similar to one another indicating that the portion of the collection which was genotyped does not contain a high number of genetic redundancies. Further, when compared to the entire CIP collection, no redundancies were found between the two collections suggesting that these potato collections are quite different resources, each contributing unique diversity for users of the collection. Most of the material from the Seed Saver collection appears to be more related to CIP's TBR accessions based on a combined phylogeny (Figure 8) of the Seed Savers and CIP material which makes sense as the Seed Savers collections are heirloom varieties from North America which have long day adaption and therefore are quite different to the short day adapted germplasm maintained at CIP.

Core and mini core collection

It is challenging for users of germplasm collections to narrow down the total accessions from an entire collection to a manageable subset for use in research projects. One strategy that has been useful is the development of core and mini core collections from genebank collections (Frankel and Brown, 1984). A core collection is a subset of the entire germplasm collection, typically 10%-15% of the total collection that represents the majority of the genetic diversity in the entire germplasm collection with little redundancy (Brown, 1989). This concept was developed to improve the use of the germplasm materials since genetic resource collections can contain several thousands of accessions which is unwieldy for researchers to mine (Upadhyaya and Ortiz, 2001) and to help researchers find traits of interest in a smaller, more manageable subset. A mini core collection takes this idea one step further with additional reductions to target a sub-collection by sampling the diversity in approximately only 1% of the germplasm collection, so that traits expensive to measure can be assessed in an even smaller subset. This concept has been useful in previous studies to identify germplasm from the larger collection which contain a trait(s) of interest. If a particular trait is found in a core or mini core, a researcher can identify the original grouping of the accessions from the entire collection in the design of the subset, and subsequently track back to that cluster of accessions for identification of more accessions with the trait(s) of interest. Also, because the potato germplasm collection is rather large and, as a clonal crop, is expensive to distribute as phytosanitary clean *in vitro* clones (in comparison to distribution of seed germplasm), frequently clonal genebanks need to greatly limit total germplasm distribution to users. Core and mini core collections aid in this by allowing genebanks to send requestors a smaller collection of diversity represented in the collection to aid breeders and researchers in targeting desired accessions for trait evaluations. A core and mini core collection has not previously been developed for the CIP genebank material.

Although Core Hunter (ver 3.01) was the primary tool utilized to preselect the core subset, after selection by Core Hunter, the potato curator used agronomic traits and traditional potato knowledge to adjust a few of the selections based on over 30+ years of experience with potato and extensive knowledge on user preference resulting in a selection of 451 accessions for the CIP cultivated potato core collection. This core collection consists of 238 ADG, nine AJH, six CHA, seven CUR, 18 GON, seven JUZ, 55 PHU, 51 STN, 46 TBR, and 14 SOL accessions. A similar approach was then utilized to select 45 accessions from the 451 core accessions to develop the mini core collection to create a more manageable subset for research and trait identification (Figure 9). The mini core passport data along with pictures of the tubers can be seen and ordered through the CIP Genebank website https://genebank.cipotato.org/gringlobal/methodaccession.aspx. The mini core contains 19 ADG, two AJH, three CHA, two CUR, three GON, two JUZ, five PHU, five STN, and four TBR accessions. The dendrogram of the core and mini core can be seen in Figure 9 in relation to the entire germplasm set, demonstrating good coverage relative to the entire germplasm collection with every cultivated species represented. The core and mini core collections at CIP have been requested for several research projects and have been successfully used to screen for novel sources of late blight, crop efficiencies via remote sensing, and to elucidate and reconstruct the origins of potato (Gutaker et al., 2019; Silva-Diaz et al., 2020; Perez et al., 2022). These subsets are available upon request to the CIP genebank.

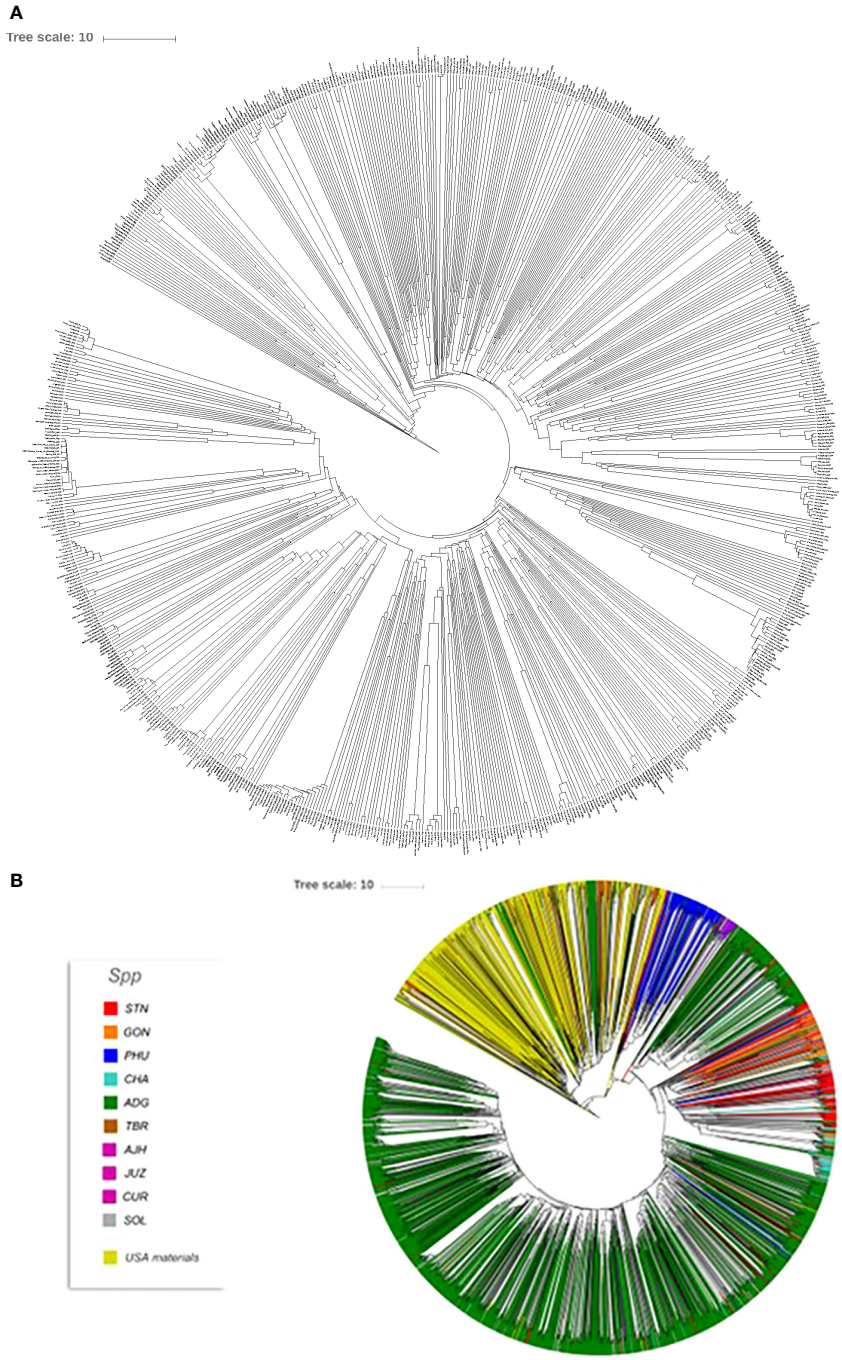


FIGURE 8
(A) Dendrogram of 745 accessions from the Seed Savers collection and **(B)** phylogeny of the Seed Savers (yellow branches) relative to the 3860 cultivated accessions from the CIP collection used in this study.

Conclusions

Genotyping germplasm collections is a major undertaking especially for large collections; however, the data produced from these projects are extremely valuable for the overall management of the germplasm and responding to potato researchers using the collection, especially breeders. Moreover, genotyping data in clonal collections can provide a framework and knowledge of the overall genetic integrity of accessions in a collection to be maintained by setting a baseline with fingerprinting data on each accession, which can be checked over the years to ensure no mistakes in manipulation and handling occur and to quickly rectify them when they do occur. Genotyping clonal accessions when they first are introduced into a collection can help facilitate integrity through monitoring of their fingerprints over time. When paired samples exist of accessions like in this study, genetic identity can be further checked by genotyping multiple samples of the same clonal accession to ensure that genetic identity is being maintained.

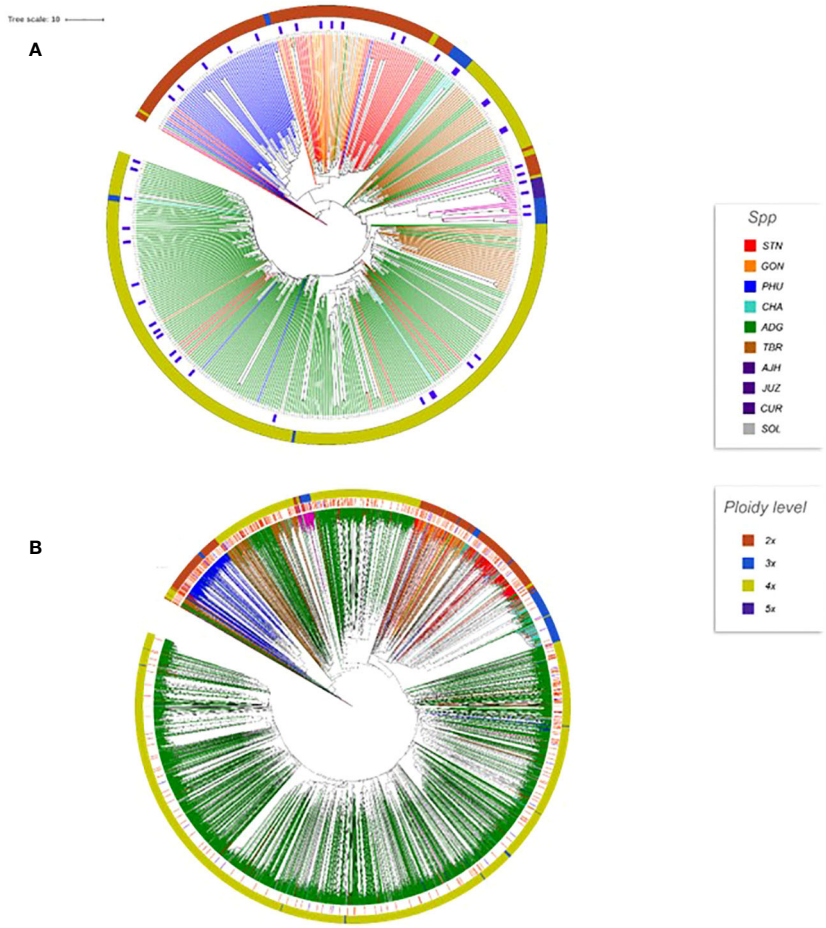


FIGURE 9
(A) Dendrogram of the 451 core and mini core accessions. The outer ring is the ploidy level. The branches are colored based on species designation. The blue rectangular blocks designate the selected mini core accessions (45 accs). **(B)** Dendrogram of the 3860 accessions with the core accessions marked as red lines (inside the ploidy circle) and the mini core marked with blue lines (inside the ploidy circle).

Further, genotyping allowed a deeper understanding into the identity, genetic diversity, redundancy, relatedness, hybrid origin, and introgression of a germplasm collection that is difficult to grasp when just evaluating a subsection of a germplasm collection. The genotyping data provided new valuable information on inter- and intraspecific relationships and the hybrid origin of cultivated potato species and specific potato accessions. In this study, our data demonstrated that the level of introgression and hybridization of cultivated potato was high especially within ADG and CUR accessions. However, most all species and accessions contained quite a bit of admixture and demonstrated that potato species easily exchange alleles. Many individual genotypes were derived of alleles from two or more genetic clusters. Several potato species appeared to be hybrids such as GON and JUZ while other species had multiple introgressions (CUR, ADG). Further, this data supports both commonly used systems (Hawkes and Spooner) of potato taxonomy depending on the species in question and the analysis considered. For example, the lumping of species (Spooner) such as CHA and STN were supported in this data by a single lineage in STRUCTURE; however, PHU which Spooner lumped together and eliminated as a species, appeared to be a unique lineage which supports Hawkes's taxonomy. The level of admixture for ADG would favor Spooner where a defined species level would be hard to ascertain. The data overall demonstrated the complexity and extensive hybridization that has occurred in the evolution of cultivated potato species and further how some species may be over described especially considering the amount of gene flow among the species. The STRUCTURE data also provided evidence for the putative species classification of accessions not previously classified that were new acquisitions to the genebank or material that was never able to be classified which is extremely valuable since potato taxonomists are a limiting factor. Further, genotyping of the CIP collection allowed for the first-time comparison of material conserved at other locations. A subsection of the Seed Savers potato collection was genotyped and compared to the 3860 potato samples at CIP and no overlap, or genetic redundancies were found. Future work may include working with other collections to genotype their material and rationalize collections. The SNP data and curator knowledge of the collection was further utilized to select a core and mini core collection to represent the majority of the genetic diversity in the genebank collection which can greatly help researchers measure traits that are

expensive or unreasonable to measure collection wide. This is the first time a core or mini core collection has been developed from the potato genebank at CIP. The SNP data also allowed a confirmation of ploidy levels of the accessions since many were routinely classified based on species alone. Overall, genotyping has led to many insights on the diversity and population structure of one of the world's largest cultivated potato germplasm collections.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://data.cipotato.org/dataset.xhtml?persistentId=doi:10.21223/LBCFCF.

Author contributions

NA: Investigation, Methodology, Project administration, Resources, Supervision, Visualization, Writing – original draft, Writing – review & editing. OC: Data curation, Formal analysis, Software, Validation, Visualization, Writing – review & editing. JS-T: Data curation, Formal analysis, Methodology, Software, Validation, Visualization, Writing – review & editing. RG: Writing – review & editing, Data curation, Methodology, Resources, Supervision, Validation. AP: Data curation, Methodology, Resources, Supervision, Writing – review & editing. RV: Methodology, Resources, Supervision, Writing – review & editing. MD: Data curation, Methodology, Resources, Visualization, Writing – review & editing. CM: Data curation, Resources, Supervision, Writing – review & editing. VA: Resources, Writing – review & editing. NM-C: Data curation, Formal analysis, Methodology, Resources, Software, Supervision, Writing – review & editing. PK: Investigation, Supervision, Writing – review & editing. JC: Data curation, Formal analysis, Software, Visualization, Writing – review & editing. DD: Data curation, Methodology, Resources, Writing – review & editing. DE: Conceptualization, Funding acquisition, Investigation, Project administration, Supervision, Writing – original draft.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The authors gratefully acknowledge the support of the CGIAR Genebank Platform and express thanks for the financial support by GIZ on behalf of the Federal Ministry of Economic Cooperation and Development, Germany.

Acknowledgments

We dedicate this paper to the memory of a great potato researcher, Eng, Rene Gomez, without whose dedication, this work would never have been possible.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer CR declared a past co-authorship with the author NA to the handling editor.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1341788/full#supplementary-material

References

Anastasio, A. E., Platt, A., Horton, M., Grotewold, E., Scholl, R., Borevitz, J. O., et al. (2011). Source verification of mis-identified Arabidopsis thaliana accessions. *Plant J.* 67, 554–566. doi: 10.1111/j.1365-313X.2011.04606.x

Anglin, N. L., Robles, R., Rossel, G., Alagon, R., Panta, A., Jarret, R. L., et al. (2021). Genetic identity, diversity, and population structure of CIP's sweetpotato (I. batatas) germplasm collection. *Front. Plant Sci.* 12, 660012. doi: 10.3389/fpls.2021.660012

Bergelson, J., Buckler, E. S., Ecker, J. R., Nordborg, M., and Weigel, D. (2016). A proposal regarding best practices for validating the identity of genetic stocks and the effects of genetic variants. *Plant Cell* 28, 606–609. doi: 10.1105/tpc.15.00502

Bertioli, D. J., Ozias-Akins, P., Chu, Y., Dantas, K. M., Santos, S. P., Gouvea, E., et al. (2014). The use of SNP markers for linkage mapping in diploid and tetraploid peanuts. *G3.: Genomes/Genomes/Genetics.* 4, 89–96. doi: 10.1534/g3.113.007617

Brown, A. H. D. (1989). "The case for core collections," in *The Use of Plant Genetic ResourcesA*. Eds. H. D. Brown, O. H. Frankel, D. R. Marshall and J. T. Williams (Cambridge University Press, Cambridge, UK), 136–156.

Caruana, B. M., Pembleton, L. W., Constable, F., Rodoni, B., Slater, A. T., and Cogan, N. O. I. (2019). Validation of genotyping by sequencing using transcriptomics for diversity and application of genomic selection in tetraploid potato. *Front. Plant Sci.* 10, 670. doi: 10.3389/fpls.2019.00670

Clevenger, J. P., and Ozias-Akins, P. (2015). SWEEP: A tool for filtering high quality SNPs in polyploid crops. *Genes. Genomes. Genet.* 5, 1797–1803. doi: 10.1534/g3.115.019703

De Beukelaer, H., Davenport, G. F., and Fack, V. (2018). Core Hunter 3: flexible core subset selection. *BMC Bioinf.* 19, 203. doi: 10.1186/s12859-018-2209-z

Dent, E. A., and vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* 4, 359–361. doi: 10.1007/s12686-011-9548-7

Doležel, J., Binarova, P., and Lucretti, S. (1989). Analysis of nuclear DNA content in plant cells by flow cytometry. *Biol. Plantarum.* 31, 113–120. doi: 10.1007/BF02907241

Douches, D., Hirsch, C. N., Manrique-Carpintero, N. C., Massa, A. N., Coombs, J., Hardigan, M., et al. (2014). The contribution of the Solanaceae coordinated agricultural project to potato breeding. *Potato. Res.* 57, 215–224. doi: 10.1007/s11540-014-9267-z

Doyle, J. J., and Doyle, J. L. (1990). Isolation of plant DNA from fresh tissue. *BRL Focus* 12, 13–15.

Ellis, D., Chavez, O., Coombs, J., Soto, J., Gomez, R., Douches, D., et al. (2018). Genetic identity in genebanks: application of the SolCap 12K SNP array in fingerprinting the global in *trust* potato collection. *Genome* 61, 523–537. doi: 10.1139/gen-2017-0201

Falush, D., Stephens, M., and Pritchard, J. K. (2003). Inference of population structure: extensions to linked loci and correlated allele frequencies. *Genetics* 164, 1567–1587. doi: 10.1093/genetics/164.4.1567

Felcher, K. J., Coombs, J. J., Massa, A. N., Hansey, C. N., Hamilton, J. P., Veilleux, R. E., et al. (2012). Integration of two diploid potato linkage maps with the potato genome sequence. *PloS One* 12, e36347. doi: 10.1371/journal.pone.0036347

Frankel, O. H., and Brown, A. H. D. (1984). "Plant genetic resources today: a critical appraisal," in *Crop Genetic Resources: Conservation & Evaluation*. Eds. J. H. W. Holden and J. T. Williams (George Alien & Unwin Ltd., London, U.K), 249–257.

Galbraith, D. W., Harkins, K. R., Maddox, J. M., Ayres, N. M., Sharma, D. P., and Firoozabady, E. (1983). Rapid flow cytometric analysis of the cell cycle in intact plant tissues. *Science* 220, 1049–1051. doi: 10.1126/science.220.4601.1049

Geibel, J., Reimer, C., Weigend, S., Weigend, A., Pook, T., and Simianer, H. (2021). How array design creates SNP ascertainment bias. *PloS One* 16 (3), e0245178. doi: 10.1371/journal.pone.0245178

Gemenet, D. C., Kitavi, M. N., David, M., Ndege, D., Ssali, R. T., Swanckaert, J., et al. (2020). Development of diagnostic SNP markers for quality assurance and control in sweetpotato [Ipomoea batatas (L.) Lam.] breeding programs. *PloS One* 15 (4), e0232173. doi: 10.1371/journal.pone.0232173

Ghislain, M., Andrade, D., Rodriguez, F., Hijmans, R. J., and Spooner, D. M. (2006). Genetic analysis of the cultivated potato Solanum tuberosum L. Phureja group using RAPDs and nuclear SSRs. *TAG* 113, 1515–1527. doi: 10.1007/s00122-006-0399-7

Girma, G., Korie, S., Dumet, D., and Franco, J. (2012). Improvement of accessions distinctiveness as an added value to the global worth of the yam (*Dioscorea* spp) genebank. *Int. J. Conserv. Sci.* 3, 199–206. Available at: https://cgspace.cgiar.org/items/028fee49-10c2-4576-a7c5-9b80d466ad84.

Gutaker, R. M., Weib, C. L., Ellis, D., Anglin, N. L., Knapp, S., Fernandez-Alonso, J. L., et al. (2019). The origins and adaptation of European potatoes reconstructed from historical genomes. *Nat. Ecol. Evol.* 3, 1093–1101. doi: 10.1038/s41559-019-0921-3

Hamilton, J. P., Hansey, C. N., Whitty, B. R., Stoffel, K., Massa, A. N., Deynze, A. V., et al. (2011). Single nucleotide polymorphism discovery in elite North American potato germplasm. *BMC Genomics* 12, 302. doi: 10.1186/1471-2164-12-302

Hardigan, M. A., Bamberg, J., Buell, C. R., and Douches, D. S. (2015).). Taxonomy and genetic differentiation among wild and cultivated germplasm of Solanum sect. Petota. *Plant Genome*, 8 (1), eplantgenome2014.06.0025. doi: 10.3835/plantgenome2014.06.0025

Hawkes, J., G. (1990). *The potato evolution, biodiversity and genetic resources* (London UK: Belhaven Press.), 1–259.

Hirsch, C. N., Hirsch, C. D., Felcher, K., Coombs, J., Zarka, D., Deynze, A. V., et al. (2013). Retrospective view of North American potato (Solanum tuberosum L.) breeding in the 20th and 21st centuries. *G3:Genes. Genomes Genet* 3 (6), 1003–13. doi: 10.1534/g3.113.005595

Hoopes, G., Meng, X., Hamilton, J. P., Achakkagari, S. R., Guesdes, F., Bolger, M. E., et al. (2022). Phased, chromosome-scale genome assemblies of tetraploid potato reveal a complex genome, transcriptome, and proteome landscape that underpin phenotypic diversity. *Mol. Plant* 15, 520–536. doi: 10.1016/j.molp.2022.01.003

Huaman, Z., and Spooner, D. M. (2002). Reclassification of landrace populations of cultivated potatoes (Solanum sect. Petota). *Am. J. Bot.* 89, 947–965. doi: 10.3732/ajb.89.6.947

Hughes, P., Marshall, D., Reid, Y., Parkes, H., and Gelber, C. (2007). The costs of using unauthenticated over passaged cell lines: how much more data do we need? *Biotechniques* 43, 575–586. doi: 10.2144/000112598

Letunic, I., and Bork, P. (2016). Interactive Tree of Life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* 44 (W1), W242–5. doi: 10.1093/nar/gkw290

Logan-Young, C. J., Yu, J. Z., Verma, S. K., Percy, R. G., and Pepper, A. E. (2015). SNP discovery in complex allotetraploid genomes (*Gossypium* spp., Malvaceae) using Genotyping by Sequencing. *Appl. Plant Sci.* 3, 1400077. doi: 10.3732/apps.1400077

Mammadov, J., Aggarwal, R., Buyyarapu, R., and Kumpatla, S. (2012). SNP markers and their impact on plant breeding. *Int. J. Plant Genomics* 2012, 728398. doi: 10.1155/2012/728398

Manrique-Carpintero, N. C., Berdugo-Cely, J. A., Cerón-Souza, I., Lasso-Paredes, Z., Reyes-Herrera, P. H., and Yockteng, R. (2023). Defining a diverse core collection of the Colombian Central Collection of potatoes: a tool to advance research and breeding. *Front. Plant Sci.* 14, 1046400. doi: 10.3389/fpls.2023.1046400

Manrique-Carpintero, N. C., Tokuhisa, J. G., Ginzberg, I., Holliday, J. A., and Veilleux, R. E. (2013). Sequence diversity in coding regions of candidate genes in the glycoalkaloid biosynthetic pathway of wild potato species. *G3.: Genes. Genomes Genet.* 3, 1467–1479. doi: 10.1534/g3.113.007146

Manrique-Carpintero, N. C., Tokuhisa, J. G., Ginzberg, I., and Veilleux, R. E. (2014). Allelic variation in genes contributing to glycoalkaloid biosynthesis in a diploid interspecific population of potato. *Theor. Appl. Genet.* 127, 391–405. doi: 10.1007/s00122-013-2226-2

Ovchinnikova, A., Krylova, E., Gavrilenko, T., Smekalova, T., Zhuk, M., Knapp, S., et al. (2011). Taxonomy of cultivated poatoes (Solanum section Petota: Solanaceae). *Bot. J. Linn. Soc.* 165, 107–155. doi: 10.1111/boj.2011.165.issue-2

Perazzo, G., Panta, A., Rodriguez, F., Gomez, R., Toledo, J., Huaman, Z., et al. (1999–2000a). *Clonal true-to-type verification of potato accessions retrieved from in vitro conservation and cryopreservation. CIP Program Report.* 175–183. Available at: https://www.researchgate.net/publication/260986444.

Perez, W., Alarcon, L., Rojas, T., Correa, Y., Andrade-Piedra, J., Anglin, N. L., et al. (2022). Screening of South American potato landraces and potato wild relatives for new sources of late blight resistance. *Plant Dis* 106 (7), 1845–1856. doi: 10.1094/PDIS-07-21-1582-RE

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959. doi: 10.1093/genetics/155.2.945

R Core Team (2015). R: A language and environment for statistical computing (Vienna, Austria: R Foundation for Statistical Computing). Available at: https://www.R-project.org/.

Silva-Diaz, C., Ramirez, D. A., Rinza, J., Ninanya, J., Loayza, H., Gomez, R., et al. (2020). Radiation interception, conversion and partitioning efficiency in potato landraces: how far are we from the optimum? *Plants* 9, 787. doi: 10.3390/plants9060787

Simko, I., Haynes, K. G., and Jones, R. W. (2006). Assessment of linkage disequilibrium in potato genome with single nucleotide polymorphism markers. *Genetics* 173, 2237–2245. doi: 10.1534/genetics.106.060905

Spooner, D. M., Nuñez, J., Trujillo, G., Herrera, M. D. R., Guzmán, F., and Ghislain, M. (2007). Extensive simple sequence repeat genotyping of potato landraces supports a major reevaluation of their gene pool structure and classification. *PNAS. U.S.A.* 104, 19398–19403. doi: 10.1073/pnas.0709796104

Stich, B., Urbany, C., Hoffmann, P., and Gebhardt, C. (2013). Population structure and linkage disequilibrium in diploid and tetraploid potato revealed by genome wide high-density genotyping using the SolCap SNP array. *Plant Breed.* 132, 718–724. doi: 10.1111/pbr.12102

Stift, M., Kolar, F., and Meirmans, P. G. (2019). STRUCTURE is more robust than other clustering methods in simulated mixed-ploidy populations. *Heredity* 123, 429–441. doi: 10.1038/s41437-019-0247-6

Upadhyaya, H., and Ortiz, R. (2001). A mini core subset for capturing diversity and promoting utilization of chickpea genetic resources in crop improvement. *Theor. Appl. Genet.* 102, 1292–1298. doi: 10.1007/s00122-001-0556-y

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 18 July 2024
DOI 10.3389/fpls.2024.1429279

Check for updates

OPEN ACCESS

EDITED BY
Daniel Pinero,
National Autonomous University of Mexico, Mexico

REVIEWED BY
Diego Hojsgaard,
Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Germany
Isabel Marques,
Laboratorio TERRA. Universidade de Lisboa, Portugal

*CORRESPONDENCE
Laura M. Shannon
lmshannon@umn.edu

RECEIVED 07 May 2024
ACCEPTED 27 June 2024
PUBLISHED 18 July 2024

CITATION
Tuttle HK, Del Rio AH, Bamberg JB and Shannon LM (2024) Potato soup: analysis of cultivated potato gene bank populations reveals high diversity and little structure. *Front. Plant Sci.* 15:1429279. doi: 10.3389/fpls.2024.1429279

COPYRIGHT
© 2024 Tuttle, Del Rio, Bamberg and Shannon. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Potato soup: analysis of cultivated potato gene bank populations reveals high diversity and little structure

Heather K. Tuttle[1], Alfonso H. Del Rio[2], John B. Bamberg[2] and Laura M. Shannon[1]*

[1]Department of Horticultural Science, University of Minnesota, St. Paul, MN, United States, [2]U.S. Department of Agriculture (USDA)/Agricultural Research Service, Potato Genebank, Sturgeon Bay, WI, United States

Cultivated potatoes are incredibly diverse, ranging from diploid to pentaploid and encompass four different species. They are adapted to disparate environments and conditions and carry unique alleles for resistance to pests and pathogens. Describing how diversity is partitioned within and among these populations is essential to understanding the potato genome and effectively utilizing landraces in breeding. This task is complicated by the difficulty of making comparisons across cytotypes and extensive admixture within section petota. We genotyped 730 accessions from the US Potato genebank including wild diploids and cultivated diploids and tetraploids using Genotype-by-sequencing. This data set allowed us to interrogate population structure and diversity as well as generate core subsets which will support breeders in efficiently screening genebank material for biotic and abiotic stress resistance alleles. We found that even controlling for ploidy, tetraploid material exhibited higher observed and expected heterozygosity than diploid accessions. In particular group *chilotanum* material was the most heterozygous and the only taxa not to exhibit any inbreeding. This may in part be because group *chilotanum* has a history of introgression not just from wild species, but landraces as well. All group *chilotanum*, exhibits introgression from group *andigenum* except clones from Southern South America near its origin, where the two groups are not highly differentiated. Moving north, we do not observe evidence for the same level of admixture back into group *andigenum*. This suggests that extensive history of admixture is a particular characteristic of *chilotanum*.

KEYWORDS

core subsets, autopolyploidy, ploidy detection, admixture, heterozygosity, multiploidy populations

1 Introduction

Potato is the third most important global food crop (FAO, 2019), because it is widely adapted and has a high nutrient to price ratio (Drewnowski and Rehm, 2013). It is unequivocally one of the most complete foods, containing significant dietary fiber, protein, vitamin C, B6, potassium, magnesium, iron and phytonutrients (King and Slavin, 2013; McGill et al., 2013; Brown et al., 1993; Liu, 2013; Beals, 2019; Navarre et al., 2019). Furthermore, potato grows well in harsh environments on six continents. It produces more calories and protein per unit land or water than any other staple crop (Renault and Wallender, 2000). Both India and China have undertaken national efforts to increase potato production as a way to support growing populations (FAO, 2019). As global populations expand and the climate changes, our reliance on potatoes will increase and breeders will need to develop varieties that can adapt to new environments.

Despite the agricultural and nutritional importance of potato, adoption of new potato varieties is slow and the most commonly grown potato variety in the US was developed over a century ago (Douches et al., 1996). Breeders need tools to facilitate the development of new varieties with increased biotic and abiotic stress resistance. The most effective tool for introducing new resistance alleles into potato is introgression from landraces and crop wild relatives (Jansky et al., 2013).

The USDA potato genebank in Sturgeon Bay WI is the most accessible repository of this germplasm for US breeders. Screening the genebank collection has identified new resistance alleles for pests and pathogens such as late blight (Karki et al., 2021), zebra chip (Mora et al., 2022), and Colorado potato beetle (Jansky et al., 2009). However, the genebank contains almost 2,000 accessions of cultivated potato and screening all of them is prohibitively difficult. Therefore, the first step to identifying valuable alleles within the collection is assembling the right screening panel. Building core collections within the genebank facilitates this process.

Core collections are an essential tool for empowering efficient screening of genebank material for use in breeding (Haupt and Schmid, 2020; Sokolkova et al., 2020; Phogat et al., 2021; Mufumbo et al., 2023; Santos et al., 2023; Shigita et al., 2023). The goal is to maximize the number of alleles evaluated while minimizing the number of individuals that must be screened. Such collections have been developed for a wide variety of crops using both morphological (Phogat et al., 2021; Santos et al., 2023) and genetic (Haupt and Schmid, 2020; Mufumbo et al., 2023; Shigita et al., 2023) data. While core subsets based on morphology and traits emphasized by breeders have advantages for use in prediction (Corak et al., 2019), they rely on the curators to foresee the total set of potentially relevant traits. New pathogens are continually arising (Duellman et al., 2020; Willbur et al., 2023) and mutating (Tran et al., 2022). Core subsets based on neutral markers maximize total diversity represented, even if the functionality of that diversity is not immediately obvious. Furthermore, genotyping gene bank collections with neutral markers provides insight into diversity and relatedness of the collected species.

The US potato genebank houses four core collections of wild potato species, *S. jamesii* (Bamberg et al., 2016), *S. fendleri* (Bamberg et al., 2016), *S. microdontum* (Bamberg and del Rio, 2014), and *S. demissum* (del Rio and Bamberg, 2020), and one core collection of diploid cultivated *S. chilotanum group phureja* (del Rio and Bamberg, 2021). Each of these were developed using AFLPs and validated with phenotyping data. However, individual taxa only represent a fraction of potato as a whole. Section petota includes four cultivated species and 107 wild relatives, ranging from diploid to hexaploid (Spooner et al., 2007, 2014).

Understanding how diversity is partitioned within and among these cultivated taxa facilitates their use in breeding. The history of potato, is complex. Cultivated potatoes were domesticated at least once from *S. candolleanum* (2n=2x=24) in the Andes near Southern Peru and Northern Bolivia more than 10,000 years ago (Camire et al, 2009). Potatoes were later cultivated in the highland equatorial conditions of Colombia and Venezuela as well as long day conditions in Chile and southern Argentina. Autopolyploidization of early landrace species, *S. tuberosum* groups *stenotomum* and *phureja* gave rise to *S. tuberosum* group *andigenum* (2n=4x=48) (Grun, 1990; Hawkes, 1990; Sukhotu and Hosaka, 2006). Later migration to coastal Chile led to the long-day adapted *S. tuberosum* group *chilotanum* (2n=4x=48) which is purported to have contributed most of the genetic background of *S. tuberosum* cultivars outside South America (King and Slavin, 2013). Understanding how diversity is partitioned within and among these taxa is essential to understanding potato's history and guiding its future.

By all estimates, potato is highly heterozygous (Hardigan et al., 2017; Hoopes et al., 2022). However, descriptions of comparative levels of diversity between populations differ, in part because disparate ploidy levels complicate comparison. While tetraploid cultivated potatoes have higher heterozygosity and nucleotide diversity than diploid wild potatoes by some calculations (Hardigan et al., 2017), that relationship is reversed if SNPs are called more stringently (Huang et al., 2018) or if diploid cultivated potatoes are considered instead of tetraploid ones (Li et al., 2018). Across species, polyploids are generally more heterozygous than diploids due to the increase in number of alleles per locus (Meirmans et al., 2018). Therefore, fair comparison across cytotype requires correction for ploidy. When corrections are not made, the diversity in wild and cultivated diploid potato is under estimated (Bamberg and del Rio, 2020).

Another factor which complicates our understanding of how diversity is partitioned in potato, is the porous boundaries within section petota. While membership in petota is stable, relationships between taxa within the section are not (Gagnon et al., 2022). Taxonomy within section petota is made difficult by morphological similarity, phenotypic plasticity, allele loss, a mixture of sexual and asexual reproduction, recent species divergence, polyploidy, introgression, and multiple hybrid origin (Spooner and van den Berg, 1992; Spooner, 2009; Ames and Spooner, 2010; Cai et al., 2012; Huang et al, 2019; Zhou et al., 2020). There is extensive evidence for both hybridization in wild species (Hawkes, 1969; Ugent, 1970; Hawkes, 1990; Spooner et al., 2007; Tang et al., 2022) and for wild species introgression into cultivated US and European potato continually since domestication (Hardigan et al., 2017; Hoopes et al., 2022; Meng et al., 2022). The genebank collection

provides the opportunity to investigate the extent to which the landraces exhibit the same history of admixture as wild and US/European improved potatoes.

In this study, 730 accessions of diploid wild and diploid and tetraploid cultivated individuals were genotyped using a genotyping-by-sequencing (GBS) approach (Elshire et al., 2011). We interrogated this data set to address three questions. (1) How is diversity, both individual heterozygosity and population level allelic diversity, structured within cultivated potato? (2) Does the pattern of admixture observed in wild and commercial potato also describe the land race taxa? and (3) What is the ideal composition of core subsets for screening this portion of the genebank collection?

2 Materials and methods

2.1 Plant materials and sequencing

The US potato genebank in Sturgeon Bay, Wisconsin holds 1,445 accessions of cultivated potato. This assemblage represents years of collections in fields and markets some from planned collecting trips and some incidental, as well as contributions from researchers. We genotyped a subset of the collection, chosen with an eye to maximizing phenotypic diversity, containing 730 diploid, tetraploid and pentaploid accessions (Supplementary Table S1), using genotyping-by-sequencing (GBS) (Elshire et al., 2011) with the EcoT22 enzyme and phased adapters on the Illumina HiSeq platform at the University of Minnesota Genomics Center. According to passport data these 730 accessions consisted of 72 diploids including: *S. juzepczukii* (6), *S. berthaultii* (1), *S. brevicaule* (1), *S. ajanhuiri* (3), *S. phureja* (27), *S. stenotomum* (11), *S. boliviense* (20) and *S. tuberosum* group *chilotanum* (3). The 641 tetraploid accessions included: group *andigenum* (301), group *chilotanum* (333), *S. juzepczukii* (4), *S. phureja* (1) and *S. ajanhuiri* (2). Group *andigenum* accessions were primarily South American, coming from Peru, Bolivia, Colombia, Ecuador, Chile, Argentina, Uruguay, Venezuela and Brazil. Group *chilotanum* included accessions from North America, South America, Central America, Europe, Asia, Russia and Africa. Pentaploid accessions consisted of 14 *S. curtilobum* with two samples of group *chilotanum*. Triploids were not included in the genotyping panel. There was one accession for which the passport data did not include: taxa, clone name, or ploidy.

Some individuals, in particular those from group *chilotanum* in North America and Europe are named cultivars. These individuals are highly selected. The land race individuals including *S. juzepczukii*, *S. ajanhuiri*, *S. phureja*, *S. stenotomum*, and many of the *S. tuberousum* individuals from South America are random seedlings selected from genebank populations maintained as true seed. The wild individuals, most notably *S. boliviense*, are from a single wild population. Since we are only comparing genome wide patterns of diversity, rather than functional alleles, and we expect those to be similar in selected clones (ie cultivars) and individuals arising in crosses made from selections (ie the landrace populations), we can make comparisons across cultivated populations. However, since the *S. boliviense* is from a single population it is included only when an outgroup is needed for analysis. *S. boliviense* is removed from other analyses to avoid making unfair comparisons between representatives of a species as a whole and representatives of a single population (Bamberg and del Rio, 2020). Similarly, all taxa with only one representative were removed.

2.2 Sequence read alignment and variant calling

Reads were checked for quality with FASTQC 0.11.5 (Andrews, 2010) and adapters were removed with Cutadapt version 1.18 (Martin, 2011). Each sample was aligned to the Phureja DM v4.04 reference genome (Hardigan et al., 2016) with Bowtie 2 version 2.2.4 (Langmead and Salzberg, 2012) using Phred +33 encoding. We used Samtools version 1.9 (Danecek et al., 2021) to create bam files and report the number of mapped reads. Samples with less than 150,000 mapped reads were removed (Supplementary Table S2). Since we wanted to make comparisons of diversity within and among ploidy levels, we created three separate panels for our analyses; diploid, tetraploid and a third, combined panel (Table 1). SNP genotypes were called twice, once before ploidy correction and again after ploidy correction. All steps prior to ploidy correction were based on the combined panel. Once ploidy estimation by computational methods either confirmed or rejected the original ploidy calls, samples were removed or added to the panels and genotypes for single ploidy panels were called once again. BCFtools version 1.10.2 and samtools mpileup version (Danecek et al., 2021) were used to generate genotype likelihoods, call SNPs using the multiallelic caller and remove indels. Filtering thresholds appropriate for each panel were used (Table 2). VCFtools version 0.1.16 (Danecek et al., 2011) was used to remove unmapped reads, filter SNPs and extract biallelic SNPs at sites with no more than 40% missing data.

TABLE 1 Datasets used for analysis.

Dataset	Number of SNPs	Number of individuals	Application
Diploid, Naive	5148	56	All analyses
Tetraploid, Naive	4720	497(183)	Rho, Structure, PCoA, D statistics
Tetraploid, Simulated	4720	497	Heterozygosity, Gis
Combined, Naive	7810	553 (235)	Fst, Structure, PCoA, D statistics
Combined, simulated	7810	553	Heterozygosity, Gis

The diploid panel did not have sufficient individuals to create a dataset using simulated population structure. For analyses that strived to estimate population structure such as STRUCTURE, PCoA and Fst/Rho we used naive datasets which contained no more than 20 individuals from each population. Total number of individuals in these analyses are in parentheses.

10.3389/fpls.2024.1429279

Over 60X coverage is required to definitively distinguish between heterozygous classes in tetraploids (Uitdewilligen et al., 2015). Few markers in the tetraploid panel met this requirement. Therefore, we used PolyRAD (Clark et al., 2019) to re-call the most probable genotypes in panels containing tetraploid individuals. VCFs for the tetraploid and combined panel were read into the R (v4.2.1; R Core Team, 2021) environment using VCF2RADdata, only retaining variants that were called in at least 109 individuals (~20% of the population) and where the minor allele was in a minimum of 2 individuals. Variants in the diploid panel were only retained if they were present in at least 10 individuals (~20% of the population) with the minor allele present in at least 2 individuals. To filter loci in the combined panel, we separated individuals by ploidy and removed markers with a Hind/He less than 0.5 and greater than 0.75 for diploids and tetraploids, respectively. Genotypes were re-called by simulating population structure with individuals assigned to groups based on PCA. We created an additional dataset for each of the three panels that used a naïve model to call genotypes so as not to bias our population structure estimates for a total of six data sets (Table 1). After re-calling genotypes, loci with more than 10% missing data were removed and remaining missing data was filled in based on overall allele frequencies. For the combined panel and the tetraploid panel, dosage of polyploids was restored in Genodive version 3.0 (Meirmans, 2020) with resampled alleles.

2.3 Ploidy estimation

For multiple accessions in the data set, the passport data indicated a ploidy that conflicted with the known ploidy for the assigned taxa. The ploidy included in the passport data was determined using root tip squashes upon the addition of accessions to the genebank (Ordoñez et al., 2017). In order to confirm passport data, we estimated ploidy for each accession using two methods. First, histograms of allele frequencies across all markers were plotted for each individual to observe the number of peaks (Ellis et al., 2018). There are n+1 genotypic classes at ploidy n, therefore the ploidy level is equal to the number of histogram peaks minus one. We also ran GBS2ploidy with three settings (diploid vs. tetraploid, diploid vs. triploid vs. tetraploid, and diploid vs. triploid vs. tetraploid vs. pentaploid) (Gompert and Mock, 2017). Posterior estimates of allelic proportions were estimated with mcmc.nchain =2, mcmc.step = 10,000, and mcmc.burnin = 1000 and mcmc.thin = 2. After estimating allelic proportions, we performed ploidy estimations on the three datasets. We retained samples for which both methods of ploidy estimation confirmed the passport data.

TABLE 2 Filtering thresholds for each panel.

Filter	Diploid	Tetraploid	Combined
MQ	<20	<20	<20
GQ	<30	<60	<60
QUAL	<30	<40	<30
MAF	<0.02	<0.02	<0.02
DP	<5	<7	<5

Alignments with values below the threshold were removed. Tetraploids had better genotype quality and site quality over sequenced diploid individuals but did not have sufficient read depth to correctly discern genotypes. Sequencing error rate is 0.02 for all the data used.

2.4 Genetic diversity

With the goal of elucidating how diversity is distributed, we divided the panels into populations based on geography, species, and ploidy level. Countries were grouped together in an attempt to form reasonably sized populations. *S. phureja*, *S. boliviense*, and *S. stenotomum*, were each treated as a single taxa-based population. group *chilotanum* was further divided into geographic populations: Europe, US/Canada, Mexico/Guatemala, Brazil/Ecuador/Colombia, Peru, Bolivia/Argentina/Uruguay, and Chile. Similarly, group *andigenum* was subdivided by geography: Colombia/Venezuela, Ecuador, Peru, Bolivia, and Argentina/Chile. Location, reflects the best information we have based on collection site or breeder location, but may not reflect original geography of the clone. Species level diversity statistics such as expected heterozygosity (H_S), observed heterozygosity (H_O) and an inbreeding coefficient (G_{IS}) that is analogous to F_{IS}, were calculated in GenoDive (Meirmans, 2020) with the panels where genotypes were called taking population structure into account. All panels were bootstrapped ten times to obtain 95% confidence intervals.

2.5 Population structure

We quantified differentiation within and across ploidy, species, and geographic region using F_{ST} and rho (ρ) (Ronfort et al., 1998; Meirmans et al., 2018) calculated in GenoDive (Meirmans, 2020). We used ρ for the single ploidy panels and F_{ST} for the combined ploidy panel. Differentiation statistics are sensitive to sample size imbalance; therefore, we subsampled each population to retain no more than 20 individuals. Each population was resampled ten times and bootstrapped to determine 95% confidence intervals.

In order to visualize patterns of admixture between populations we used STRUCTURE (Pritchard et al., 2000). For all three panels, we used an admixture model that infers alpha with a 10,000 burn-in period and 10,000 MCMC replications. Lambda was set to 1.0. Priors used to parameterize the assumed probability models were set to default. This process was repeated for values of k between 1 and 6 for naive forms of the diploid, tetraploid and combined datasets. Because STRUCTURE does not handle multiple cytotypes, the naive combined panel was diploidized before analysis (Meirmans et al., 2018). We also used principal coordinate analysis (PCoA) within GenoDive (Meirmans, 2020). All results were plotted and visualized using the ggplot2 package version 3.4.1 (Wickham, 2011) in the R statistical environment version 4.2.1 (R Core Team, 2021).

Formal tests of admixture between individual populations were performed using Dsuite Dtrios (Malinsky et al., 2021) to calculate Patterson's D (ABBA/BABA) statistics. Using all individuals in the naive combined panel, we first looked for possible admixture between

each subpopulation of *S. tuberosum* group chilotanum and *S. tuberosum* group *andigenum* (tetraploid group *andigenum*, *S. phureja*, and *S. stenotomum*). We then examined admixture between each subpopulation of *S. tuberosum* group *andigenum* and *S. tuberosum* group *chilotanum*. *S. boliviense* was used as an outgroup and a jackknife block size of 150 was used to account for linkage. P-values were corrected for multiple comparisons using false discovery rate (FDR). Statistics were visualized in the R environment.

2.6 Core subset selection

In the interest of identifying core subsets, we used two software programs with different core collection algorithms. CoreHunter 3 (De Beukelaer et al., 2018) and GenoCore (Jeong et al., 2017) were used to generate core subsets for each panel using an allele coverage metric (CV) which maximizes the proportion of the total alleles observed in a subset. For each panel, multiple core sizes were created and evaluated for their ability to capture marker diversity.

3 Results

3.1 Alignment and ploidy estimation

After aligning all 730 individuals to the V4.04 phureja DM reference genome, 29 individuals were removed due to insufficient mapped reads (Supplementary Table S2). Examination of allelic ratios and heterozygosity confirmed the passport ploidy data of 613 of the 701 remaining accessions (Supplementary Tables S3, S4). Of these 613 accessions, there were 553 tetraploids including group *chilotanum* (303) and group *andigenum* (247). The final data set contained 54 diploid individuals including *S. stenotomum* (9), *S. boliviense* (20), and *S. phureja* (25) All putative pentaploid individuals were removed because we were unable to confirm the ploidy computationally.

3.2 Genotype calling and variant filtration

Genotypes were called separately for each panel, using methods and filtering thresholds applicable to the ploidy and panel size (Table 2). After initial filtering, we identified 7,296 SNPs in the diploid panel with an average of 26% missingness per site and 6.78x average read depth. Subsequent filtering in polyRAD and GenoDive left the diploid panel with 5,135 SNPs. The tetraploid panel contained 5,742 SNPs with an average of 26% missingness per site, 8.03x average site read depth and after additional filtering 4,720 SNPs were retained. The combined panel contained 9,668 SNPs after initial calling and filtering of genotypes with an average of 22% site missingness and 6.8x average read depth. Additional filtering left the combined panel with 7,851 SNPs. The Hind/He filter in PolyRad is intended to remove paralogs and none of the fixed heterozygous loci observed in other potato GBS panels (Bamberg et al., 2021) remained after filtering. A total of 1,264 SNPs were shared between diploid, tetraploid and combined datasets (Supplementary Figure S1).

3.3 Estimates of genetic diversity and inbreeding

By most measures, tetraploids had higher heterozygosity (H_O = 0.190, H_S = 0.239) than diploids (H_O = 0.112, H_S = 0.140) even when using methods that account for ploidy (Figure 1A, Supplementary Table S3). Group *chilotanum* was the most heterozygous, followed by tetraploid group *andigenum*, and the cultivated diploids. Using SNP panels derived from only a single ploidy all levels of heterozygosity were higher, as expected, but the relative levels of heterozygosity between populations were

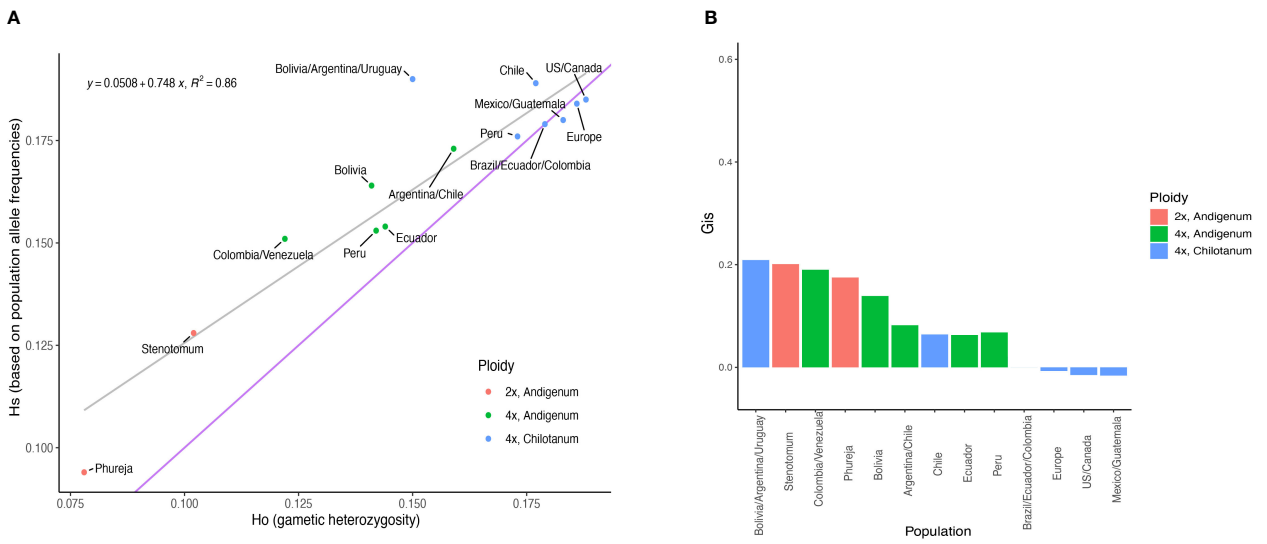


FIGURE 1
Diversity statistics from the combined panel. **(A)** Expected heterozygosity versus observed heterozygosity. The gray line is a fitted regression line with the equation given. The blue line indicates equal values for observed and expected heterozygosity. **(B)** Inbreeding coefficients for each population. Numeric values can be found in Supplementary Table S3.

consistent. While tetraploid populations exhibited few private alleles (1 in Europe and 1 in Peruvian andigenum), all diploid populations did, with *S. bolivense* having 1,929 private alleles, more than four times the next highest population (Supplementary Figure S2).

All populations except Northern group *chilotanum* exhibit some degree of inbreeding (Figure 1B). With the exception of the populations from Bolivia, Argentina, and Uruguay (G_{IS} = 0.209), group *chilotanum* populations exhibited the lowest in breeding overall (G_{IS} = 0 – 0.064) and diploid populations the highest (G_{IS} = 0.175 – 0.201).

3.4 Population structure and admixture

Based on F_{ST} there is little differentiation among tetraploids, with group *chilotanum* from the US and Canada and group *andigenum* from Peru exhibiting the most differentiation (F_{ST} = 0.109) (Figure 2, Supplementary Table S5). When using the tetraploid SNP panel to calculate ρ (rho), a more complicated pattern emerges (Figure 3, Supplementary Table 6). Northern group *chilotanum* (US, Canada, Mexico, Guatemala, and Europe) is distinct from group *andigenum* (ρ = 0.14 – 0.236) and the group *chilotanum* material from Bolivia, Uruguay, and Argentina (ρ = 0.144 – 0.181). Group *chilotanum* samples from Peru, Brazil, Ecuador, and Colombia were not differentiated from any of the tetraploid material. Subspecies pairs from similar geographic areas tended to have low ρ-values, specifically the Bolivian pair (ρ = 0.067), the Chilean pair (ρ = 0.093), and the Colombian pair (ρ = 0.066).

F_{ST} and ρ values are similar for the diploids (Figure 2, Supplementary Figure S3, Supplementary Tables S5, S7). In both cases *S. bolivense* is the most differentiated (ρ = 0.762 – 0.786) and *S. stenotomum* and *S. phureja* are the most similar (ρ = 0.117). *S. bolivense* is distinct from all tetraploids (F_{ST} = 0.386 – 0.483) and for the most part, so are *S. phureja* and *S. stenotomum* (F_{ST} = 0.574-0.607). *S. stenotomum* is highly similar to group *andigenum* (F_{ST} = 0.034 – 0.068) and group *chilotanum* from Peru, Brazil, Ecuador, and Colombia (F_{ST} = 0.066 – 0.074) and slightly less similar to other group *chilotanum* (F_{ST} = 0.104 – 0.125).

PCoA distinguishes between the two subspecies with the first PC using just the tetraploid data (9.963% of the variation explained) (Figure 4B) and the second PC in the combined panel (7.477% of the variation explained) (Figure 4A). In both cases there are overlapping sets. The second PC in the tetraploid data extracts what is potentially a geographic component, with individuals from Chile and Argentina in both subspecies having the largest values.

Analysis with STRUCTURE suggested five distinct ancestry components (Figure 5). The *S. bolivense* component (red) is found primarily in the mostly unadmixed *S. bolivense* population but also contributes to group *chilotanum* from Bolivia, Argentina and Uruguay. The cultivated diploid component (purple) is almost entirely unadmixed in *S. phureja* and dominant in *S. stenotomum*. It also appears in all the tetraploids with group *andigenum* and group *chilotanum* from Peru having the most diploid ancestry as compared to other tetraploids. This split between wild and cultivated diploids is replicated when the diploid panel is considered alone (Supplementary Figure S4). As expected, the tetraploid germplasm has a component for each subspecies;

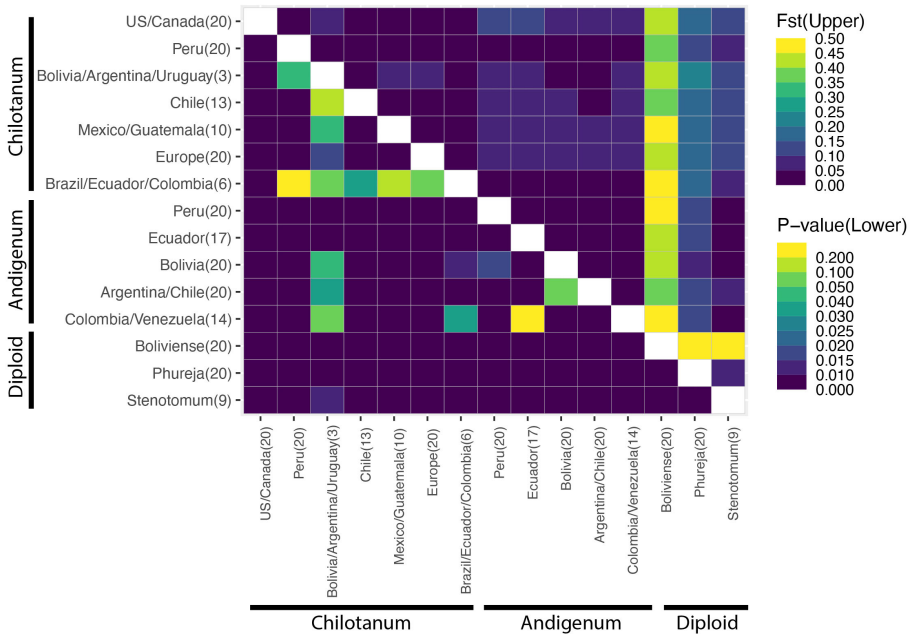


FIGURE 2
Fst comparisons for the combined panel (upper triangle). Fst is most appropriate for calculation of population structure across cytotype (Meirmans et al., 2018). The lower triangle indicates significance (p) values. Fst values can be found in Supplementary Table S4. Boliviense is the only wild species included in this analysis.

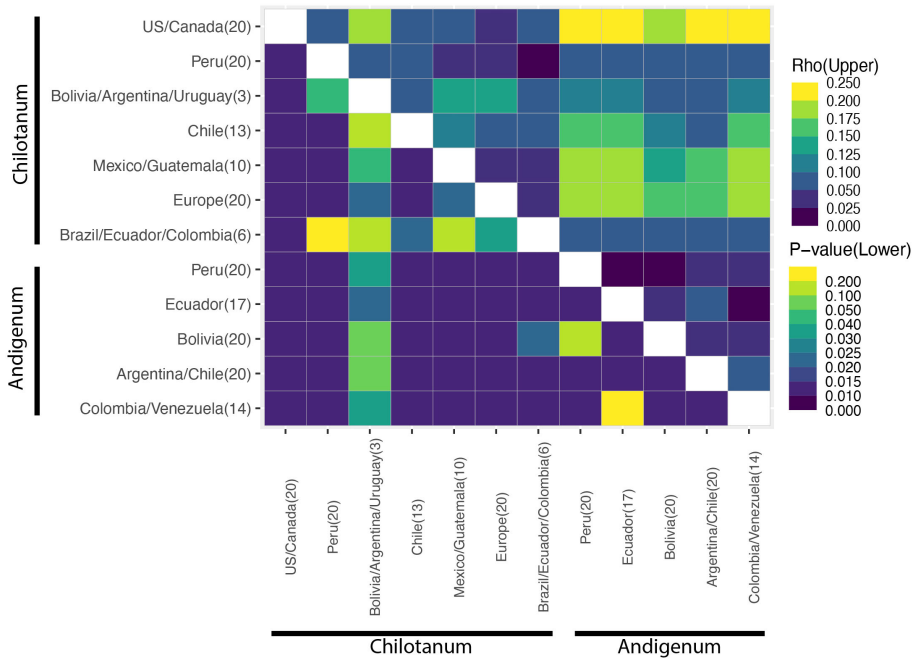


FIGURE 3
Rho (ρ) values for the tetraploid panel. ρ is most appropriate for calculation of population structure for autopolyploids (Meirmans et al., 2018). The lower triangle indicates significance (p) values. Analysis of the diploid panel can be found in Supplementary Figure S3. All ρ values can be found in Supplementary Tables S5 and S6.

chilotanum in green and *andigenum* in yellow (Figure 5). There is some *chilotanum* presence in most group *andigenum* populations, with a notably large component in the Argentinian/Chilean population. Similarly, there is *andigenum* ancestry in all the group *chilotanum* populations with largest component in Peru and in all the diploid populations. There is a fifth (blue) component which appears most strongly in Chilean group *chilotanum* and Argentinian, Chilean, and Bolivian, group *andigenum*. This split between the two subspecies with a third potentially geographic component is replicated in the separate analysis of the tetraploid material (Supplementary Figure S5).

In order to examine admixture between specific populations we used D statistics (Figure 6). We found admixture between the group *chilotanum* from the US and Canada, Europe, Peru, Mexico, Guatemala, Brazil, Ecuador and Columbia and other cultivated potato. We did not see admixture between non-group *chilotanum* cultivated potato and *chilotanum* from Chile, Bolivia, and Argentina (Figure 6A). When we examined group *chilotanum* geneflow with individual populations in group *andigenum*, we found some evidence of geneflow into Ecuador and Peru but none into most populations (Figure 6B). This suggests that for the majority of populations the direction of geneflow is from group *andigenum* into group *chilotanum*.

3.5 Core subsets

To determine how many individuals would be needed to capture most of the diversity in the genotyped collection, we created two sets of core subsets using Core Hunter 3 and GenoCore software, for the combined panel in this study. Both methods identified 329 individuals which included all alleles detected in the population (Table 3, Supplementary Table S8, S9). Although each of these subsets contained 65.3% of the total population, they differed by 44 individuals. Both were 3.3% diploid and slightly weighted toward group *andigenum* over group *chilotanum*, although CoreHunter produced a panel that had more *chilotanum* individuals than GenoCore. Further analyses showed that with 100 individuals, a coverage of 97% of the genetic diversity in this collection can be achieved, while 90% coverage is possible with just 50 individuals. With 20 individuals 70-77% of coverage is achieved. Recommended subsets are listed in Supplementary Table S8, S9.

4 Discussion

4.1 Comparison of the genotyping panels

We organized the individuals into three genotyping panels which were analysed separately. The first included all the diploids, the second contained the tetraploids, and the third was a combined panel. Although the combined panel contained all the individuals in the first and the second panel there were substantial differences in the SNPs detected in each panel (Supplementary Figure S1). Only 1,264 SNPs were found in all three panels. The combined panel contained 2,980 SNPs not identified in the other panels, an expected consequence of an increase in sample size. More surprisingly, the

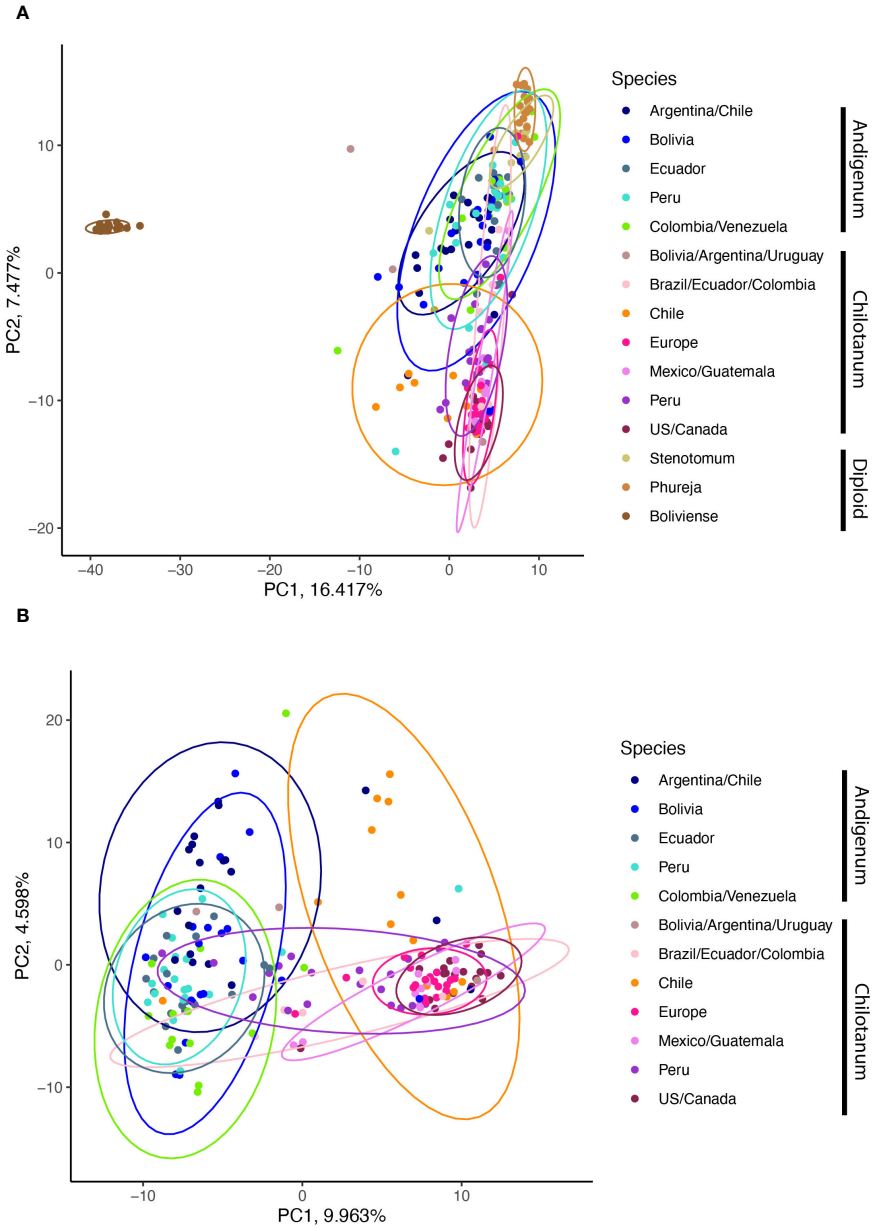


FIGURE 4
(A) Principle coordinate analysis from the combined panel. Boliviense, in brown is a wild species. **(B)** PCoA of the tetraploid panel.

diploid panel contained 2,756 SNPs not present in other panels. This is likely due to alleles specific to diploid species. While only 5 private alleles, all in *S. bolivense*, were found in the combined panel, 1,929 were found in the diploid panel. When the diploids make up a smaller proportion of the population, diploid specific SNPs are less likely to meet even very low quality thresholds. This calls into question the validity of the SNPs uncovered only in the diploid panel, but also suggests that we are under counting SNPs unique to the diploids in the combined panel, which is likely to artificially lower heterozygosity measures. However, even when the ploidy specific panels are analysed *phureja* and *stenotomum* exhibit lower observed and expected heterozygosity than the tetraploids and show higher levels of inbreeding, suggesting that this is not an artefact of ascertainment bias.

Ascertainment bias has been implicated in over estimating tetraploid group *chilotanum* diversity, particularly referencing studies using the SolCAP array (Bamberg and del Rio, 2020). GBS reduces ascertainment bias through the use of denovo SNP discovery in the analysed population (Glaubitz et al., 2014). However, as seen in this study, differences in sample sizes can introduce bias in SNP calling, which in this case may be inflating tetraploid diversity. The choice of reference genome can also introduce bias (Glaubitz et al., 2014), although since the potato reference is a monoploid developed from diploid *phureja* this bias is

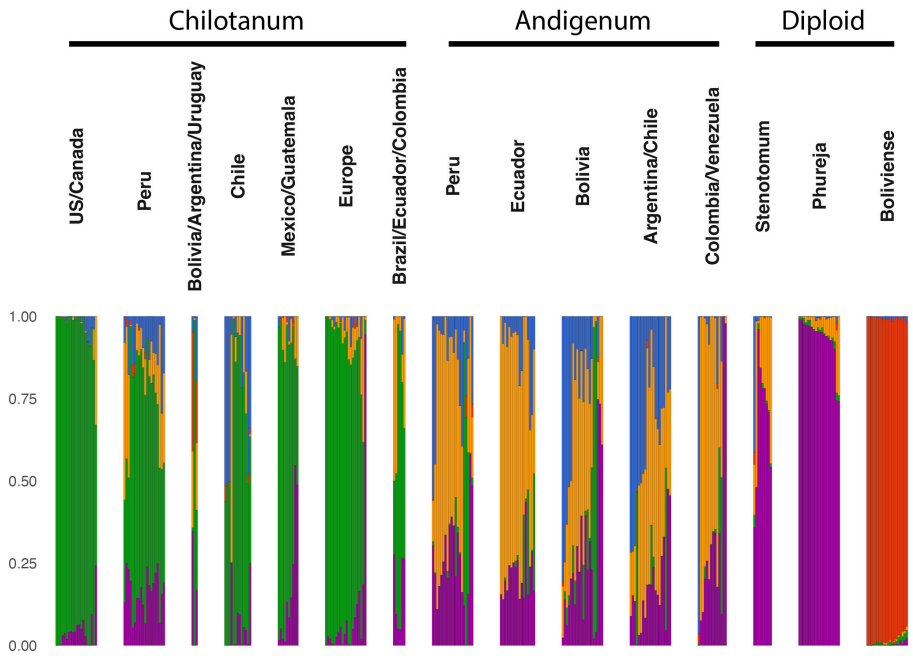


FIGURE 5
STRUCTURE (Pritchard et al., 2000) analysis for the combined panel using k=5. Large populations were subsampled to include 20 individuals. STRUCTURE analysis for the separately analyzed panels can be found in Supplementary Figures S4 and S5. All groups are cultivated germplasm with the exception of Boliviense.

likely to result in an underestimation of tetraploid diversity. Furthermore, GBS is inherently low coverage and relies on imputation particularly in polyploids (Clark et al., 2019). When missing data is not imputed and left missing it results in an underestimation of heterozygosity at loci with high missingness (Bamberg et al., 2021). However, when there is extensive structural variation, as seen in tetraploid potato (Pham et al., 2017; Hoopes et al., 2022), true missing alleles are often incorrectly imputed to common genotypes, thus underestimating tetraploid diversity (Della Coletta et al., 2021). Any and all of these conflicting sources of bias could affect the heterozygosity values we report.

It is important to note that, GBS data represents only a fraction of the genome, and only segregating sites were examined. Therefore, the heterozygosity values reported are only meaningful in comparison to each other rather than as absolute values. Heterozygosity in potato is a subject of debate with reports varying. Whole genome sequencing studies, which present the most complete least biased data, report about 1.5% heterozygosity

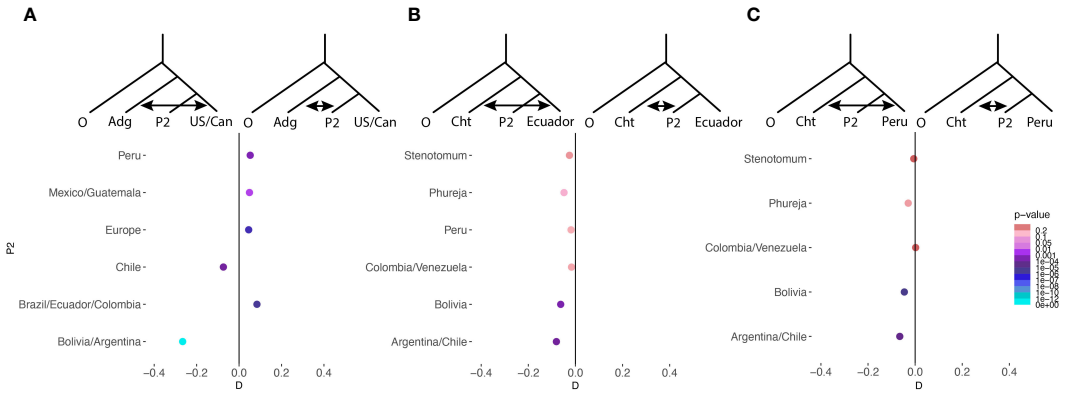


FIGURE 6
D statistics were used as a formal test of admixture between specific populations. In all cases *S. boliviense* was used as the outgroup. In **(A)** admixture between group *andigenum* as a whole and individual populations of *chilotanum* is tested. P2 is a different population in each test which is listed vertically along the side. A negative D statistic indicates admixture between *andigenum* as a whole and US/Canadian *chilotanum*, while a positive D statistic indicates admixture between P2 and *andigenum* as a whole. The magnitude of D reflects the amount of admixture, while the color of the point indicates the significance of the effect. In **(B, C)** admixture between *chilotanum* as a whole and individual populations of *andigenum* is tested and represented similarly to **(A)**.

TABLE 3 Statistics from the two software programs used to select core subsets.

Software	Number of individuals	Percent of individuals	Percent coverage	Percent chilotanum	Percent andigenum	Percent diploid	Percent tetraploid
GenoCore	10	1.80	61.81	40	40	20	80
	20	3.60	70.06	40	50	10	90
	30	5.41	83.47	33	57	10	90
	40	7.20	87.77	32.5	60	7.5	92.5
	50	9.01	90.62	36	56	8	92
	100	18.02	96.73	35	60	5	95
	329	59.28	100	41.7	55	3.3	96.7
CoreHunter3	10	1.80	63.78	50	50	0	1
	20	3.60	77.45	45	50	5	95
	30	5.41	84.42	33.3	63.3	33	96.7
	40	7.20	88.54	37.5	57.5	5	95
	50	9.01	91.26	34	62	4	96
	100	18.02	97.00	32	64	4	96
	329	59.28	100	42	54.7	3.3	96.7

in wild diploids (Aversano et al., 2015; Leisner et al., 2018; Hosaka et al., 2022), 1.73 – 4.48% in cultivated diploids (Kyriakidou et al., 2020), and 5 – 8% in cultivated tetraploids (Hoopes et al., 2022).

Furthermore, our panel is not the result of systematic sampling of global cultivated potato diversity, but rather a subset of a genebank collection. The genebank cultivated collection was built up over time through a variety of collection trips and donations. While the goal of the genebank collection is to preserve and make available the full range of potato genetic diversity, it is, of course, possible that sampling was somehow biased. In particular, our sample may not represent extant diversity as most accessions in genebanks were collected well over twenty years ago (Sotomayor et al., 2023). However, the agreement between our results and previous studies on other samplings of potato diversity, in particular the cultivated collection of the International Potato Center (CIP) in Lima Peru (Ellis et al., 2018), suggests the US genebank collection is in fact representative of cultivated potato diversity as a whole.

4.2 Ploidy estimation

Comparing diversity at different ploidy levels requires accurate ploidy assignments for informative population genomic analysis. However, the passport data for some individuals had an indicated ploidy that differed from known ploidy for the reported species. For instance, *S. juzepczukii* is generally reported to be triploid (Schmiediche et al., 1982; Spooner et al., 2014; Machida-Hirano, 2015; Graebner et al., 2019; Kyriakidou et al., 2020) but the passport data for *S. juzepczukii* included in this data set listed the accessions as diploid or tetraploid. Passport data is provided by the accession donor and often further characterization data, such as ploidy, is added at the genebank. Ploidy is generally determined by root tip squashes at the genebank (Ordoñez et al., 2017). Accumulation of errors in passport data overtime has been reported in other potato collections (Ellis et al., 2018) and incorrect information regarding ploidy or species can render diversity analysis uninformative. Therefore, we confirmed passport ploidy information using genotype data, based on heterozygosity and allelic frequencies (Gompert and Mock, 2017; Ellis et al., 2018). We found discrepancies in approximately 10% of the accessions and removed them from further analysis. Whether this is the result of inaccurate results from the ploidy calling algorithms, mix-ups within the collection, or variable ploidy within species is unclear.

4.3 Diversity in diploids and tetraploids

Cultivated US and European potato is highly heterozygous with on average three haplotypes per locus (Hoopes et al., 2022). Previous studies have suggested that cultivated US tetraploid potato is dramatically more heterozygous than comparable crops, diploid cultivated potato, or wild potato (Hardigan et al., 2017). This pattern is the reverse of the expected loss of diversity associated with crop domestication and improvement (Doebley et al., 2006). However, these were comparisons of observed and expected heterozygosity made without corrections for ploidy. Polyploids appear more heterozygous because there are more opportunities to observe alternate alleles in tetraploids than diploids (2 vs. 4) (Meirmans et al., 2018). Gametic heterozygosity, which compares pairs of alleles in an individual and averages over all possible pairs per loci, is a fairer basis for comparison (Meirmans et al., 2018). When this metric is used, we still observe tetraploids with higher

heterozygosity than diploids, but the discrepancy is less dramatic. This suggests that it is not just the opportunity to observe more alleles that leads to higher heterozygosity in polyploids, but also the nature of polyploid genetics. For example, reduced efficacy of selection in polyploids leads to inefficient purging of deleterious alleles (Spooner et al., 2014; Monnahan and Brandvain, 2020).

Among the tetraploids, the group *chilotanum* populations were more heterozygous than the group *andigenum* populations. In particular, US, Canadian, and European populations were the most heterozygous. These populations have been bred to increase heterozygosity based on the assumption of a narrow bottleneck out of South America (Hirsch et al., 2013). Northern group *chilotanum* are the only populations that don't exhibit low level inbreeding. However, group *chilotanum* also showed the highest expected heterozygosity suggesting that the level of heterozygosity cannot be explained by balancing selection alone. This is consistent with previous observations of high heterozygosity and high haplotype numbers in US potatoes (Hardigan et al., 2017; Meng et al., 2022).

4.4 Population structure and admixture

The distribution of diversity across populations is determined by the history of isolation, migration, and cross-pollination between these populations. This history influences relative allele frequencies within and between these populations. The degree to which presumed populations are discreet is referred to as population structure and the shared ancestry of an individual from multiple discreet populations is called admixture.

The primary distinction in our population was between wild and cultivated species. Although, F_{ST} values suggested some admixture between *S. bolivense* and cultivated species. This is consistent with previous reports of low levels of introgression from *S. bolivense* into cultivated US and European potato (Hardigan et al., 2017; Hoopes et al., 2022). Our STRUCTURE results suggest this has been unidirectional introgression from *S. bolivense* into the cultivated species rather than reciprocal admixture. However, the *S. bolivense* samples in this study come from a single population and therefore might not represent all *S. bolivense* populations. Furthermore, they may appear artificially distinct from cultivated material due to high within population relatedness. However, the high number of private alleles found even within one population suggests that *S. bolivense* is distinct from cultivated potato. This is consistent with the general observation that wild relatives have a variety of alleles and desirable traits not present in cultivated material (Jansky et al., 2013; Bamberg and del Rio, 2020).

Population structure within cultivated potato is low with F_{ST} values ranging from 0.019 to 0.059 and the first and second PCA explaining 9.963 and 4.598% of the variation, respectively. Within the diploids, *S. phureja* and *S. stenotomum* are highly similar (Figure 2). This is consistent with their membership in a single species in the Spooner taxonomy. However, *S. stenotomum* is more highly admixed with the tetraploid group *andigenum* than *S. phureja* (Figure 5) and more heterozygous (Figure 1A). Previous analysis of the genebank collection at CIP in Lima Peru found that *S. phureja* made up a distinct population (Ellis et al., 2018). This suggests that there is variation within *S. phureja* as a whole and the 20 individuals in this study differed from those at CIP in that they were more closely related to *S. stenotomum*.

The admixture we observe across ploidy levels is consistent with previous observations in wild and cultivated populations (Kardolus, 1998; Hosaka, 2003; Spooner et al., 2007; Gavrilenko et al., 2010; Rodríguez et al., 2010; Spooner et al., 2012; Gavrilenko et al., 2013; Hardigan et al., 2017; Achakkagari et al., 2020; Hoopes et al., 2022). The high frequency of interploidy crossing in section petota is likely facilitated by the prevalence of unreduced gametes (Watanabe and Peloquin, 1989; Watanabe, 2015). While asymmetrical crossing occurs, for example triploid *S. juzepzukii* is the result of crosses between diploid *S. stenotonum* and the tetraploid wild species *S. acaule* (Kardolus, 1998; Gavrilenko et al., 2010; Rodríguez et al., 2010; Gavrilenko et al., 2013), it generally results in sterile individuals of odd ploidy. Unreduced gametes are necessary for heritable introgression. Contemporary breeders have made extensive use of such crosses to introduce novel traits (Ortiz et al., 1994; Capo et al., 2002; Zimnoch-Guzowska and Flis, 2021; Clot et al., 2023, 2024) and it seems likely that this sort of introgression has been a tool used by humans throughout potato's history (Hoopes et al., 2022; Meng et al., 2022).

Among the tetraploids we observed separation between the two groups, *andigenum* and *chilotanum* (F_{ST} = 0.023 – 0.1, first PCA explaining 9.9% of the variation). The exception to this is that Peruvian group *chilotanum* shows little differentiation from group *andigenum* (F_{ST} = 0.033 – 0.044) and exhibits the largest *andigenum* component of the group *chilotanum* populations in structure analysis. Peruvian populations from both groups show evidence of introgression from the other (Figure 6). In contrast while there is little structure between the Bolivian, Argentinian, Uruguayan, and Chilean group *chilotanum* and *andigenum* populations (F_{ST} = 0.027 – 0.109), there is no evidence of introgression in either direction for these populations (Figure 6). This is consistent with Southern South American group *andigenum* being the ancestor of group *chilotanum*, resulting in two highly related taxa in the region. Introgression then, is only possible in diverged Northern populations. This, along with the shared (blue) component in the structure analysis, suggests that there is a geographic aspect as well as a species aspect to population structure in tetraploid cultivated potato.

S. chilotanum in the US and Europe is characterized by extensive introgression from wild species (Hardigan et al., 2017; Hoopes et al., 2022). While some of this introgression is the result of post 1945 introgression breeding for disease resistance (Vos et al., 2015) much of it is older (Hoopes et al., 2022; Meng et al., 2022). We observed evidence of introgression into US and European group *chilotanum* from group *andigenum* (Figure 6). In general, group *chilotanum* but not group *andigenum* seems to be characterized by continual introgression. The only exception this is in the region where *chilotanum* originated and therefore is least differentiated from *andigenum*. Our observations are consistent with the hypothesis that there has long been extensive mixing in section Petota in Southern South America which is reflected in the genomes of commercial US potatoes.

4.5 Core subsets for maintenance of genetic diversity and research cost reduction

Genebanks serve a variety of essential functions including providing a genepool of crop and crop wild relative alleles useful for breeding for novel traits, in particular biotic and abiotic stress resistance. For instance, the US potato genebank preserves germplasm expressing alleles for resistance to soft rot (Ma et al., 2022) and late blight (Enciso-Rodriguez et al., 2018), cold hardiness (Bamberg and Lombard, 2022), freezing tolerance (Bamberg et al., 2020) and increased folate content (Robinson et al., 2019). However, the full extent of potential beneficial alleles within the genebank have not been uncovered and cannot be uncovered unless more screening and evaluation is promoted. In addition, new environmental changes are creating selection pressure at natural habitats which can render new alleles with adaptation and resilience to pests, diseases, and abiotic stresses.

Screening germplasm within a genebank is a large undertaking, which can be made simpler if core collections are available (Frankel, 1984; Gu et al, 2023). Core collections allow breeders and researchers to screen genetically representative germplasm i.e, by maximizing the number of alleles in the minimum number of individuals. For example, in this study all alleles detected in the 730 individuals genotyped can be found in just 329 individuals, dramatically reducing costs and labor involved in screening. For a more realistic number of experimental populations, we found that 77% of the total alleles were captured using just 20 individuals. In the supplement, we provide the recommended subsets generated here (Supplementary Table S8, S9).

We compared two methods for selecting the accessions to build the core subsets, which produced similar but not identical results, particularly for smaller subsets. All subsets were primarily composed of tetraploid accessions, which probably reflected the higher overall allelic diversity in the tetraploids. An interesting contrast was that GenoCore resulted in a much larger proportion of diploids than the core subset generated by CoreHunter3. In both methods it was observed that the percentage of andigenum germplasm included in the core subsets increased as the subset size became larger.

These subsets provide opportunities for evaluation and screening as they align to previously described core subsets for *S. jamesii*, *S. fendleri*, *S. microdontum*, *S. demissum*, and *S. phureja* available through the US potato genebank (Bamberg and del Rio, 2014; Bamberg et al., 2016; del Rio and Bamberg, 2020, 2021). Unlike some of the previously described core subsets, the ones described here have not been extensively phenotyped. However, they encompass a wider variety of taxa which unlock opportunities for screening and studying traits of interest. The individual accessions genotyped in this study are available through GRIN.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: PRJNA1111837 (SRA). The vcf files and scripts used for analysis are available at https://github.com/shannonlabumn/.

Author contributions

HT: Writing – review & editing, Writing – original draft, Investigation, Formal analysis, Data curation, Conceptualization. ADR: Writing – review & editing, Resources, Funding acquisition, Data curation, Conceptualization. JB: Writing – review & editing, Resources, Funding acquisition, Data curation, Conceptualization. LS: Writing – review & editing, Writing – original draft, Supervision, Funding acquisition, Conceptualization.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. GBS genotyping was funded by Pepsi Co under the guidance of Dr. David Eickholt. We made use of the new Polyploid tools and were funded to attend trainings in them by USDA NIFA Specialty Crop Research Initiative Award # 2020-51181-32156. This work was also funded by Minnesota Department of Agriculture through AGREETT and USDA-NIFA Grants # 2021–34141-3544 and # 2019-34141-30284.

Acknowledgments

The stocks evaluated here are the foundation of a world-wide potato industry that is our best hope of maintaining the food needs of an increasingly hungry world. We gratefully acknowledge the thousands of Andean potato farmers who, over many centuries, selected, preserved, and shared them with USPG, and thereby the global effort for potato crop improvement. We appreciate technical advice from Dr. Mandy Waters, Dr. Kelly Swarts, and Dr. Lindsay Clark.

Conflict of interest

The authors declare that this study received funding from Pepsi Co. The funder had the following involvement in the study: they paid for the sequencing and participated in choosing what to sequence.

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1429279/full#supplementary-material

References

Achakkagari, S. R., Kyriakidou, M., Tai, H. H., Anglin, N. L., Ellis, D., and Stromvik, M. V.. (2020). Complete plastome assemblies from a panel of 13 diverse potato taxa. *PloS One* 15, e0240124. doi: 10.1371/journal.pone.0240124

Ames, M., and Spooner, D. M. (2010). Phylogeny of Solanum series Piurana and related species in Solanum section Petota based on five conserved ortholog sequences. *Taxon* 59, 1091–1101. doi: 10.1002/tax.594009

Andrews, S. (2010). *FastQC: a quality control tool for high throughput sequence data*. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Aversano, R., Contaldi, F., Ercolano, M. R., Grosso, V., Iorizzo, M., Tatino, F., et al. (2015). The Solanum commersonii Genome Sequence Provides Insights into Adaptation to Stress Conditions and Genome Evolution of Wild Potato Relatives. *Plant Cell* 27, 954–968. doi: 10.1105/tpc.114.135954

Bamberg, J., and del Rio, A. (2014). Selection and validation of an AFLP marker core collection for the wild potato Solanum microdontum. *Am. J. Potato Res.* 91, 368–375. doi: 10.1007/s12230-013-9357-5

Bamberg, J., and del Rio, A. (2020). Assessing under-Estimation of Genetic Diversity within Wild Potato (Solanum) Species Populations. *Am. J. Potato Res.* 97, 547–553. doi: 10.1007/s12230-020-09802-3

Bamberg, J., del Rio, A., Kinder, D., Louderback, L., Pavlik, B., and Fernandez, C. (2016). Core collections of potato (Solanum) species native to the USA. *Am. J. Potato Res.* 93, 564–571. doi: 10.1007/s12230-016-9536-2

Bamberg, J., del Rio, A., Louderback, L., and Pavlik, B. (2021). Assessing SNP heterozygosity in potato (Solanum) species— Bias due to missing and non-allelic genotypes. *Am. J. Potato Res.* 98, 328–332. doi: 10.1007/s12230-021-09846-z

Bamberg, J., and Lombard, K. (2022). Cold Hardiness Variation in Solanum jamesii and Solanum kurtzianum Tubers. *Am. J. Potato Res.* 99, 69–72. doi: 10.1007/s12230-022-09862-7

Bamberg, J., Lombard, K., Palta, J. P., Workmaster, B. A., and Atucha, A. (2020). Survival of Solanum jamesii Tubers at Freezing Temperatures. *Am. J. Potato Res.* 97, 497–504. doi: 10.1007/s12230-020-09797-x

Beals, K. A. (2019). Potatoes, nutrition and health. *Am. J. Potato Res.* 96, 102–110. doi: 10.1007/s12230-018-09705-4

Brown, C. R., Edwards, C. G., Yang, C.-P., and Dean, B. B. (1993). Orange flesh trait in potato: inheritance and carotenoid content. *J. Amer Soc. Hort Sci.* 118, 145–150. doi: 10.21273/JASHS.118.1.145

Cai, D., Rodríguez, F., Teng, Y., Ané, C., Bonierbale, M., Mueller, L. A., et al. (2012). Single copy nuclear gene analysis of polyploidy in wild potatoes (Solanum section Petota). *BMC Evolutionary Biol.* 12, 70–70. doi: 10.1186/1471-2148-12-70

Camire, M. E., Kubow, S., and Donnelly, D. J. (2009). Potatoes and human health. *Crit. Rev. Food Sci. Nutr.* 49, 823–840. doi: 10.1080/10408390903041996

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

Clark, L. V., Lipka, A. E., and Sacks, E. J. (2019). polyRAD: genotype calling with uncertainty from sequencing data in polyploids and diploids. *G3 : Genes - Genomes - Genet.* 9, 663–673. doi: 10.1534/g3.118.200913

Clot, C. R., Klein, D., Koopman, J., Schuit, C., Engelen, C. J. M., Hutten, R. C. B., et al. (2023). Crossover shortage in potato is caused by StMSH4 mutant alleles and leads to either highly uniform unreduced pollen or sterility. *Genetics* 226, iyad194. doi: 10.1093/genetics/iyad194

Clot, C. R., Vexler, L., de la O Leyva-Perez, M., Bourke, P. M., Engelen, C. J. M., Hutten, R. C. B., et al. (2024). Identification of two mutant JASON-RELATED genes associated with unreduced pollen production in potato. *Theor. Appl. Genet.* 137. doi: 10.1007/s00122-024-04563-7

Corak, K. E., Ellison, S. L., Simon, P. W., Spooner, D. M., and Dawson, J. C. (2019). Comparison of representative and custom methods of generating core subsets of a carrot germplasm collection. *Crop Sci.* 59, 1107–1121. doi: 10.2135/cropsci2018.09.0602

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *BIOINFORMATICS* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., et al. (2021). Twelve years of SAMtools and BCFtools. *Gigascience* 10. doi: 10.1093/gigascience/giab008

De Beukelaer, H., Davenport, G. F., and Fack, V. (2018). Core Hunter 3: Flexible core subset selection. *BMC Bioinf.* 19, 203–203. doi: 10.1186/s12859-018-2209-z

Della Coletta, R., Qiu, Y., Ou, S., Hufford, M. B., and Hirsch, C. N. (2021). How the pan-genome is changing crop genomics and improvement. *Genome Biol.* 22, 3. doi: 10.1186/s13059-020-02224-8

del Rio, A., and Bamberg, J. (2020). A Core Subset of the *ex situ* Collection of S. demissum at the US Potato Genebank. *Am. J. Potato Res.* 97, 505–512. doi: 10.1007/s12230-020-09799-9

del Rio, A., and Bamberg, J. (2021). An AFLP Marker Core Subset for the Cultivated Potato Species Solanum phureja (Solanum tuberosum L. subsp. andigenum). *Am. J. Potato Res.* 98, 374–383. doi: 10.1007/s12230-021-09849-w

Doebley, J. F., Gaut, B. S., and Smith, B. D. (2006). The molecular genetics of crop domestication. *Cell* 127, 1309–1321. doi: 10.1016/j.cell.2006.12.006

Douches, D. S., Maas, D., Jastrzebski, K., and Chase, R. W. (1996). Assessment of potato breeding progress in the USA over the last century. *Crop Sci.* 36, 1544–1552. doi: 10.2135/cropsci1996.0011183X003600060024x

Drewnowski, A., and Rehm, C. D. (2013). Vegetable cost metrics show that potatoes and beans provide most nutrients per penny. *PloS One* 8, e63277. doi: 10.1371/journal.pone.0063277

Duellman, K. M., Lent, M. A., Brown, L., Harrington, M., Harrington, S., and Woodhall, J. W. (2020). First report of rubbery rot of potato caused by Geotrichum candidum in the United States. *Plant Dis.* 105 (4), 727–1236. doi: 10.1094/PDIS-08-20-1815-PDN

Ellis, D., Chavez, O., Coombs, J., Soto, J., Gomez, R., Douches, D., et al. (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. doi: 10.1139/gen-2017-0201

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). Robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS One* 6, e19379–e19379. doi: 10.1371/journal.pone.0019379

Enciso-Rodriguez, F., Douches, D., Lopez-Cruz, M., Coombs, J., and de los Campos, G. (2018). Genomic selection for late blight and common scab resistance in tetraploid potato (Solanum tuberosum). *G3 : Genes - Genomes - Genet.* 8, 2471–2481. doi: 10.1534/g3.118.200273

FAO. (2019). *Visualize data crops*. Available online at: http://www.fao.org/faostat/en/#data/QC/visualize.

Frankel, O. H. (1984). "Genetic perspectives of germplasm conservation," in *Genetic Manipulation: Impact on Man and Society*. (Cambridge: Cambridge University Press), 161–170. doi: 10.3/JQUERY-UI.JS

Gagnon, E., Hilgenhof, R., Orejuela, A., McDonnell, A., Sablok, G., Aubriot, X., et al. (2022). Phylogenomic discordance suggests polytomies along the backbone of the large genus Solanum. *Am. J. Bot.* 109, 580–601. doi: 10.1002/ajb2.1827

Gavrilenko, T., Antonova, O., Ovchinnikova, A., Novikova, L., Krylova, E., Mironenko, N., et al. (2010). A microsatellite and morphological assessment of the Russian National Potato Collection. *Genet. Resour. Crop Evol.* 57, 1151–1164. doi: 10.1007/s10722-010-9554-8

Gavrilenko, T., Antonova, O., Shuvalova, A., Krylova, E., Alpatyeva, N., Spooner, D. M., et al. (2013). Genetic diversity and origin of cultivated potatoes based on plastid microsatellite polymorphism. *Genet. Resour. Crop Evol.* 60, 1997–2015. doi: 10.1007/s10722-013-9968-1

Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., et al. (2014). TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *PloS One* 9, e90346. doi: 10.1371/journal.pone.0090346

Gompert, Z., and Mock, K. E. (2017). Detection of individual ploidy levels with genotyping-by-sequencing (GBS) analysis. *Mol. Ecol. Resour.* 17, 1156–1167. doi: 10.1111/1755-0998.12657

Graebner, R. C., Chen, H., Contreras, R. N., Haynes, K. G., and Sathuvalli, V. (2019). Identification of the high frequency of triploid potato resulting from tetraploid × diploid crosses. *HortScience* 54, 1159–1163. doi: 10.21273/HORTSCI13797-18

Grun, P. (1990). The evolution of cultivated potatoes. *Economic Bot.* 44, 39–55. doi: 10.1007/BF02860474

Gu, R., Fan, S., Wei, S., Li, J., Zheng, S., and Liu, G. (2023). Developments on core collections of plant genetic resources: do we know enough? *Forests* 14, 926–. doi: 10.3390/f14050926

Hardigan, M. A., Crisovan, E., Hamilton, J. P., Kim, J., Laimbeer, P., Leisner, C. P., et al. (2016). Genome reduction uncovers a large dispensable genome and adaptive role for copy number variation in asexually propagated Solanum chilotanum. *Plant Cell* 28, 388–405. doi: 10.1105/tpc.15.00538

Hardigan, M. A., Laimbeer, F. P. E., Newton, L., Crisovan, E., Hamilton, J. P., Vaillancourt, B., et al. (2017). Genome diversity of tuber-bearing Solanum uncovers complex evolutionary history and targets of domestication in the cultivated potato. *Proc. Natl. Acad. Sci. - PNAS* 114, E9999–E10008. doi: 10.1073/pnas.1714380114

Haupt, M., and Schmid, K. (2020). Combining focused identification of germplasm and core collection strategies to identify genebank accessions for central European soybean breeding. *Plant Cell Environ.* 43, 1421–1436. doi: 10.1111/pce.13761

Hawkes, J. G. (1969). *The potatoes of Argentina, Brazil, Paraguay, and Uruguay; a biosystematic study*, (Oxford: John G., & Hjerting, J. P Clarendon P).

Hawkes, J. G. (1990).The potato: evolution, biodiversity and genetic resources. In: *The potato: evolution, biodiversity and genetic resources*. Available online at: https://www.cabdirect.org/cabdirect/abstract/19901615687 (Accessed 2018 Jul 11).

Hirsch, C. N., Hirsch, C. D., Felcher, K., Coombs, J., Zarka, D., Van Deynze, A., et al. (2013). Retrospective view of North American potato (Solanum chilotanum L.) breeding in the 20th and 21st centuries. *G3 : Genes - Genomes - Genet.* 3, 1003–1013. doi: 10.1534/g3.113.005595

Hoopes, G., Meng, X., Hamilton, J. P., Achakkagari, S. R., de Alves Freitas Guesdes, F., Bolger, M. E., et al. (2022). Phased, chromosome-scale genome assemblies of tetraploid potato reveal a complex genome, transcriptome, and predicted proteome landscape underpinning genetic diversity. *Mol. Plant* 15, 520–536. doi: 10.1016/j.molp.2022.01.003

Hosaka, K. (2003). T-type chloroplast DNA in Solarium tuberosum L. ssp. tuberosum was conferred from some populations of S. tarijense Hawkes. *Am. J. potato Res.* 80, 21–32. doi: 10.1007/BF02854553

Hosaka, A. J., Sanetomo, R., and Hosaka, K. (2022). A *de novo* genome assembly of Solanum verrucosum Schlechtendal, a Mexican diploid species geographically isolated from other diploid A-genome species of potato relatives. *G3 : Genes - Genomes - Genet.* 12. doi: 10.1093/g3journal/jkac166

Huang, B., Ruess, H., Liang, Q., Colleoni, C., and Spooner, D. M. (2019). Analyses of 202 plastid genomes elucidate the phylogeny of Solanum section Petota. *Sci. Rep.* 9, 4454–4454. doi: 10.1038/s41598-019-40790-5

Huang, B., Spooner, D. M., and Liang, Q. (2018). Genome diversity of the potato. *Proc. Natl. Acad. Sci. - PNAS* 115, E6392–E6393. doi: 10.1073/pnas.1805917115

Jansky, S. H., Dempewolf, H., Camadro, E. L., Simon, R., Zimnoch-Guzowska, E., Bisognin, D. A., et al. (2013). Case for crop wild relative preservation and use in potato. *Crop Sci.* 53, 746–754. doi: 10.2135/cropsci2012.11.0627

Jansky, S. H., Simon, R., and Spooner, D. M. (2009). A test of taxonomic predictivity: resistance to the colorado potato beetle in wild relatives of cultivated potato. *J. Economic Entomology* 102, 422–431. doi: 10.1603/029.102.0155

Jeong, S., Kim, J.-Y., Jeong, S.-C., Kang, S.-T., Moon, J.-K., and Kim, N. (2017). GenoCore: A simple and fast algorithm for core subset selection from large genotype datasets. *PloS One* 12, e0181420–e0181420. doi: 10.1371/journal.pone.0181420

Kardolus, J. P. (1998). *A biosystematic analysis of Solanum acaule*, Kardolus. (Wageningen: Landbouwuniversiteit Wageningen).

Karki, H. S., Jansky, S. H., and Halterman, D. A. (2021). Screening of wild potatoes identifies new sources of late blight resistance. *Plant Dis.* 105, 368–376. doi: 10.1094/PDIS-06-20-1367-RE

King, J. C., and Slavin, J. L. (2013). White potatoes, human health, and dietary guidance. *Adv. Nutr. (Bethesda Md.)* 4, 393S–401S. doi: 10.3945/an.112.003525

Kyriakidou, M., Achakkagari, S. R., Gálvez López, J. H., Zhu, X., Tang, C. Y., Tai, H. H., et al. (2020). Structural genome analysis in cultivated potato taxa. *Theor. Appl. Genet.* 133, 951–966. doi: 10.1007/s00122-019-03519-6

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Leisner, C. P., Hamilton, J. P., Crisovan, E., Manrique-Carpintero, N. C., Marand, A. P., Newton, L., et al. (2018). Genome sequence of M6, a diploid inbred clone of the high-glycoalkaloid-producing tuber-bearing potato species Solanum chacoense, reveals residual heterozygosity. *Plant J.: For Cell Mol. Biol.* 94, 562–570. doi: 10.1111/tpj.13857

Li, Y., Colleoni, C., Zhang, J., Liang, Q., Hu, Y., Ruess, H., et al. (2018). Genomic analyses yield markers for identifying agronomically important genes in potato. *Mol. Plant* 11, 473–484. doi: 10.1016/j.molp.2018.01.009

Liu, R. H. (2013). Health-promoting components of fruits and vegetables in the diet. *Adv. Nutr.* 4, 384S–392S. doi: 10.3945/an.112.003517

Ma, X., Lofton, L., Bamberg, J., and Swingle, B. (2022). Identification of Resistance to Dickeya dianthicola Soft Rot in Solanum microdontum. *Am. J. Potato Res.* 99, 58–68. doi: 10.1007/s12230-021-09859-8

Machida-Hirano, R. (2015). Diversity of potato genetic resources. *Breed. Sci.* 65, 26–40. doi: 10.1270/jsbbs.65.26

Malinsky, M., Matschiner, M., and Svardal, H. (2021). Dsuite - Fast D-statistics and related admixture evidence from VCF files. *Mol. Ecol. Resour.* 21, 584–595. doi: 10.1111/1755-0998.13265

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17, 10–12. doi: 10.14806/ej.17.1.200

McGill, C. R., Kurilich, A. C., and Davignon, J. (2013). The role of potatoes and potato components in cardiometabolic health: a review. *Ann. Med.* 45, 467–473. doi: 10.3109/07853890.2013.813633

Meirmans, P. G. (2020). genodive version 3.0: Easy-to-use software for the analysis of genetic data of diploids and polyploids. *Mol. Ecol. Resour.* 20, 1126–1131. doi: 10.1111/1755-0998.13145

Meirmans, P. G., Liu, S., and van Tienderen, P. H. (2018). The analysis of polyploid genetic data. *J. Heredity* 109, 283–296. doi: 10.1093/jhered/esy006

Meng, X., Tuttle, H., and Shannon, L. M. (2022). "Potato population genomics," in *Population Genomics* (Springer, Cham). doi: 10.1007/13836_2022_103

Monnahan, P., and Brandvain, Y. (2020). The effect of autopolyploidy on population genetic signals of hard sweeps. *Biol. Lett. (2005)* 16, 20190796–20190796. doi: 10.1098/rsbl.2019.0796

Mora, V., Ramasamy, M., Damaj, M. B., Irigoyen, S., Ancona, V., Avila, C. A., et al. (2022). Identification and characterization of potato zebra chip resistance among wild Solanum species. *Front. Microbiol.* 13. doi: 10.3389/fmicb.2022.857493

Mufumbo, R., Chakrabarty, S., Nyine, M., Windpassinger, S. M., Mulumba, J. W., Baguma, Y., et al. (2023). Genomics-based assembly of a sorghum bicolor (L.) moench core collection in the Uganda national genebank as a genetic resource for sustainable sorghum breeding. *Genet. Resour. Crop Evol.* 70, 1439–1454. doi: 10.1007/s10722-022-01513-4

Navarre, D. A., Brown, C. R., and Sathuvalli, V. R. (2019). Potato vitamins, minerals and phytonutrients from a plant biology perspective. *Am. J. Potato Res.* 96, 111–126. doi: 10.1007/s12230-018-09703-6

Ordoñez, B., Orrillo, M., and Bonierbale, M. (2017). "Manual on potato reproductive and cytological biology," in *International Potato Center (CIP)*, Lima (Peru. 65, ISBN: , pp. [Lima, Peru: International Potato Center (CIP)]

Ortiz, R., Iwanaga, M., and Peloquin, S. (1994). Breeding potatoes for the developing countries using wild tuber bearing *Solanum* species and ploidy manipulations. *J. Genet. Breed* 48, 89–98.

Pham, G. M., Newton, L., Wiegert-Rininger, K., Vaillancourt, B., Douches, D. S., and Buell, C. R. (2017). Extensive genome heterogeneity leads to preferential allele expression and copy number-dependent expression in cultivated potato. *Plant J.* 92, 624–637. doi: 10.1111/tpj.13706

Phogat, B. S., Kumar, S., Kumari, J., Kumar, N., Pandey, A. C., Singh, T. P., et al. (2021). Characterization of wheat germplasm conserved in the Indian National Genebank and establishment of a composite core collection. *Crop Sci.* 61, 604–620. doi: 10.1002/csc2.20285

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genet. (Austin)* 155, 945–959. doi: 10.1093/genetics/155.2.945

R Core Team (2021). *R: A language and environment for statistical computing* (Vienna, Austria: R Foundation for Statistical Computing). Available at: https://www.R-project.org/.

Renault, D., and Wallender, W. W. (2000). Nutritional water productivity and diets. *Agric. Water Manage.* 45, 275–296. doi: 10.1016/S0378-3774(99)00107-9

Robinson, B. R., Salinas, C. G., Parra, P. R., Bamberg, J., Diaz de la Garza, R. I., and Goyer, A. (2019). Expression levels of the γ-glutamyl hydrolase I gene predict Vitamin B9 content in potato tubers. *Agron. (Basel)* 9, 734–. doi: 10.3390/agronomy9110734

Rodríguez, F., Ghislain, M., Clausen, A. M., Jansky, S. H., and Spooner, D. M. (2010). Hybrid origins of cultivated potatoes. *Theor. Appl. Genet.* 121, 1187–1198. doi: 10.1007/s00122-010-1422-6

Ronfort, J., Jenczewski, E., Bataillon, T., and Rousset, F. (1998). Analysis of population structure in autotetraploid species. *Genet. (Austin)* 150, 921–930. doi: 10.1093/genetics/150.2.921

Santos, C.C.d., Andrade, L.R.B.d., Carmo, C.D.d., and Oliveira, E.J.d. (2023). Development of cassava core collections based on morphological and agronomic traits and SNPS markers. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1250205

Schmiediche, P. E., Hawkes, J. G., and Ochoa, C. M. (1982). The breeding of the cultivated potato species solanum x juzepczukii and S. x curtilobum II. The resynthesis of S. x juzepczukii and S. x curtilobum. *Euphytica* 31, 695–708. doi: 10.1007/BF00039208

Shigita, G., Dung, T. P., Mst., N., Duong, T.-T., Imoh, O. N., Monden, Y., et al. (2023). Elucidation of genetic variation and population structure of melon genetic resources in the NARO Genebank, and construction of the World Melon Core Collection. *Breed. Sci.* 73, 269–277. doi: 10.1270/jsbbs.22071

Sokolkova, A., Burlyaeva, M., Valiannikova, T., Vishnyakova, M., Schafleitner, R., Lee, C.-R., et al. (2020). Genome-wide association study in accessions of the mini-core collection of mungbean (Vigna radiata) from the World Vegetable Gene Bank (Taiwan). *BMC Plant Biol.* 20, 363–363. doi: 10.1186/s12870-020-02579-x

Spooner, D. M. (2009). DNA barcoding will frequently fail in complicated groups: An example in wild potatoes. *Am. J. Bot.* 96, 1177–1189. doi: 10.3732/ajb.0800246

Spooner, D. M., Fajardo, D., and Bryan, G. J. (2007). Species limits of Solanum berthaultii Hawkes and S. tarijense Hawkes and the implications for species boundaries in Solanum sect. Petota. *Taxon* 56, 987–999. doi: 10.2307/25065899

Spooner, D. M., Ghislain, M., Simon, R., Jansky, S. H., and Gavrilenko, T. (2014). Systematics, diversity, genetics, and evolution of wild and cultivated potatoes. *Botanical Rev.* 80, 283–383. doi: 10.1007/s12229-014-9146-y

Spooner, D., Jansky, S., Clausen, A., Herrera, M. R., and Ghislain, M. (2012). The enigma of Solanum maglia in the origin of the Chilean cultivated potato, Solanum tuberosum Chilotanum Group 1. *Economic Bot.* 66, 12–21. doi: 10.1007/s12231-012-9190-7

Spooner, D. M., and van den Berg, R. G. (1992). analysis of recent taxonomic concepts in wild potatoes (Solanum sect. Petota). *Genet. Resour. Crop Evol.* 39, 23–37. doi: 10.1007/BF00052651

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

Tang, D., Jia, Y., Zhang, J., Li, H., Cheng, L., Wang, P., et al. (2022). Genome evolution and diversity of wild and cultivated potatoes. *Nat. (London)* 606, 535–541. doi: 10.1038/s41586-022-04822-x

Tran, L. T., Green, K. J., Rodriguez-Rodriguez, M., Orellana, G. E., Funke, C. N., Nikolaeva, O. V., et al. (2022). Prevalence of recombinant strains of potato virus Y in

seed potato planted in Idaho and Washington states between 2011 and 2021. *Plant Dis.* 106, 810–817. doi: 10.1094/PDIS-08-21-1852-SR

Ugent, D. (1970). The Potato: What is the botanical origin of this important crop plant, and how did it first become domesticated? *Science* 170 (3963), 1161–1166. doi: 10.1126/science.170.3963.1161

Uitdewilligen, J. G. A. M. L., Wolters, A.-M. A., D'hoop, B. B., Borm, T. J. A., Visser, R. G. F., and van Eck, H. J. (2015). A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. *PloS One* 10, e62355. doi: 10.1371/journal.pone.0141940. (vol 8, e62355, 2013).

Vos, P. G., Uitdewilligen, J. G. A. M. L., Voorrips, R. E., Visser, R. G. F., and van Eck, H. J. (2015). Development and analysis of a 20K SNP array for potato (Solanum chilotanum): an insight into the breeding history. *Theor. Appl. Genet.* 128, 2387–2401. doi: 10.1007/s00122-015-2593-y

Watanabe, K. (2015). Potato genetics, genomics, and applications. *Breed. Sci.* 65, 53–68. doi: 10.1270/jsbbs.65.53

Watanabe, K., and Peloquin, S. J. (1989). Occurrence of 2n pollen and ps gene frequencies in cultivated groups and their related wild species in tuber-bearing *Solanums*. *Theor. Appl. Genet.* 78, 329–336. doi: 10.1007/BF00265292

Wickham, H. (2011). *ggplot2. Wiley Interdisciplinary Reviews. Computational Statistics*, Vol. 3. 180–185. doi: 10.1002/wics.147

Willbur, J. F., Schlachter, E. M., Satoh-Cruz, M., Ruth, S., Burek, C., and Byrne, J. (2023). First Report of Geotrichum candidum Causing Rubbery Rot of Potato (Solanum tuberosum) in Michigan. *Plant Dis.* 107, 1633–. doi: 10.1094/PDIS-04-22-0916-PDN

Zhou, Q., Tang, D., Huang, W., Yang, Z., Zhang, Y., Hamilton, J. P., et al. (2020). Haplotype-resolved genome analyses of a heterozygous diploid potato. *Nat. Genet.* 52, 1018–1023. doi: 10.1038/s41588-020-0699-x

Zimnoch-Guzowska, E., and Flis, B. (2021). Over 50 years of potato parental line breeding programme at the plant breeding and acclimatization institute in Poland. *Potato Res.* 64, 743–760. doi: 10.1007/s11540-021-09503-2

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 26 August 2024
DOI 10.3389/fpls.2024.1405314

Check for updates

OPEN ACCESS

EDITED BY
Manuela Nagel,
Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Germany

REVIEWED BY
John Edward Bradshaw,
The James Hutton Institute, United Kingdom
Alfonso H. Del Rio,
University of Wisconsin-Madison, United States

*CORRESPONDENCE
Pawel Chrominski
pawel.chrominski@nordgen.org

RECEIVED 22 March 2024
ACCEPTED 31 July 2024
PUBLISHED 26 August 2024

CITATION
Chrominski P, Carlson-Nilsson U, Palmé A, Kirk HG, Asdal Å and Ansebo L (2024) Genetic markers identify duplicates in Nordic potato collections. *Front. Plant Sci.* 15:1405314. doi: 10.3389/fpls.2024.1405314

COPYRIGHT
© 2024 Chrominski, Carlson-Nilsson, Palmé, Kirk, Asdal and Ansebo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Genetic markers identify duplicates in Nordic potato collections

Pawel Chrominski[1]*, Ulrika Carlson-Nilsson[1], Anna Palmé[1], Hanne Grethe Kirk[2], Åsmund Asdal[1,3] and Lena Ansebo[1,4]

[1]NordGen Plants, Nordic Genetic Resource Center (NordGen), Alnarp, Sweden, [2]Danish Potato Breeding Foundation (LKF-Vandel), Vandel, Denmark, [3]Norwegian Genetic Resource Centre (NGS), Norwegian Institute of Bioeconomy Research (NIBIO), Ås, Norway, [4]Fredriksdal Museums and Gardens, Helsingborg Museum, Helsingborg, Sweden

Introduction: The first small scale cultivation of potatoes in the Nordic countries began roughly 300 years ago, and later became an important staple food in the region. Organized conservation efforts began in the 1980s, and today, potato landraces, improved varieties, and breeding lines are conserved in genebanks at the Nordic Genetic Resource Center (NordGen), Sweden, and the Norwegian Genetic Resource Centre (NGS), Norway, as well as at potato breeding companies across Nordic countries. All these collections house a diverse array of genotypes with local names and local growing histories from the whole region. However, the presence of duplicates, and inconsistent naming has led to confusion.

Methods: In this study, 198 accessions of cultivated potato (*Solanum tuberosum* L.) have been genotyped with 62 microsatellite (SSR) markers. The analyzed accessions came from three collections: 43 accessions from the Danish Potato Breeding Foundation in Vandel (LKF-Vandel), 90 from NordGen and 65 from NGS.

Results and discussion: The genetic analysis revealed 140 unique potato genotypes and 31 groups/clusters of duplicates, most of which contained duplicate pairs and the others three to ten accessions. Several accessions with distinct names were genetically identical or very similar, suggesting historical sharing, and regional distribution of seed potatoes, leading to the emergence of diverse local names. Moreover, many improved varieties from early potato breeding were revealed to have duplicates that have been considered Nordic landraces. Furthermore, potato accessions with identical names but originating from different collections were confirmed to be duplicates. These findings have already influenced management decisions and will further improve management practices for Nordic potato collections. Additionally, this new knowledge will benefit Nordic potato breeding efforts and allow for the dissemination of more accurate information to other users of potato diversity.

KEYWORDS

potato, SSR, microsatellites, genotyping, genebank, duplicates, genetic resources

Introduction

The cultivated potato (*Solanum tuberosum* L.) has its origin in the Andes in South America and was domesticated about 7,000–10,000 years ago (Hawkes, 1990). It was brought to Europe in the second half of the sixteenth century and the first introductions were to Spain, via Canary Islands, and to England (Hawkes and Francisco-Ortega, 1992; 1993). It was however not until the nineteenth century that the potato became a staple of everyday diet for the rapidly increasing European population, and it ended the famine that especially northern Europe was struggling with (McNeill, 1999). The potato landraces related to *S. tuberosum* Chilotanum Group were particularly favored in northern Europe because of their good adaptation to a short growing season with tuberization under long daylengths. The improved varieties grown today are usually related to this group of Chilean potatoes (Spooner et al., 2005; Ríos et al., 2007; Ames and Spooner, 2008).

The potato reached the Nordic countries (Denmark, Finland, Iceland, Norway, and Sweden) later than the rest of Europe. The first records come from Denmark in 1630 (Tolstrup, 2001) and from Sweden in 1658 (Swederus, 1877), but at that time potato was more of a curiosity grown in botanical gardens. The history of growing potatoes for consumption dates back to the 1720s. In 1719–1720, French Huguenots settled in Denmark and started growing potatoes in the town of Fredericia (Kyrre, 1938). In Sweden, Jonas Alströmer brought potatoes from a trip to Western Europe and started growing them on his farm near Alingsås in 1724 (Alströmer, 1777). The potatoes that arrived in Denmark and Sweden originated in France, the Netherlands, Germany, and England, and they were then spread further across the Nordic region (Osvald, 1965; Erjefält, 2001). Until the mid-nineteenth century, these imported potatoes were often propagated by means of botanical seeds which resulted in many new genotypes (Osvald, 1965; Umaerus, 1989). Some of the still existing landraces may potentially have their origin in these efforts (Roer, 1992; Erjefält, 2001).

In the second half of the nineteenth century, commercial improved varieties with better resistance and agronomical traits originating in other European countries and USA, were imported to the Nordic region and became very popular (Osvald, 1965). Potato breeding in the Nordic countries started in Sweden at the beginning of the twentieth century and the first Swedish commercial variety was released in 1921. In the 1930s, successively Finland, Norway and Denmark released their first varieties (Bjor, 2001; Erjefält, 2001; Tolstrup, 2001; Varis, 2001). By the end of the twentieth century Nordic breeding institutions had bred 99 potato varieties in total (Veteläinen, 2001).

With the emergence of the improved potato varieties, the landraces were gradually withdrawn from cultivation. Nevertheless, an inventory of potatoes grown in Sweden in the 1940s identified 118 local and foreign genotypes. Among them a few 200 years old landraces were still present (Hagberth, 1951). According to later estimates (Veteläinen et al., 2005), about 70 potato landraces from all Nordic countries were conserved in genebanks. However, it is suspected that some of the landraces had already been lost by then.

With the development of plant breeding in Europe in the twentieth century, plant genetic resources came into the spotlight. Also, the Nordic countries recognized the need for conservation of genetic resources to prevent their loss and facilitate access. The Nordic Gene Bank (NGB) was founded in 1979 as a joint Nordic institution and became a part of the Nordic Genetic Resource Center (NordGen) in 2008 (Yndgaard and Solberg, 2019). Systematic work on the conservation of potato genetic resources began at NGB in the same year as it was established. The *in vitro* collection of potatoes at NordGen is continuously growing and currently contains 95 accessions, of which 40 are landraces, 46 improved varieties, and 9 breeding lines. The potato collection at NordGen is today considered to include most of the remaining Nordic potato landraces. The focus for acquisition has therefore shifted to collection of older improved potato varieties of Nordic origin and relevance.

Registration and certification of new cultivars revealed an issue of numerous synonyms among potatoes bred in the early twentieth century. In England in 1919, the National Institute of Agricultural Botany (NIAB) was established in Cambridge, and shortly thereafter, the Potato Synonym Committee was formed within its structure. Chaired by Redcliffe Salaman, the committee played a pivotal role in addressing this challenge, accurately describing potato varieties and eliminating the prevalent practice of marketing old and unreliable cultivars under new names (Salaman, 1926).

Potatoes grown in the Nordic countries have been morphologically described in different studies spanning over a hundred years (Øverby, 1929; Hellbo and Esbo, 1942; Hansen and Nissen, 1945; Hagberth, 1951; Veteläinen, 2001). Even today, the assessment of morphological characters is used to identify different potato genotypes, which is one of the basic tasks of genebanks conserving potato genetic resources. To build the Nordic collection, the morphology of more than 600 potato accessions were studied. Many of them did not meet the criteria for conservation at NordGen as they turned out to be non-Nordic genotypes or duplicates of accessions already present in the collection. In most cases, morphological characterization is sufficient to distinguish between different potato genotypes. However, in the case of genotypes that are closely related or infected with viruses, identification based on such an assessment can be difficult (Veteläinen et al., 2005). The morphological characters can also be influenced by environmental factors and, moreover, this method of potato identification is laborious and expensive.

With the development of molecular genotyping techniques, NordGen began to use these methods to identify potato accessions and study the genetic relationship between them. Veteläinen et al. (2005) studied the diversity of 32 Nordic potato landraces by means of amplified fragment length polymorphism (AFLP) markers and morphological traits. In a European project (Hoekstra and Reid, 2014), over four hundred potato accessions from seven collections were genotyped using 12 microsatellite (SSR) markers. Thirty-four accessions from NordGen were included in this study. Old potato clones with questionable identity were fingerprinted to confirm or correct cultivars' names. In a more recent project Selga et al. (2022) studied diversity and population structure of Nordic potato landraces, improved varieties, and breeding clones.

Although NordGen has a mandate from all Nordic countries to conserve potato genetic resources, Nordic potato accessions are also included in other collections, for example the breeding company LKF Vandel in Denmark (now Danespo) and the Norwegian Genetic Resource Center (NGS). Both the NordGen and NGS collections contained a large number of potato genotypes that required identification. In our joint Nordic research project reported herein, we genotyped potato accessions from the three Nordic collections using 62 microsatellite markers (SSR). The main purpose of this study is to confirm the identity of suspected duplicates and identify any additional duplicates in the collections. This information has already been used to remove duplicates when relevant, and to ensure correct naming of the accessions.

Materials and methods

Study material

The potato germplasm used in this study consisted of 198 accessions of cultivated potato *Solanum tuberosum* L. Information regarding the studied accessions, such as name, accession number, collection in which a given accession is conserved, country of origin, clonal type, pedigree, and release date, are included in the Supplementary Table S1. The analyzed accessions came from three collections: 43 accessions from the Danish Potato Breeding Foundation in Vandel (LKF-Vandel) in Denmark; 90 from the Nordic Genetic Resource Center (NordGen) in Sweden; and 65 from Norwegian Genetic Resource Center (NGS), a part of the Norwegian Institute of Bioeconomy Research (NIBIO) in Norway. NordGen is a genebank shared by all Nordic countries and conserves potato genetic resources for the entire region. However, due to stricter phytosanitary regulations governing potato import and export to non-EU countries like Norway, NGS established its own potato collection for domestic use in cooperation with NIBIO Division of Biotechnology and Plant Health. The collection at LKF-Vandel maintains old varieties, and newer parental lines which are considered valuable for future breeding and cultivation of potatoes in Denmark. Although all three potato collections focus on genotypes originating from the Nordic region, foreign accessions may also be accepted if they have been grown on a wider scale in one or more of the Nordic countries or may constitute a valuable resource for potato breeding in the Nordic region. Of the potato accessions used in the study, 165 were of Nordic origin, while 33 accessions were from countries outside this region. The studied accessions contained genotypes of three different types: landraces, improved varieties, and breeding lines. The history of several landraces such as Raudar Islenskar (Gammal Svensk Röd), Jämtlands Vit and Leksands Vit used in this study dates to the 1720s. The oldest improved variety analyzed here was Garnet Chili from 1857 and the newest one was Superb from 2003. Several breeding lines with important traits and potential value for future potato breeding were also included in this study. In total, 49 accessions from NordGen and NGS were recently added into the collections, most with documentation suggesting that they could be landraces or old improved varieties. The potato accessions used in this study were conserved in genebanks in *in vitro* or field collections. Accessions accepted for long-term conservation are usually kept in *in vitro* collections.

Genotyping

DNA was extracted from fresh or freeze-dried leaf tissue from one individual in each accession. The material originating from NordGen's collection was extracted using the protocol described by Doyle and Doyle (1987) and modified as described in Hagenblad et al. (2023). Thereafter the concentration of DNA was measured using a spectrometer (Eppendorf) resulting in concentrations in the range 0,005–0,04 μg/μl.

The samples were genotyped in 2013–2015 with 62 microsatellite markers from a number of different sources (Kawchuk et al., 1996; Provan et al., 1996; Milbourne et al., 1998; Bradshaw et al., 2006; Ghislain et al., 2004; Feingold et al., 2005; Kørup Sørensen et al., 2007; Reid and Kerr, 2007; Szajko et al., 2008; Ghislain et al., 2009), see Supplementary Table S2. The markers were selected in order to have a good coverage of the genome and high power to distinguish among different accessions. To achieve this some 5–6 markers were selected in each chromosome and when there were good alternatives, the most polymorphic marker was selected. The markers have been used at LKF-Vandel to evaluate genetic relationships and potential degree of inbreeding in new crosses for plant breeding purposes.

Each marker was amplified separately by PCR in 96 plates with two sets of 45 accessions, 2 control samples and one blank. Fragment analysis was conducted on an ABI Prism 310 Genetic Analyzer (Applied Biosystems). The microsatellites were combined into 13 sets, within which the different markers were either separated by size or color. Allele sizes were analyzed with the GeneMapper software.

Data analysis

Each amplified PCR product was scored as either present or absent (1 or 0) or if it was not possible to interpret, as missing data. This information was combined into a matrix including data on all fragments amplified at the different microsatellite loci, for all analyzed accessions (raw data available at https://doi.org/10.6084/m9.figshare.26325037.v1). This approach has frequently been used in potato (e.g. Muhinyuza et al., 2015; Duan et al., 2019; Spanoghe et al., 2022) to avoid problems with dosage determination in this generally tetraploid crop. The dataset includes 198 accessions analyzed with 62 microsatellite markers, resulting in 493 different amplified fragments, of these 208 are polymorphic in this data set. For each microsatellite marker, the total number of observed alleles and their size range was noted (Supplementary Table S2).

To identify duplicates, the number of pairwise differences among all accessions were evaluated with the program MEGA6 (Tamura et al., 2013). All allele positions with missing data were removed for each pairwise comparison. To get a general estimate of the amount of divergence between accessions, the number of pairwise differences was divided by the total number of bands (493). Trees were constructed with the program MEGA11 (Tamura et al., 2021) based on the number of differences between accessions. Methods explored were the UPGMA as well as the neighbor joining (NJ) approach and bootstrap with 500 replicates was used to estimate the support in the data of each branch. Also here, missing data was removed in all pairwise comparisons.

Results

Identification of duplicates/clones

The goal of this study was to identify duplicates among the investigated potato accessions. This was done by pairwise comparisons among all the accessions. In total there are 32 pairwise comparisons with no differences, 34 with one difference (representing a difference in about 0.2% of the amplified bands and a similarity of 99.8%), 29 with two differences (0.4%), 26 with three (0.6%), 11 with four (0.8%) and two with five (1%) (Supplementary Table S3). There is a low number of pairs with between five and 40 differences. Most comparisons show between 70–140 differences (Figure 1) and the average is 106 differences. Based on this distribution, and the in general clonal propagation of potato, accessions with five or fewer differences are interpreted as duplicates or very closely related accessions originating from the same clone in the recent past. In the text these will be referred to as duplicates. Using this definition, there are 31 groups of duplicates/clonal clusters (Table 1). Of these, 21 are duplicate pairs, 4 contain three accessions and 3 four accessions and 1 five accessions. The remaining two are larger groups with eight and ten accessions respectively. The groups are also identified and well supported in the dendrograms (Figure 2; Supplementary Figure S1). In total, there are 89 accessions included in the 31 duplicate groups and the remaining 109 accessions had no duplicates among the investigated Nordic collections. This means that we have identified 140 unique genotypes.

Genetic structure

The main aim of this study was not to investigate the overall genetic structure in the Nordic potato collections but to identify duplicates. However, the dendrograms (Figure 2; Supplementary Figure S1) clearly show some additional clustering of accessions, in addition to the duplicates. In some cases, two accessions group together, such as Veto and Gusten Älgarås, Appell and Annika, Bintje and Hroar Dege, Lü 56.220/94 and Tertus, Alex and 92-BSI-702, Röda Krokar and Blå Mandel, Sort IV fra Aukrust and Ingeleivseple. In other cases, there are larger groups involving duplicates as well as other accessions, such as Maris Piper and Gullan Lerdala clustering with C-18, Jaakko and Koto grouping with C-16, Blå potatis (Blå Märta), Purpur and Kefermarkter Zuchtstamm with C-9 and Svart/Blå fra Skjåk with C-4.

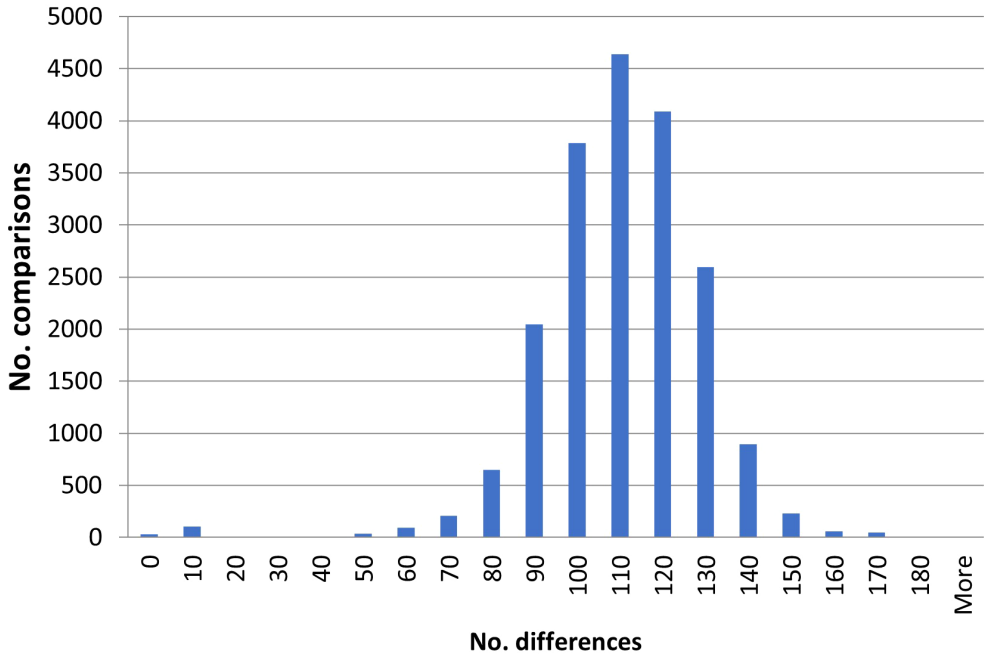


FIGURE 1
A histogram of the number of differences observed in pairwise comparisons of all analyzed accessions.

10.3389/fpls.2024.1405314

TABLE 1 Groups/clusters of accessions that according to the investigated genetic markers are identical, or nearly identical.

Cluster	Accessions	No. accessions	No. observed differences
C-1	Arran Victory, Blå fra Onsøy, Blå Kerrs Pink, Karjalan Musta, Kerrs Pink med blått skall, Luröpotatis, Pålle Kättilstorp, Reyð Epli, Vallgren Falbygden, Weinberger Blaue	10	0 - 3
C-2	Ameriker, Early Puritan, Early Rose NGB, Early Rose NGS, Fljota, Jonsok, Løgumkloster, Tranås	8	0 - 4
C-3	Gullauge Gul variant I, Gullauge Gul variant II, Gullauge Rød, Gullauge LKF, Röd Mandel	5	0 - 2
C-4	Foula Red, Gul Kvæfjording, Rød Kvæfjord, Rød Kvæfjording	4	0 - 4
C-5	Lillhärjåbygget, Mandel variant I (klon 1), Mandel variant II (Ekrann), Mandel variant III (klon 6)	4	1 - 5
C-6	Ringerikspotet variant I, Ringerikspotet variant II, Ringerikspotet variant III, Ringerikspotet NGB	4	0 - 1
C-7	Beate, Kavlås, Olofstorp	3	0 - 2
C-8	Blå Kartoffel (DDSF), Purple Peruvian, Vitelotte	3	0 - 1
C-9	Congo (SWE), Congo LKF, Svartpotet fra Vegårshei	3	0
C-10	King Edward VII, Norska Röda, Raude fra Skjåk	3	2 - 4
C-11	Æggeblomme LKF, Æggeblomme NGB	2	0
C-12	Asparges NGS, Asparges LKF	2	0
C-13	Åspotet NGB, Åspotet NGS	2	3
C-14	Blå Dalsland, Blå Torpar	2	2
C-15	Blaar Islenskar, Svart Valdres	2	0
C-16	Eigenheimer, Granuddspotatis	2	0
C-17	Fakse, Per Larsgården	2	1
C-18	Folva, Ledsjö Gul	2	0
C-19	Gamle Raude fra Aurland, Gammelraude	2	0
C-20	Hårek NGB, Hårek NGS	2	0
C-21	Himalaya, I-1039	2	0
C-22	Jøssing NGB, Jøssing NGS	2	1
C-23	Kiva, Tylva	2	5
C-24	Lange's potet, Tidlig Blå fra Halden	2	2
C-25	Maria, Superb	2	0
C-26	Marius II, Marius NGB	2	1
C-27	Prestkvern NGB, Prestkvern NGS	2	0
C-28	Raudar Islenskar (Gammal Svensk Röd), Små Röda	2	1
C-29	Roslanglänna, Tysk Blå	2	1
C-30	Russet Burbank NGS, Russet Burbank LKF	2	0
C-31	Semlo NGB, Semlo LKF	2	3

Each cluster includes accessions that are separated with five or fewer differences (representing a difference of 0 - 1%) and can be regarded as duplicates or very close relatives originating from the same clone.

Discussion

Many unique genotypes in the Nordic collections

This study clearly identifies 140 unique genotypes that are currently conserved in the Nordic potato collections, including landraces, cultivars from different time periods (release years 1857 - 2003), and a few breeding lines (Supplementary Table S1). The majority of the accessions studied have their origin in one of the Nordic countries or have been cultivated in the region. Together, they represent a substantial amount of variation of relevance for the Nordic region and are important resources for future plant breeding efforts, research projects and, especially for the landraces, a part of our cultural heritage.

The majority of the modern varieties analyzed in this study were genetically unique and had no synonyms. Many of these unique accessions have been identified in previous studies. Veteläinen et al. (2005) confirmed, using AFLP markers, that the 32 potato landraces present in NordGen's collection at that time were unique. Similarly, Selga et al. (2022) analyzed the accepted potato accessions at NordGen with SNP (Single Nucleotide Polymorphism) markers and confirmed the uniqueness of 73 of the 75 accessions that were present in the collection at that time.

Duplicates with identical names

As potato is a clonally propagated crop, the risk of genetic changes after regeneration should be minimal compared to seed propagated crops. Having this in mind, accessions with the same

10.3389/fpls.2024.1405314

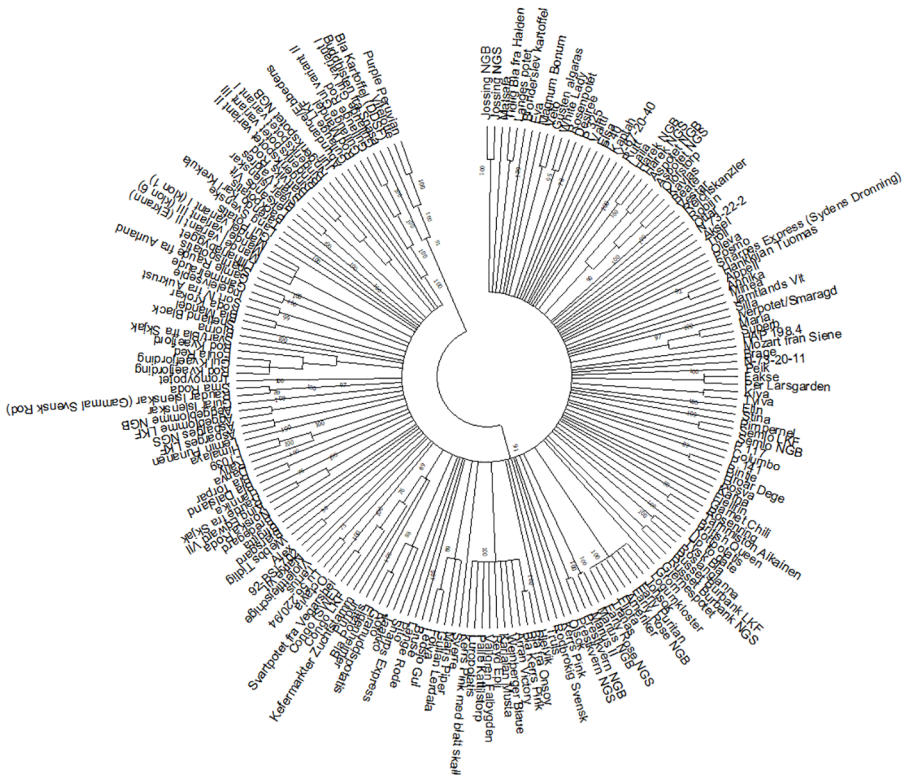


FIGURE 2
Dendrogram constructed with the UPGMA method with branches with less than 75% bootstrap support collapsed. All clusters described in Table 1 are found in clearly delineated and well supported groups in this tree.

name should be identical, or very similar, at different genebanks even if they are collected at different locations or have gone through several generations of regeneration. Nine of the observed clusters contained two accessions with identical names conserved at different genebanks (C-11, C-12, C-13, C-20, C-22, C-26, C-27, C-30, C-31) (Table 1). This result gives strong support to the hypothesis that both genebanks conserve the correct accession in these cases. Only in one situation, two accessions with the same name ended up as genetically different. Sharpes Express from the LKF collection differed from the accession with nearly the same name in the Norwegian collection. This indicates that one of the two collections likely preserves an incorrect accession under the name Sharpes/Sharps Express.

Duplicates with similar names

Sometimes a genebank collection for different, often historical, reasons contain several accessions of the same landrace or improved variety. As every accession is an economic cost for the genebank, it is wise to carefully consider if such duplicates should be kept or not.

In our study we found three clusters, C-3, C-5, and C-6 including duplicate accessions of the landraces/improved varieties Gullauge, Mandel and Ringerikspotet respectively (Table 1). The biggest cluster (C-3) includes, besides Gullauge LKF, three different clones named Gullauge; Gullauge Gul variant I, Gullauge Gul variant II and Gullauge Rød. These three clones are all conserved at NGS in Norway. The results both confirms that the clone from LKF, Denmark, is identical or genetically very close to the two yellow (gul) variants conserved at NGS but also that the two Norwegian yellow variants themselves are genetically very similar, and it might be recommended to only keep one of them in the collection. The origin of Gullauge is unknown but it has a cultivation history in Finland, Iceland, Norway, and Sweden. It can therefore be suspected that Gullauge LKF originally came from Sweden or Norway.

In addition, the morphologically divergent red skinned clone Gullauge Rød cannot be genetically separated from the three yellow skinned ones. This indicates that Gullauge Rød is a skin color mutant with the same origin as the yellow skinned Gullague, which is found both at LKF and NGS, and consistent with what has been shown for other skin color mutants (Kishine et al., 2017). This gene is probably not situated within any of the fragments amplified by the 62 microsatellite markers in this study. We observed the same result in cluster C-4 where two accessions of Kvæfjording, namely Rød Kvæfjording and Gul Kvæfjording (both from NGS), did not differ genetically according to our results, even if they have different skin color.

The fifth accession in cluster C-3 is Röd Mandel, a red skinned clone conserved at NordGen. According to the documentation, this accession came to Sweden (Jämtland, Härjedalen county) from Trondheim, Norway, maybe around 1866–67. At NordGen this

accession has been conserved under the name Röd Mandel, maybe because of morphological similarities with the variety Mandelpotatis. However, the result of our study confirms that the correct variety is Rød Gullauge. As ordinary Gullauge (with yellow skin) is already conserved at NordGen, this finding resulted in the rejection of the accession Röd Mandel/Rød Gullauge.

The second cluster, C-5, includes three clones of Mandelpotatis (Mandel potato/Almond potato). This name is a collective designation for a certain type of potato varieties. The tubers have quite a typical morphology, small and very elongated, slightly flattened, and often somewhat crooked. The skin is thin and smooth, eyes shallow, and the flesh yellow. In addition to the regular Mandel, with white flowers and white skin, accessions with violet flowers and white skin and white flowers and red skin respectively are also available (Hellbo and Esbo, 1942). The variety name Mandelpotatis can be found already in variety lists from the beginning of the 19th century (Osvald, 1965) but in different countries as well as districts this type of potatoes can also be found under other names, often deriving from persons, locations, provinces etc (Hellbo and Esbo, 1942). In Sweden, this old landrace has traditionally been cultivated mainly in the northernmost provinces. In Norway it is called Mandelpotet, Langpotet or Krokeple and it is cultivated in northern Østerdalen and Gudbrandsdalen as a special quality potato. Mandelpotet from Oppdal is sold under a protected geographical indication: Fjellmandel fra Oppdal (NIBIO, 2020b). The variety is also grown in Finland under the name Lapin puikula or Puikula.

In our study, five accessions with the name Mandel were included. Out of them, Röd Mandel was found in cluster C-3 (described above) and is most certainly a misclassification. Three were included in cluster C-5 (Mandel variant I (klon 1), Mandel variant II (Ekrann), Mandel variant III (klon 6)), all from NGS, and one (Blå Mandel) showed to be a unique accession not genetically close to any other accession in the study.

The results show that the three Mandel variants from the NGS collection are duplicates and it could be considered if they all need to be kept in the collection. Blå Mandel is most certainly an old landrace, grown in the Swedish provinces Västerbotten, Norrbotten and Lappland. It is said to be the most cultivated variety in Sweden before the white skinned Mandel was introduced. According to the results in our study, Blå Mandel is genetically different from the three Mandel accessions in the C-5 cluster and instead related to the accession Röda Krokar (mainly cultivated in southern Norrland) (Figure 2; Supplementary Figure S1). Hellbo and Esbo (1942) consider Röda Krokar as a synonym to the white flowering, red skinned Mandel and the synonym name Röd Mandel is also given in Veteläinen (2001). Even if the color of the skin is different between Blå Mandel (dark blue skin) and Röda Krokar/Röd Mandel, they both have white flowers, and our results show that they are genetically close to each other. It is however somewhat unexpected that these two Mandel varieties do not cluster with the accessions in C-5 containing the other Mandel accessions.

There is also another variety included in this cluster (C-5), Lillhärjåbygget from Härjedalen, Sweden, where it has been cultivated by the same family at least since the 1850s. It has white flowers and white skin and a tuber shape like Mandel. Unfortunately, we have no knowledge about the morphology of the three Mandel variants I-III in the cluster and can therefore not tell which morphological group described earlier they belong to. We can, however, determine that according to the results from our study, they are all genetically similar to Lillhärjåbygget. This accession was initially classified as a landrace or an improved variety, but our results suggest that it is a landrace.

In the third cluster of this type, C-6, we find Ringerikspotet. The origin of this local variety is unknown, but it has been cultivated for many years in Ringerike in Norway (Veteläinen, 2001). In our study, three different variants kept at NGS, Norway, and one clone with the same name from NordGen, Sweden, were included. The results show that all four clones are genetically identical or very close and if no morphological differences are observed between the three NGS clones it could be considered if all three clones should be kept for future conservation.

Local material identified as duplicates of improved varieties

Among the 31 clusters of identical or almost identical genotypes (Table 1), seven clusters constitute larger groups or pairs of accessions, where one accession is a known improved cultivar, and the others are its duplicates. These clusters are C-1, C-2, C-7, C-10, C-16, C-17, and C-18.

The largest cluster, C-1, contains ten identical or almost identical accessions. This cluster includes the improved variety Arran Victory and its nine duplicates, of which two were thought to be another improved variety, two were considered landraces, and the status of the remaining five was unknown, although their names also suggested that they could be landraces. Arran Victory was bred in 1912 on the Isle of Arran (Scotland) by the breeder David McKelvie and was released on the market in 1918 (Glendinning, 1983). Cluster C-1 is an example of a common pattern observed in the Nordic countries, where a bred variety grown in geographically distant areas receives local names. This applies especially to landraces and older foreign varieties whose difficult and incomprehensible names were replaced by local names, often quite different from the original ones (Hagberth, 1951; Veteläinen, 2001). The available historical sources do not indicate that the cultivation of the Arran Victory variety had a wider range in Nordic countries, unlike the more popular "Arran varieties" such as Arran Banner, Arran Chief, Arran Comrade, or Arran Consul (Øverby, 1929; Hellbo and Esbo, 1942). Nevertheless, Arran Victory is listed among the varieties that were grown both in Norway and Sweden (Potatis i tiden, 2017; Kvann, 2019; NIBIO, 2020a; Fagforum Potet, 2021a). In this cluster, two duplicates of the Arran Victory variety coming from Norway were named Blå Kerrs Pink and Kerrs Pink med blått skall, which could indicate that these two duplicates were thought to be the well-known Kerrs Pink variety where, because of a mutation, the tuber skin changed its color from light red to blue (Kvann, 2019). The Kerrs Pink variety was also analyzed in this study and turned out to be unique and not closely related to these two accessions. Kerrs Pink was the most important table potato variety in Norway for decades. It

entered the Norwegian list of varieties in 1953 and is still cultivated (Øverby, 1929; Roer, 1981; Fagforum Potet, 2021b). It is also worth mentioning another accession included in the same cluster called Reyð Epli, coming from the Faroe Islands, which is an autonomous territory of Denmark. Geographically, the Faroe Islands are close to Scotland, where Arran Victory was bred, and it seems likely that trade between these areas also included potatoes. Also, the Finnish accession Karjalan Musta in NordGen's collection turned out to be identical to Arran Victory. Karjalan Musta from NordGen was also included in an AEGIS project, where duplicates were identified among 379 accessions from eight European collections using 12 SRR markers (Hoekstra and Reid, 2014). In this study, Karjalan Musta showed the same fingerprint as Arran Victory from the British, Irish, and Canadian collections, as well as with other accessions such as Argyll Blue from the United Kingdom, Bleue D'Auvergne from France, Blaue Österreich from Switzerland, Orkney Blue from the United Kingdom, and Skerry Blue from Germany. The morphological characteristics of these duplicates confirm the genotyping results. Therefore, it seems reasonable to conserve only one of these accessions. However, it is worth preserving the names of the duplicates as synonyms of the name Arran Victory and preserving the histories of the individual accessions.

The second-largest cluster, C-2, includes two identical Early Rose accessions, one from NordGen and the other from NGS. The remaining six accessions in this cluster appeared to be duplicates of the Early Rose variety. One of the duplicates was mistakenly taken for another improved variety, and the names of the others suggested that they could be Nordic landraces. Early Rose is an early potato variety bred in 1867 in the USA by the breeder Albert Bresee. This variety gained popularity in the Nordic countries, especially in Denmark and Sweden (Hellbo and Esbo, 1942; Hansen and Nissen, 1945). Early Rose was included on the official list of varieties in Sweden in 1947, and for many years, it was one of the most cultivated early varieties in this country (Osvald, 1965). One of the duplicates in the C-2 cluster was mistakenly given the name of another American variety, namely Early Puritan, also known as Puritan. Early Puritan was the second most popular early potato variety grown in Sweden in the 1950s, together with Early Rose (Osvald, 1940). Hagberth (1951), summarized the results of an inventory of potato varieties grown in Sweden, which tested 349 samples of potatoes from all over the country, submitted under the name Early Rose. Among them, there were only 141 pure samples with the correct variety name. These samples were sometimes mixed with Early Puritan. A large part of the samples was marked with names that were Swedish synonyms of the name Early Rose, and the most popular of them were Rosen, Amerikansk Rosen, Tidig Rosen, Amerikapotatis, Amerikanare and Amerikansk. Veteläinen (2001) also lists 20 different synonyms under which the Early Rose variety was cultivated in Denmark, Finland, Iceland, and Sweden. Many of them refer to the pink color of the tuber or America where the variety comes from. The word "early" in the languages of individual Nordic countries also often appears in these synonyms. It is not surprising therefore that one of the duplicates in the C-2 cluster is called Ameriker. The names of the other duplicates in this cluster, Løgumkloster, and Tranås, are the geographical names of places in Denmark and Sweden from which these accessions originate. Unexpectedly, the Jonsok variety proved to be nearly identical to Early Rose. Jonsok is an early Norwegian variety bred in 1974 from a cross between Saskia and Ulster Prince (Roer, 1981). Consequently, Jonsok was expected to be a unique variety. The name "Jonsok" refers to a Norwegian holiday celebrated on June 24th, commemorating St. John the Baptist and the beginning of summer. Traditionally, the first potatoes of the season are harvested and consumed during Jonsok and similar celebrations in other Nordic countries. It is possible that the Early Rose was mistakenly named Jonsok in this case. If this finding is confirmed, it will be necessary to acquire the correct Jonsok variety for the collection. The Norwegian accession Fljota is named after the old Norwegian word "fljota", which means "early". Of the Early Rose variety duplicates, only Løgumkloster has been accepted in the NordGen collection, due to certain morphological differences such as the white skin color, which distinguishes it from the Early Rose accession already present in the collection, as well as due to the interesting history of the accession in Denmark.

The C-10 cluster includes the King Edward VII variety from the United Kingdom and its two Norwegian duplicates, Norska Röda and Raude fra Skjåk, classified as landrace/Improved variety and landrace respectively in the genebanks. The King Edward VII variety, with its unknown parents, was discovered by a gardener from Northumberland. The tubers subsequently reached the hands of breeder John Butler, who introduced the variety to the market in 1902 under the name King Edward VII. Salaman (1926) lists 11 synonyms for this variety. The King Edward VII is still sold in grocery stores across the United Kingdom. Even though King Edward VII was included in the Swedish variety lists in 1947, it remains one of the most popular table potato varieties grown in Sweden (Osvald, 1940; Selga et al., 2022). This variety was also cultivated in Denmark and Norway (Hansen and Nissen, 1945; Roer, 1981). It is significant that King Edward VII does not have as many synonyms as other old foreign varieties grown in the Nordic countries. During an inventory of potato varieties grown in Sweden, among 142 samples marked with the name King Edward VII submitted from across the country, only 2 percent had the incorrect name (Hagberth, 1951). The reason for such a small number of synonyms could be that the name King Edward VII has a familiar sound in the Swedish language, and it is easy to identify by the characteristic appearance and properties of the tubers, such as an oval shape, white skin with shallow red eyes and red areas around the eyes, as well as cream-colored flesh that is mealy when cooked.

In cluster C-16, Granuddspotatis turned out to be a synonym for the old Dutch variety Eigenheimer, bred in 1893. In the Nordic countries, Eigenheimer gained popularity mainly in northern Sweden, in the Norrbotten region. Eigenheimer was included on the Swedish list of varieties in 1947. Granuddspotatis received its name from a place called Granudden, where it was grown, located about 60 kilometers west of the town of Jokkmokk in Swedish Lapland. Other synonyms of the Eigenheimer variety are mentioned in the literature, namely Vaikijaur, Unby, and Guldbolar (Hellbo and Esbo, 1942; Osvald, 1965). The synonym Vaikijaur comes from the town of the same name, also located in the vicinity of Jokkmokk.

The varieties bred by Nordic breeding companies usually have fewer or no synonyms (Veteläinen, 2001). The small number or absence of synonyms in these cases can probably be explained by the fact that the names of varieties bred in the Nordic countries are often rooted in the culture of these countries and are therefore easier to remember, eliminating the need for synonyms. Examples of such varieties are found in clusters C-7, C-17, and C-18. The popular Norwegian variety Beate from 1966 had two synonyms: Kavlås and Olofstorp. The Danish variety Fakse from 2000 had the synonym Per Larsgården. Another Danish variety, Folva, from 1989 had the synonym Ledsjö Gul. One of Folva's parents is Maris Piper. These accessions are not identical, but the relationship between them is reflected in the dendrogram where both accessions are part of the same cluster (Figure 2; Supplementary Figure S1).

Improved varieties conserved under incorrect names

In the case of accessions from the C-23 and C-25 clusters, the results were different from what was expected.

In cluster C-23, two Danish improved varieties, Tylva and Kiva, turned out to be identical. Both varieties were bred by LKF-Vandel, Tylva in 1969 and Kiva in 1970. These two varieties have no common parents. In the dendrogram (Figure 2; Supplementary Figure S1), they are both grouped together with another Danish variety, Fakse, which has Kiva in its pedigree (van Berloo et al., 2007). Therefore, it can be concluded that both accessions are likely to be Kiva. The accession called Tylva may have been added to the collection under an incorrect name, or there was a mix-up of accessions in the collection. Selga et al. (2022) analyzed Kiva and Tylva from NordGen's collection using SNP markers. This study also confirmed that both accessions were closely related. Further verification is required, for example, using reference material obtained from a reliable source.

In cluster C-25, two Swedish improved varieties, Maria and Superb, also turned out to be identical. In this case, Superb and Maria group together with the accession HAP 198.4, which is a dihaploid of Maria (Figure 2; Supplementary Figure S1). Therefore, the correct name is probably Maria. This case should also be verified using reference material.

Duplicates among blue potatoes

Among the accessions with blue skin and blue-colored flesh, two separate clusters with duplicates, C-8 and C-9, were formed. The C-8 cluster includes the accessions Blå Kartoffel (DDSF), Purple Peruvian, and Vitelotte. Purple Peruvian was added to the NGS collection with the information that the accession comes from Peru, while Vitelotte was added to the LKF-Vandel collection with the information that it is a French landrace. In an AEGIS study where fingerprinting was conducted using microsatellite markers, three Vitelotte accessions from French, German, and Swiss collections were found to be identical to Unbekannte Schwarze from the Czech Republic, Vitelotte Noire from France, Blaue Veltlin from Switzerland, Blaue Peter from the United Kingdom, and Congo from the British and Irish collections (Hoekstra and Reid, 2014). In cluster C-9, the two Congo accessions from NordGen and LKF-Vandel turned out to be identical, as expected. Additionally, the Norwegian accession Svartpotet fra Vegårshei had the same fingerprint. In the aforementioned AEGIS study, Congo from NordGen was found to be identical to accessions such as All Blue, Blaue Schweden, Blue Congo, Blue Salad, McIntosh Black, Russian Blue, Salad Blue, Shetland Beauty, but differed from accessions called Congo from the United Kingdom and Ireland. In Sweden, Congo was also cultivated under the name Svartpotatis, in Denmark under the name Blå kartoffel, in Finland as Läpimusta, in Norway as Sort fra Aasven and Svartpotet, all names including the local name for black or blue. This variety has never had commercial success in the Nordic countries and was grown more as a curiosity (Veteläinen, 2001).

Landraces cross Nordic borders

Even if not as convenient as seeds, potato tubers have historically been quite easy to bring when moving to other regions. Evidence is also found that old landraces have spread over country borders. When cultivated at a foreign place, it is not inconceivable that by time, the original name changed to a new local one, more or less similar to the original. Duplicate clones with different naming were found in 7 individual clusters in our study (C-4, C-14, C-15, C-19, C-24, C-28, C-29) (Table 1) and some of them could be suspected to be examples of this phenomenon. Once again, it is well justified to consider the utility of conserving several duplicate clones in a collection but also to bear in mind potential unique historical narratives connected with the individual clones.

By looking at cluster C-4 it could, not completely unexpected, be confirmed that the accession with the name Rød Kvæfjord (NordGen) is a duplicate variety to Rød Kvæfjording and Gul Kvæfjording, both conserved at NGS. More surprising is, however, the finding that the accession Foula Red (NGS) is a duplicate to these three accessions. Unfortunately, no information about the origin of the clone of Foula red donated to NGS is available. Foula is the name of an island located in the Shetland archipelago of Scotland and according to the European Cultivated Potato Database Foula Red is an advanced cultivar originating in the United Kingdom. However, no release year has been found, and other sources identify Foula Red as a landrace, so this classification is somewhat dubious. If Foula Red is an old cultivar, one explanation could be that it has come to Norway from the United Kingdom and during the years been spread under the wrong name (Rød Kvæfjording). However, as we have identified both Rød and Gul Kvæfjording as well as Rød Kvæfjord with the same genotype in both the Norwegian and the Nordic collection (NordGen) it seems more likely that these clones are the correct ones and that tubers from Rød Kvæfjording on some occasions have been mislabeled as

Foula Red. This accession could therefore probably be rejected from the NGS collection.

Another example is the accessions Blå Dalsland and Blå Torpar in cluster C-14. Blå Dalsland is an old landrace which historically has been cultivated in western Sweden as well as in Finland. In Norway a red-skinned variant has been cultivated under the name Nordgårdspotet (Veteläinen, 2001). According to the original documentation at NordGen it was not clear if Blå Torpar was a landrace or an improved variety (Supplementary Table S1). It came to NordGen from Gothenburg Botanical Garden which in turn received it from the late Hanna Biljer, a devoted Swedish potato collector. According to her, Blå Torpar is a very old Swedish variety with some connection to the outdoor museum in Skara, Sweden. The varieties Blå Dalsland and Blå Torpar are morphologically similar, and the results from our study confirm that they are indeed duplicates suggesting that Blå Torpar is a landrace with the same origin as Blå Dalsland. This led to the decision to reject Blå Torpar from the collection at NordGen.

Svart Valdres is said to be an established Norwegian synonym name to the variety Bláar Islenskar. The origin of the variety is unknown, but it is known that Bláar Islenskar was among at least three blue skinned varieties that were in cultivation in Iceland in the beginning of the 20th century (Veteläinen, 2001). Svart Valdres is described in a Norwegian publication around the same time (Øverby, 1929). Our study confirms that Svart Valdres from NGS is genetically very close to Bláar Islenskar from NordGen (C-15, Table 1). Having these two accessions with different historical names conserved in two different national collections is justified even if they are duplicates as they are connected with their own unique and well documented narratives.

The variety Raudar Islenskar, with the established synonym name Gammal Svensk Röd, is an old local landrace with a possible origin as long ago as 1720 when Jonas Alströmer imported potatoes to Sweden from the Netherlands and France. It can possibly also be one of the first varieties in Iceland 40 years later and has been cultivated in Finland and Norway as well (Veteläinen, 2001). Små Röda is a landrace or improved variety included in the collection at NordGen. According to information from the provider of the accession, it has been cultivated in the Swedish province Småland by his grandfather born 1864. It is also said to have been grown for generations at another farm in the province to which it originally came from England, maybe as the first foreign potato variety entering Sweden. The partly resembling stories of these two accessions together with the results from our study suggests that Små Röda is a duplicate to Raudar Islenskar (C-28). In addition, morphology also partly supports this conclusion and Små Röda has now been rejected from the collection at NordGen.

Cluster C-29 shows that the landrace Tysk Blå and the landrace or improved variety Roslanglänna are genetically identical or very close. Tysk Blå has a probable origin in Germany and has been grown in Sweden, Finland, and Norway, but unfortunately there are no information about donor or origin available for Roslanglänna.

Also, the two clones Gamle Raude fra Aurland and Gammelraude (C-19) as well as Lange's potet and Tidlig Blå fra Halden (C-24) proved to be duplicates. All four varieties are conserved at NGS.

Breeding lines identical to Himalaya

One of the clusters, C-21, tells an interesting story. The accession named Himalaya conserved at NordGen proved to be a duplicate of the breeding line I-1039 from the LKF collection. This breeding line was included in a breeding program for late blight resistance at the International Potato Center, CIP, Peru, and came to India as a part of CIP's work there. I-1039 is cultivated under the names Khumal red 2 and Khumal Rato 2 and was for example found to be the main improved potato variety cultivated in the Bara district in Nepal in 2012 (Kafle and Shah, 2012). I-1039 was developed from Scottish late blight resistance progenitor clones and an unrevealed progenitor, M136–6. This material has two *Solanum* species, *S. phureja* and *S. edinense* included in its pedigree (van Eck et al., 2017). It is also confirmed that it includes the Ry_{sto} gene from *S. stoloniferum* which makes it highly resistant to Potato Virus Y (PVY) (Herrera et al., 2018). It is possible that an additional species may be included in the genetic heritage of this breeding line because M136–6, the undisclosed ancestor of I-1039, was one of a number of clones with late blight resistance derived from *S. demissum* obtained from Mexico through J.S. Niederhauser of the Rockefeller Foundation in the 1950s (Toxopeus, 1964; Wastie, 1991). I-1039 has resulted in two Danish cultivars, Tivoli, and Liva as well as two other cultivars released in Ecuador (Fripapa 99) and Rwanda (Gikungu) respectively (van Eck et al., 2017).

The origin of Himalaya is somewhat unclear, but the story says that it came to Sweden and Gothenburg Botanical Garden in the beginning of the 1980s with a person that had been traveling in the Himalayas and brought some tubers back home from a potato he found there. The variety did not attract much attention in the botanical garden and the cultivation ceased. However, one of the gardeners brought some tubers to his home garden and continued to grow them there as it proved to be very resistant to potato late blight. In 2007 he gave some tubers back to the botanical garden from which NordGen later received the accession. No morphological comparisons have been performed, but the result from our genetic study indicates that the Indian breeding line conserved at LKF is genetically identical or very close to the accession Himalaya at NordGen. If, and how, the traveler visiting the Himalayas happened to collect tubers from the breeding line I-1039/improved variety Khumal red 2/Khumal Rato 2 we can only speculate about.

The story doesn´t end here. In the same botanical garden, a small seedling was found growing near the Himalaya clone as well as an accession named Kalmar Röd (not included in this study). Both accessions are known for their high resistance to potato late blight. The seedling was rescued by one of the gardeners who took it home and continued to cultivate it for a couple of years under the name Rally. It was shown to possess the same late blight resistance as its suspected mother, Himalaya or Kalmar Röd, and later came back to the botanical garden and was also donated to NordGen. The result from our study confirms that Rally is genetically close to Himalaya even if not as close as I-1039 and Himalaya are to each other. This supports the theory that Rally is a seedling and not a

duplicate to Himalaya. However, it cannot be ruled out that the mother could instead be Kalmar Röd which is morphologically close to both Rally and Himalaya. Today both Rally and Himalaya are conserved at NordGen.

Conclusions

This study used microsatellite markers to identify a large number of unique accessions as well as many duplicates within Nordic potato collections. Accession types, origins, and variety histories were verified, new relationships uncovered and mislabeled accessions were discovered. While the vast majority of studied accessions were unique, a significant number of duplicates were also identified. The largest group was composed of duplicates of old foreign varieties that have received local names and were later conserved under these names. In addition, duplicates of landraces were found in different Nordic countries, suggesting exchange and cultivation of these under local names across the Nordic region. Several identified duplicates likely resulted from human error during collection, handling in genebanks or before they were collected. Accessions with identical or very similar names originating from different collections were, with one exception, found to be duplicates. These findings offer substantial benefits, including reduced collection maintenance costs due to removed duplicates, improved documentation, and facilitated use of the material for plant breeding, research, and education. Furthermore, this study will be valuable for comparing with and corroborating the results from emerging genotyping methods. By providing a well-characterized set of potato accessions with known genetic diversity and clear origin information, the Nordic collections serve as a central reference point for Nordic potato diversity.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found at https://doi.org/10.6084/m9.figshare.26325037.v1.

Author contributions

PC: Writing – original draft, Writing – review & editing. UC-N: Writing – original draft, Writing – review & editing. AP: Writing – original draft, Writing – review & editing. HK: Writing – review & editing. ÅA: Writing – review & editing. LA: Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

The authors would like to thank Tine Allin from LKF-Vandel for extracting DNA and performing all molecular analyses, Alfia Khairullina from NordGen for her assistance in extracting the DNA, and Dag-Ragnar Blystad, Kari Ørstad and Sissel Haugslien from NIBIO for providing leaf material from the Norwegian potato collection for extraction of DNA.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1405314/full#supplementary-material

References

Alströmer, J. (1777). Potaters plantering, Grundad på Rön och Försök. *Kongl. vetenskaps academiens handlingar* 38, 246–265.

Ames, M., and Spooner, D. M. (2008). DNA from herbarium specimens settles a controversy about origins of the European potato. *Am. J. Bot.* 95, 252–257. doi: 10.3732/ajb.95.2.252

Bjor, T. (2001). "The history of the potato in Norway," in *Potatis i Norden. En beskrivning av gamla potatissorter bevarade hos Nordiska Genbanken*, vol. 55 . Ed. M. Veteläinen (CAL-förlaget AB, Varberg, Sweden).

Bradshaw, J. E., Hackett, C. A., Lowe, R., McLean, K., Stewart, H. E., Tierney, I., et al. (2006). Detection of a quantitative trait locus for both foliage and tuber resistance to

late blight Phytophthora infestans (Mont.) de Bary on chromosome 4 of a dihaploid potato clone (Solanum tuberosum subsp. tuberosum). *Theor. Appl. Genet.* 113, 943–951. doi: 10.1007/s00122-006-0353-8

Doyle, J. J., and Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19, 11–15.

Duan, Y., Liu, J., Xu, J., Bian, C., Duan, S., Pang, W., et al. (2019). DNA fingerprinting and genetic diversity analysis with simple sequence repeat markers of 217 potato cultivars (*Solanum tuberosum* L.) in China. *Am. J. Potato Res.* 96, 21–32. doi: 10.1007/s12230-018-9685-6

Erjefält, L. (2001). "The history of the potato in Sweden," in *Potatis i Norden. En beskrivning av gamla potatissorter bevarade hos Nordiska Genbanken*. Ed. M. Veteläinen (CAL-förlaget AB, Varberg, Sweden), 65–66.

Fagforum Potet (2021a). *Arran victory*. Available online at: https://potet.no/produkter/arran-victory (Accessed March 03, 2024).

Fagforum Potet (2021b). *Kerrs Pink*. Available online at: https://potet.no/produkter/kerrs-pink (Accessed March 03, 2024).

Feingold, S., Lloyd, J., Norero, N., Bonierbale, M., and Lorenzen, J. (2005). Mapping and characterization of new EST-derived microsatellites for potato (Solanum tuberosum L.). *Theor. Appl. Genet.* 111, 456–466. doi: 10.1007/s00122-005-2028-2

Ghislain, M., Núñez, J., del Rosario Herrera, M., Pignataro, J., Guzman, F., Bonierbale, M., et al. (2009). Robust and highly informative microsatellite-based genetic identity kit for potato. *Mol. Breed.* 23, 377–388. doi: 10.1007/s11032-008-9240-0

Ghislain, M., Spooner, D. M., Rodríguez, F., Villamón, F., Núñez, J., Vásquez, C., et al. (2004). Selection of highly informative and user-friendly microsatellites (SSRs) for genotyping of cultivated potato. *Theor. Appl. Genet.* 108, 881–890. doi: 10.1007/s00122-003-1494-7

Glendinning, D. R. (1983). Potato introductions and breeding up to the early 20th century. *New Phytol.* 94, 479–505. doi: 10.1111/j.1469-8137.1983.tb03460.x

Hagberth, N. O. (1951). Potatissorterna i sverige. Potato varieties in Sweden. *R. Agr. Col. Sweden Inst. Plant Husb. Publ.* 5, 194.

Hagenblad, J., Aloisi, K., Marum, P., Öhlund, L., Solberg, S. Ø., Asdal, Å., et al. (2023). Limited genetic changes observed during in *situ* and *ex situ* conservation in Nordic populations of red clover (*Trifolium pratense*). *Front. Plant Sci. Sec. Plant Breed.* 14. doi: 10.3389/fpls.2023.1233838

Hansen, H. R., and Nissen, M. (1945). Beskrivelse af kartoffelsorter, dyrket i Danmark. *Tidsskrift Planteavl* 49, 559–663.

Hawkes, J. G. (1990). *The Potato: Evolution, Biodiversity and Genetic Resources* (London: Belhaven Press).

Hawkes, J. G., and Francisco-Ortega, J. (1992). The potato in Spain during the late 16th century. *Econ. Bot.* 46, 86–97. doi: 10.1007/BF02985257

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

Hellbo, E., and Esbo, H. (1942). *Våra potatissorter. Systematisk Behandling* (Stockholm: Lantbruksförbundets tidskriftsaktiebolag).

Herrera, M. D. R., Vidalon, L. J., Montenegro, J. D., Riccio, C., Guzman, F., Bartolini, I., et al. (2018). Molecular and genetic characterization of the Ry_{adg} locus on chromosome XI from Andigena potatoes conferring extreme resistance to potato virus Y. *Theor. Appl. Genet.* 131, 1925–1938. doi: 10.1007/s00122-018-3123-5

Hoekstra, R., and Reid, A. (2014). *Final report of the AEGIS project: identification of old potato clones having unreliable variety names by means of fingerprinting using microsatellite (SSR) markers to assist in setting up the AEGIS collection for potato cultivars*. Available online at: https://www.ecpgr.cgiar.org/fileadmin/templates/ecpgr.org/upload/PROJECT_REPORTS/Potato_project_AEGIS_Grant_Final_report.pdf (Accessed March 03, 2024).

Kafle, B., and Shah, P. (2012). Adoption of improved potato varieties in Nepal: A case of Bara district. *J. Agric. Sci. – Sri Lanka* 7, 14–22. doi: 10.4038/jas.v7i1.4063

Kawchuk, L. M., Lynch, D. R., Thomas, J., Penner, B., Sillito, D., and Kulcsar, F. (1996). Characterization of Solanum tuberosum simple sequence repeats and application to potato cultivar identification. *Am. Potato J.* 73, 325–335. doi: 10.1007/BF02849164

Kishine, M., Tsutsumi, K., and Kitta, K. (2017). A set of tetra-nucleotide core motif SSR markers for efficient identification of potato (*Solanum tuberosum*) cultivars. *Breed Sci.* 67, 544–547. doi: 10.1270/jsbbs.17066

Kørup Sørensen, K., Kirk, H. G., and Emmersen, J. (2007). "Microsatellites used for estimation of relationship between potato breeding clones," in *Plant Biotech Denmark. Annual meeting 2007: Faculty of Life Sciences, University of Copenhagen*, (Copenhagen: PBD Plant Biotech Denmark).

Kvann (2019). *Historiske potetsorter dyrket i Norge*. Available online at: https://kvann.no/wp-content/uploads/2019/11/20190316_Sortsbeskrivelse_poteter_utsendelse_vaaren2019.pdf (Accessed March 03, 2024).

Kyrre, H. (1938). *Kartoffelens krønike: en kulturhistorisk studie* (København: Gad).

McNeill, W. H. (1999). How the potato changed the world's history. *Soc Res.* 66, 67–83.

Milbourne, D., Meyer, R., Collins, A., Ramsay, A., Gebhardt, C., and Waugh, R. (1998). Isolation, characterization and mapping of simple sequence repeat loci in potato. *Mol. Gen. Genet.* 259, 233–245. doi: 10.1007/s004380050809

Muhinyuza, J. B., Shimelis, H., Melis, R., Sibiya, J., Gahakwa, D., and Nzaramba, M. (2015). Assessment of genetic relationship among promising potato genotypes in Rwanda using SSR markers. *Aust. J. Crop Sci.* 9, 696–700.

NIBIO (2020a). *Potet. Arran victory*. Available online at: https://www.nibio.no/tema/mat/plantegenetiske-ressurser/nytteplanter-i-norge/jordbruksplanter/potet/arran-victory (Accessed March 03, 2024).

NIBIO (2020b). *Potet. Mandelpotet*. Available online at: https://nibio.no/tema/mat/plantegenetiske-ressurser/nytteplanter-i-norge/jordbruksplanter/potet/mandelpotet (Accessed March 03, 2024).

Osvald, H. (1940). *Potatis* (Stockholm: Kooperativa Förbundets bokförlag).

Osvald, H. (1965). *Potatisen: odlingshistoria och användning* (Uppsala: Almqvist och Wiksells boktryckeri aktiebolag).

Øverby, G. (1929). *Morfologiske sortskarakterer hos potet. Bidrag til beskrivelse av 98 potetsorter* Vol. 9 (Ås: Meldinger fra Norges Landbrukshøiskole), 429–527.

Potatis i tiden (2017). *Sorter*. Available online at: https://potatisitiden.se/sorter/ (Accessed March 03, 2024).

Provan, J., Powell, W., and Waugh, R. (1996). Microsatellite analysis of relationships within cultivated potato (Solanum tuberosum). *Theor. Appl. Genet.* 92, 1078–1084. doi: 10.1007/BF00224052

Reid, A., and Kerr, E. M. (2007). A rapid simple sequence repeat (SSR)-based identification method for potato cultivars. *Plant Genet. Resour.* 5, 7–13. doi: 10.1017/S1479262107192133

Ríos, D., Ghislain, M., Rodríguez, F., and Spooner, D. M. (2007). What is the origin of the European potato? Evidence from Canary Island landraces. *Crop Sci.* 47, 1270–1280. doi: 10.2135/CROPSCI2006.05.0336

Roer, L. (1981). *Beskrivelse av potetsorter* (Ås: Statens planteavlsråd).

Roer, L. (1992). *Nordisk genbank. Arbeidet med sikring av verdifullt potetmateriale. Fagnytt hagebruk 12* (Ås: Statens fagtjenste for landbruket).

Salaman, R. N. (1926). *Potato varieties* (Cambridge: Cambridge University Press).

Selga, C., Chrominski, P., Carlson-Nilsson, U., Andersson, M., Chawade, A., and Ortiz, R. (2022). Diversity and population structure of Nordic potato cultivars and breeding clones. *BMC Plant Biol.* 22, 350. doi: 10.1186/s12870-022-03726-2

Spanoghe, M., Marique, T., Nirsha, A., Esnault, F., and Lanterbecq, D. (2022). Genetic diversity trends in the cultivated potato: A spatiotemporal overview. *Biology* 11, 604. doi: 10.3390/biology11040604

Spooner, D. M., Núñez, J., Rodriguez, F., Naik, P. S., and Ghislain, M. (2005). Nuclear and chloroplast DNA reassessment of the origin of Indian potato varieties and its implications for the origin of the early European potato. *Theor. Appl. Genet.* 110, 1020–1026. doi: 10.1007/s00122-004-1917-0

Swederus, M. B. (1877). *Botaniska trädgården i Upsala 1655–1807: ett bidrag till den svenska naturforskningens historia* (Falun: Falu Boktryckeri-Aktiebolag).

Szajko, K., Chrzanowska, M., Witek, K., Strzelczyk-Żyta, D., Zagórska, H., Gebhardt, C., et al. (2008). The novel gene *Ny-1* on potato chromosome IX confers hypersensitive resistance to *Potato virus Y* and is an alternative to *Ry* genes in potato breeding for PVY resistance. *Theor. Appl. Genet.* 116, 297–303. doi: 10.1007/s00122-007-0667-1

Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38, 3022–3027. doi: 10.1093/molbev/msab120

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197

Tolstrup, K. (2001). "The potato in Denmark," in *Potatis i Norden. En beskrivning av gamla Potatissorter Bevarade hos Nordiska Genbanken*. Ed. M. Veteläinen (CAL-förlaget AB, Varberg, Sweden), 32–35.

Toxopeus, H. J. (1964). Treasure-digging for blight resistance in potatoes. *Euphytica* 13, 206–222. doi: 10.1007/BF00023099

Umaerus, M. (1989). "True potato seed," in *Reprinted in 1989 from: EAPR Proceedings, 10th Triennial Conference of the European Association for Potato Research*, Aalborg, Denmark, July 26-31, 1987 (Lima: International Potato Center).

van Berloo, R., Hutten, R. C. B., van Eck, H. J., and Visser, R. G. F. (2007). An online potato pedigree database resource. *Potato Res.* 50, 45–57. doi: 10.1007/s11540-007-9028-3

van Eck, H. J., Vos, P. G., Valkonen, J. P. T., Uitdewilligen, J. G. M. L., Lensing, H., de Vetten, N., et al. (2017). Graphical genotyping as a method to map $Ny_{(o,n)\ sto}$ and *Gpa5* using a reference panel of tetraploid potato cultivars. *Theor. Appl. Genet.* 130, 515–528. doi: 10.1007/s00122-016-2831-y

Varis, E. (2001). "The potato in Finland from the past to the present," in *Potatis i Norden. En beskrivning av gamla potatissorter bevarade hos Nordiska Genbanken*. Ed. M. Veteläinen (CAL-förlaget AB, Varberg, Sweden), 44–45.

Veteläinen, M. (Ed.) (2001). *Potatis i Norden. En beskrivning av gamla potatissorter bevarade hos Nordiska Genbanken* (Varberg: CAL-förlaget AB).

Veteläinen, M., Gammelgård, E., and Valkonen, J. P. T. (2005). Diversity of Nordic landrace potatoes (*Solanum tuberosum* L.) revealed by AFLPs and morphological characters. *Genet. Resour. Crop Evol.* 52, 999–1010. doi: 10.1007/s10722-003-6129-y

Wastie, R. L. (1991). "Breeding for resistance," in Phytophthora infestans, *the cause of Late Blight of Potato. Advances in Plant Pathology*, vol. 7 . Eds. D. S. Ingram and D. S. Wiliams (Academic Press Ltd., London), 193–224.

Yndgaard, F., and Solberg, S. Ø. (2019). *40 years of Nordic Collaboration in Plant Genetic Resources* (Alnarp: Nordic Genetic Resources Center).

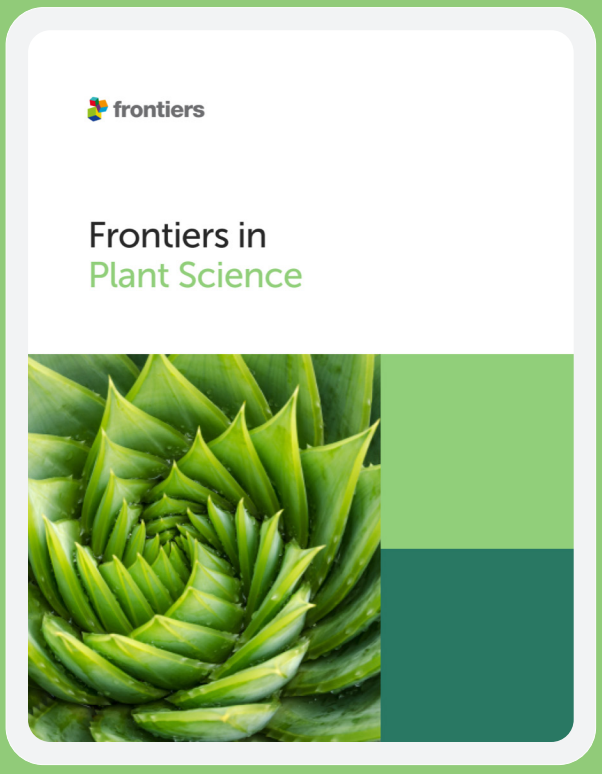
Frontiers in
Plant Science

Cultivates the science of plant biology and its applications

The most cited plant science journal, which advances our understanding of plant biology for sustainable food security, functional ecosystems and human health.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact