A genomic variation map provides insights into potato evolution and key agronomic traits

Molecular Plant

Lian, Qun; Zhang, Yingying; Zhang, Jinzhe; Peng, Zhen; Wang, Weilun et al

https://doi.org/10.1016/j.molp.2025.01.016

Please cite this article in press as: Lian et al., A genomic variation map provides insights into potato evolution and key agronomic traits, Molecular Plant (2025), https://doi.org/10.1016/j.molp.2025.01.016

Molecular Plant
Research article

CellPress
Partner Journal

A genomic variation map provides insights into potato evolution and key agronomic traits

Qun Lian[1,2,3,10], Yingying Zhang[1,10], Jinzhe Zhang[4,10], Zhen Peng[5,10], Weilun Wang[1], Miru Du[1], Hongbo Li[2,6], Xinyan Zhang[2], Lin Cheng[2], Ran Du[2], Zijian Zhou[1], Zhenqiang Yang[1], Guohui Xin[1], Yuanyuan Pu[1], Zhiwen Feng[1], Qian Wu[1], Guochao Xuanyuan[1], Shunbuer Bai[1], Rong Hu[1], Sónia Negrão[3], Glenn J. Bryan[7], Christian W.B. Bachem[6], Yongfeng Zhou[2], Ruofang Zhang[1], Yi Shang[8], Sanwen Huang[2,*], Tao Lin[9,*] and Jianjian Qi[1,*]

[1]Inner Mongolia Potato Engineering and Technology Research Center, Key Laboratory of Herbage and Endemic Crop Biology, Ministry of Education, School of Life Sciences, Inner Mongolia University, Hohhot 010021, China

[2]National Key Laboratory of Tropical Crop Breeding, Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China

[3]School of Biology and Environmental Science, University College Dublin, Dublin, Ireland

[4]State Key Laboratory of Vegetable Biobreeding, Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing 100081, China

[5]College of Plant Science and Technology, Beijing University of Agriculture, Beijing 102206, China

[6]Plant Breeding, Wageningen University & Research, P.O. Box 386, 6700 AJ Wageningen, the Netherlands

[7]Cell and Molecular Sciences, The James Hutton Institute, Invergowrie, Dundee DD2 5DA, UK

[8]Yunnan Key Laboratory of Potato Biology, CAAS-YNNU-YINMORE Joint Academy of Potato Sciences, Yunnan Normal University, Kunming, China

[9]College of Horticulture, China Agricultural University, Beijing 100193, China

[10]These authors contributed equally to this article.

***Correspondence: Sanwen Huang (huangsanwen@caas.cn), Tao Lin (lintao35@cau.edu.cn), Jianjian Qi (qijj@imu.edu.cn)**

https://doi.org/10.1016/j.molp.2025.01.016

ABSTRACT

Hybrid potato breeding based on diploid inbred lines is transforming the way of genetic improvement of this staple food crop, which requires a deep understanding of potato domestication and differentiation. In the present study, we resequenced 314 diploid wild and landrace accessions to generate a variome map of 47,203,407 variants. Using the variome map, we discovered the reshaping of tuber transcriptome during potato domestication, characterized genome-wide differentiation between landrace groups Stenotomum and Phureja. We identified a jasmonic acid biosynthetic gene possibly affecting the tuber dormancy period. Genome-wide association studies revealed a *UDP-glycosyltransferase* gene for the biosynthesis of anti-nutritional steroidal glycoalkaloids (SGAs), and a *Dehydration Responsive Element Binding* (*DREB*) transcription factor conferring increased average tuber weight. In addition, genome similarity and group-specific SNP analyses indicated that tetraploid potatoes originated from the diploid *Solanum tuberosum* group Stenotomum. These findings shed light on the evolutionary trajectory of potato domestication and improvement, providing a solid foundation for advancing hybrid potato-breeding practices.

Keywords: potato, domestication, differentiation, tuber dormancy, steroidal glycoalkaloids, average tuber weight

Lian Q., Zhang Y., Zhang J., Peng Z., Wang W., Du M., Li H., Zhang X., Cheng L., Du R., Zhou Z., Yang Z., Xin G., Pu Y., Feng Z., Wu Q., Xuanyuan G., Bai S., Hu R., Negrão S., Bryan G.J., Bachem C.W.B., Zhou Y., Zhang R., Shang Y., Huang S., Lin T., and Qi J. (2025). A genomic variation map provides insights into potato evolution and key agronomic traits. Mol. Plant. **18**, 1–20.

INTRODUCTION

Potato (*Solanum tuberosum*) is the most important non-grain crop worldwide, serving as a staple food for 1.3 billion people (Navarre and Pavek, 2014; Dudenhoefer, 2019). Although commercial potato varieties are vegetatively propagated autotetraploids, potato was first domesticated as a diploid plant from members of the wild ‘‘northern brevicaule complex’’ near Lake Titicaca, Peru (Spooner et al., 2005). To address the many issues including inbreeding depression (Zhang et al., 2019b), a narrow genetic background (Grun, 1990; Ames and Spooner, 2008), and tetrasomic inheritance, which greatly hinder the genetic improvement of tetraploid cultivars (Hirsch et al., 2013; Watanabe, 2015), diploid potato-breeding projects

Please cite this article in press as: Lian et al., A genomic variation map provides insights into potato evolution and key agronomic traits, Molecular Plant (2025), https://doi.org/10.1016/j.molp.2025.01.016

have begun to reinvent potato from a tuber-based clonally propagated tetraploid to an inbred line-based diploid crop that can be reproduced via true seeds (Zhang et al., 2021). Notably, ~70% of the 111 tuber-bearing *Solanum* species are diploid (Spooner et al., 2014), representing a rich candidate gene pool for potato improvement that remains largely uncharacterized and underutilized in breeding schemes. Therefore, a comprehensive understanding of the diversity, domestication, and differentiation of diploid potato germplasm is needed to allow breeders to take full advantage of these diverse germplasm resources.

Recent potato genome resequencing studies have expanded our understanding of modern potato demography and metabolic pathways/genes under domestication and have yielded markers for further breeding (Spooner et al., 2005; Hardigan et al., 2017; Li et al., 2018; Tang et al., 2022; Zhang et al., 2022; Bozan et al., 2023). Phylogenetic reconstructions in these studies supported *Solanum candolleanum* as the progenitor of cultivated potatoes. Still, the limited number of landrace accessions sampled (10–33) and the mixture of other wild species used for selective sweep analysis hampered the identification of genes underlying key domestication traits (Hardigan et al., 2017; Li et al., 2018; Bozan et al., 2023). Moreover, two prominent diploid potato landrace groups, *S. tuberosum* group Stenotomum (STN) and *S. tuberosum* group Phureja (PHU), have previously been identified (Huamán and Spooner, 2002). They differ markedly in tuber dormancy period, with PHU selected from STN for short dormancy (Hawkes, 1990). However, the genetic basis of differentiation between these two landrace subgroups and how they contributed to the formation of tetraploid genomes remains unclear (Sukhotu and Hosaka, 2006; Rodríguez et al., 2010).

In this study, we resequenced the genomes of 314 diploid potato accessions representing all four major clades of the tuber-bearing *Solanum* section *Petota* (Spooner and Castillo, 1997; Cai et al., 2012; Spooner et al., 2014) and performed population genomic and quantitative genetic analyses. We provided new insights into potato domestication and differentiation, and identified candidate genes associated with dormancy period, tuber steroidal glycoalkaloids (SGAs), and average tuber weight. These findings will facilitate the utilization of diploid wild and landrace accessions in potato improvement initiatives, particularly advancing diploid hybrid breeding strategies.

RESULTS

Genomic variation of cultivated and wild potatoes

To construct a comprehensive genomic variation map, we studied 314 diverse diploid potato accessions representing 47 wild species and diploid *S. tuberosum* landrace subgroups from the United States Potato Genebank and International Potato Center (Supplemental Table 1). Resequencing of these accessions yielded 3.82 trillion base pairs (bp) of sequence data, with an average read depth of ~17.67-fold and a coverage of 85.00% of the reference genome DM1-3 516 R44 v6.1 (hereafter referred to as DM v6.1) (Hardigan et al., 2016; Pham et al., 2020). In total, we detected 35,934,532 single-nucleotide polymorphisms (SNPs) and 11,268,875 small insertions and deletions (InDels, shorter than 5 bp). Of these, 1,945,099 (5.41%) SNPs and 218,200 (1.94%) InDels were located within genic regions, with 950,904 (2.64%) non-synonymous SNPs in 31,822 genes. Among these variants, 56,943 showed potentially large effects, including start codon changes, frameshift mutations, and truncated or elongated transcripts. This comprehensive variome dataset of 47,203,407 variants will enable in-depth population genomic analysis and provide a new resource for potato biology and breeding.

Evolutionary genomics of diploid potato

We investigated the phylogenetic relationships of the 314 potato accessions using 126,119 4-fold degenerate (4d) SNPs. These samples were assigned into four major clades: clade 0 (non-tuber-bearing wild species), clade 1+2 (North and Central American wild species), clade 3 (Ecuador and Northern Peru wild species), and clade 4 (South American wild and cultivated species). Clade 4 could be further divided into clade 4 north (Peru and northward) and clade 4 south (Bolivia and southward) (Figure 1A, Supplemental Figures 1–3, and Supplemental Table 1), consistent with the previously reported classification (Spooner et al., 2014; Li et al., 2018; Bozan et al., 2023). This grouping was also supported by principal-component analysis (PCA) and identity-by-descent (IBD) analysis (Figure 1B and Supplemental Figure 4). Our data also supports the notion that all diploid landrace potatoes were monophyletic, with *S. candolleanum* (CND) being their most immediate wild progenitor, reinforcing the discoveries of recent potato genetic studies (Hardigan et al., 2017; Li et al., 2018). The ancestral role of CND was also supported by its scattered PCA diagram (Figure 1B), higher nucleotide diversity (Table 1, $\pi_{CND} = 3.05 \times 10^{-3}$ and $\pi_{landraces} = 2.77 \times 10^{-3}$), more private SNPs (Table 1), and lower extent of linkage disequilibrium (LD) compared to landraces (CND = 1.06 Kb and landraces = 2.84 Kb), highlighting bottlenecks during potato domestication (Figure 1C).

Our population-structure analyses resolved the landraces into two major groups: STN and PHU (Figure 1A, Supplemental Table 1, and Supplemental Figure 4). Although phylogenic topology indicated the 115 STN accessions could be further divided into four subgroups showing an indistinctive pattern of geographic distribution, population fixation statistics (F_{ST}) and key agronomic traits (average tuber weight, yield, starch content, and dormancy period) in these subgroups suggested they were not significantly differentiated as compared to PHU (Supplemental Figures 5 and 6). Therefore, we treated them as a coherent STN group for subsequent analysis. We found that the 45 PHU accessions were monophyletic and possessed an extremely short tuber dormancy period (Supplemental Figure 7), implying a single formation event for this specific group.

Reshaping of the tuber transcriptome during potato domestication

Potatoes are highly valued for their starch-rich underground tubers, whose consumption dates to the Neolithic period, predating potato domestication (Ugent et al., 1987; Louderbacka and Pavlik, 2017). Our phenotypic evaluation showed that landrace potatoes had significantly increased average tuber weight, higher yield, and elevated tuber starch content compared to the ancestral CND (Supplemental Figure 8), consistent with the

Please cite this article in press as: Lian et al., A genomic variation map provides insights into potato evolution and key agronomic traits, Molecular Plant (2025), https://doi.org/10.1016/j.molp.2025.01.016

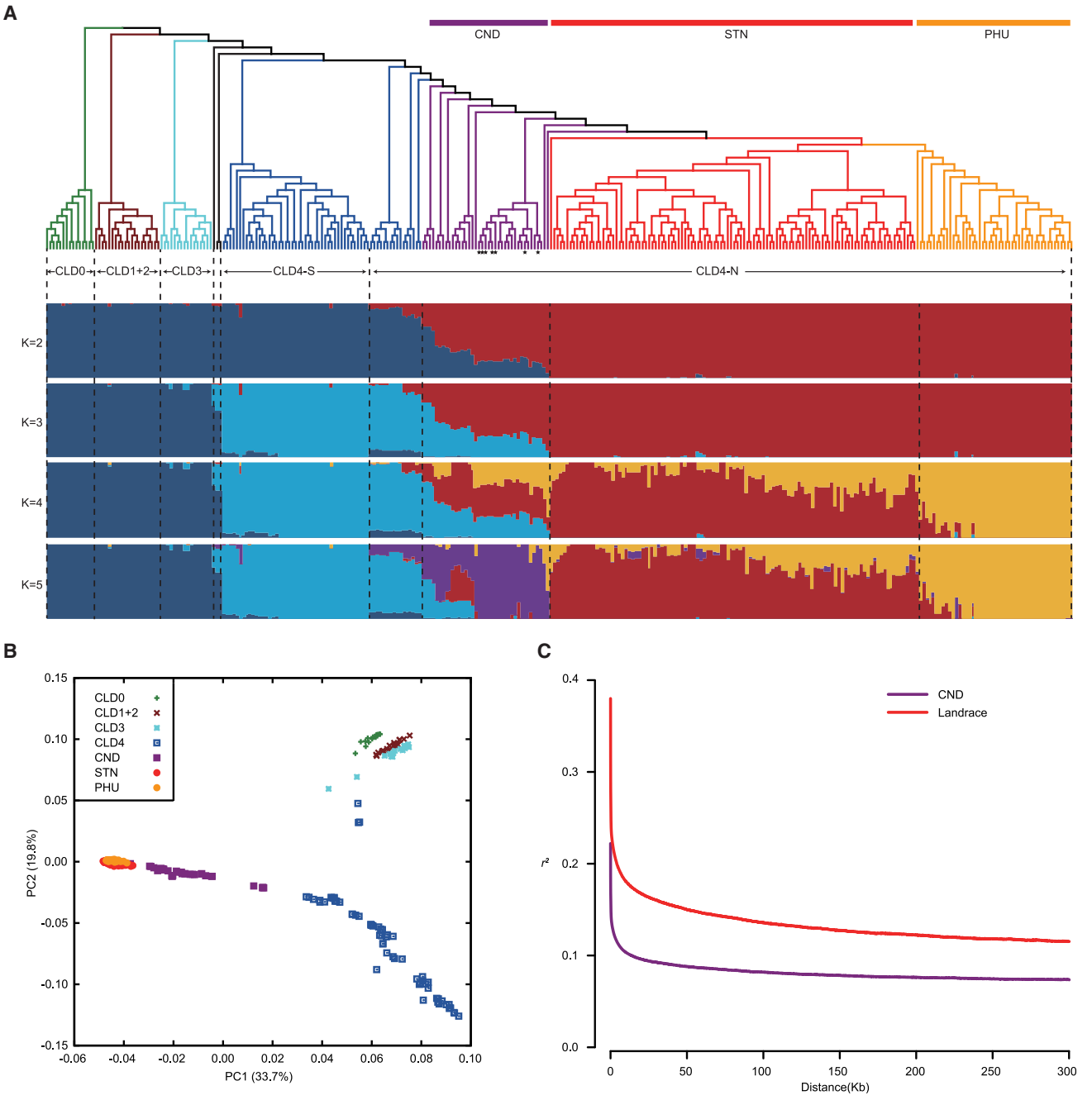

Figure 1. Population genomics of diploid potato.

(A) Upper panel: Maximum-likelihood (ML) tree of 314 accessions inferred from 126,119 4-fold degenerate nuclear SNPs genome-wide. CLD0, non-tuber-bearing *Solanum etuberosum*, *Solanum palustre*, and *Solanum fernandezianum*; CLD1+2, North and Central America wild species; CLD3, Ecuador and Northern Peru wild species; CLD4, South American wild and cultivated species. CLD4 was further separated into CLD4 south (CLD4-S) and CLD4 north (CLD4-N). CND (*Solanum candolleanum*), STN (*Solanum tuberosum* group Stenotomum), and PHU (*Solanum tuberosum* group Phureja) were grouped into CLD4-N. Asterisks indicate the phylogenetic positions of the seven accessions with insignificant levels of α-chaconine and α-solanine. The bars indicate CND (purple), STN (red), and PHU (orange) accessions. Lower panel: Model-based clustering analysis with different numbers of clusters (K = 2, 3, 4, and 5). The y-axis quantifies cluster membership, and the x-axis lists each germplasm accession. The orders and positions of these accessions on the x-axis are consistent with those for the ML tree.

(B) Principal-component analysis (PCA) with the first two eigenvectors.

(C) Decay of LD, measured by the squared correlations of allele frequency (r^2) against the distance between polymorphic sites in the CND (purple) and landrace group (red).

Please cite this article in press as: Lian et al., A genomic variation map provides insights into potato evolution and key agronomic traits, Molecular Plant (2025), https://doi.org/10.1016/j.molp.2025.01.016

Groups	CND	Landraces	
		STN	PHU
π (10^{-3})[a]	3.05	2.82	2.10
Private SNPs[b]	727,148	418,779	34,309
Comparison	CND *vs*. landrace	STN *vs*. PHU	
F_{ST}[c]	0.15	0.11	

Table 1. Population genetics statistics of *S. candolleanum* (CND) and diploid landraces.
[a]π, nucleotide diversity of each group.
[b]Private SNPs are variations specific to each group.
[c]F_{ST} values were calculated for each 100-Kb sliding window with a step size of 10 Kb between two given groups.

observed domestication syndrome of other crops (Doebley et al., 2006). To gain insight into potential selective signals during potato domestication, we performed SweeD analysis (Pham et al., 2020) and calculated the reduction of nucleotide diversity between *S. candolleanum* and the landrace group (STN and PHU). We then proposed the 184 overlapped regions harboring 2899 genes that covered 7.99% (59.19 Mb) of the DM v6.1 reference genome as candidate domestication sweeps (Figure 2A, Supplemental Tables 2 and 3). Some of the identified domestication sweeps overlapped with previously reported quantitative trait locus (QTL) regions or genes for average tuber weight, yield, and starch metabolism (Zrenner et al., 1996; Navarro et al., 2011; Schreiber et al., 2014; Endelman and Jansky, 2016; Śliwka et al., 2016; Braun et al., 2017; Hardigan et al., 2017; Van Harsselaar et al., 2017; Marand et al., 2019; Sołtys-Kalina et al., 2020), suggesting the selection of these loci during potato domestication (Figure 2A).

To gain insights into the transcriptional changes due to potato domestication, we conducted RNA sequencing (RNA-seq) for bulking-stage tubers in nine ancestral CND and 15 landrace accessions (Figure 2B) and detected 23,509 genes that were expressed in at least one of the 24 accessions (Supplemental Figure 9). PCA with fragments per Kb of transcript per million mapped reads (FPKM) of genes indicated different expression patterns between the CND and landrace groups (Figure 2C). Moreover, of the 9,225 differentially expressed genes (DEGs) between CND and landrace, 6,247 (67.7%) showed higher expression in the landrace group. Further Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis of these upregulated genes showed enrichment in "starch and sucrose metabolism", "photosynthesis", and "basal transcription factors" (Supplemental Table 4), while the significantly enriched KEGG terms in the downregulated genes were related to "biosynthesis of secondary metabolites", "phenylpropanoid biosynthesis", and "steroid biosynthesis", all of which may help explain the improved tuber traits such as yield and quality in the landrace group (Supplemental Table 5). Moreover, we built weighted gene co-expression networks and found that gene connectivity and expression diversity of the CND group was, on average, higher than that of the landrace group (Figures 2D and 2E). This indicates the reshaping of tuber expression networks during potato domestication, as reported in rice, maize, cotton, soybean, olive, and tomato (Gross and Strasburg, 2010; Swanson-Wagner et al., 2012; Koenig et al., 2013; Lu et al., 2016; Gros-Balthazard et al., 2019; Liu et al., 2019).

Moreover, 871 (30.04%) of the 2,899 genes in domestication sweeps were differentially expressed between the CND and landrace groups. Among them, 119 DEGs showed over-represented distribution within a ~5.47-Mb (72.41–77.88 Mb) interval on chromosome 1 compared to genome-wide ($p = 0.013$, hypergeometric test) (Figures 2A and 2F), which is an overlapped QTL for both average tuber weight and tuber starch content (Śliwka et al., 2016; Braun et al., 2017). These genes were clustered into two expression modules and included those encoding transcription factors (MYB, ARF, WD40, bHLH, and MADS-box), cyclin, and ubiquitin-like superfamily proteins, which could serve as high-priority candidates for further analysis (Figure 2G and Supplemental Table 6).

Differentiation of tuber dormancy period between two landrace groups STN and PHU

The most prominent trait of differentiation between landrace groups STN and PHU is the tuber dormancy period: STN produced dormant tubers that usually last 10 weeks before sprouting, whereas the PHU tubers were almost sprouting at harvest (Supplemental Figure 7). To uncover the genetic basis of differentiation between STN and PHU, we identified 174 highly divergent genomic regions containing 2,292 genes that span 59.62 Mb of the reference genome (Figure 3A, Supplemental Tables 7 and 8). Ten previously reported QTLs for tuber dormancy release or apical dominance release overlapped with these differentiation regions (Bisognin et al., 2018) (Figure 3A), indicating differential selection between STN and PHU on this trait.

We established an F_1 segregating population through a cross between a long-dormant STN (PG6027) and a short-dormant PHU (additional accession P05) to elucidate the genetic factors influencing the tuber dormancy period. This endeavor led to the identification of a major QTL (*qDR3.1*) located on chromosome 3, spanning approximately 48.91–49.88 Mb (Figure 3B and Supplemental Figure 10). Remarkably, this genomic region corresponds with a previously reported locus (*ADR3.2*) known to exert negative control over the number of days to apical dominance release (ADR) (Bisognin et al., 2018). Notably, it also encompasses a highly differential region between STN and PHU (region 27, Chr03: 49.61–49.80 Mb) (Figure 3C). Within these overlapped intervals, a total of 17 annotated genes were identified, among which only *Soltu.DM.03G024570* demonstrated significantly elevated expression levels in tuber bud of the short-dormant PHU (P05 and seven PHU accessions) in

Please cite this article in press as: Lian et al., A genomic variation map provides insights into potato evolution and key agronomic traits, Molecular Plant (2025), https://doi.org/10.1016/j.molp.2025.01.016

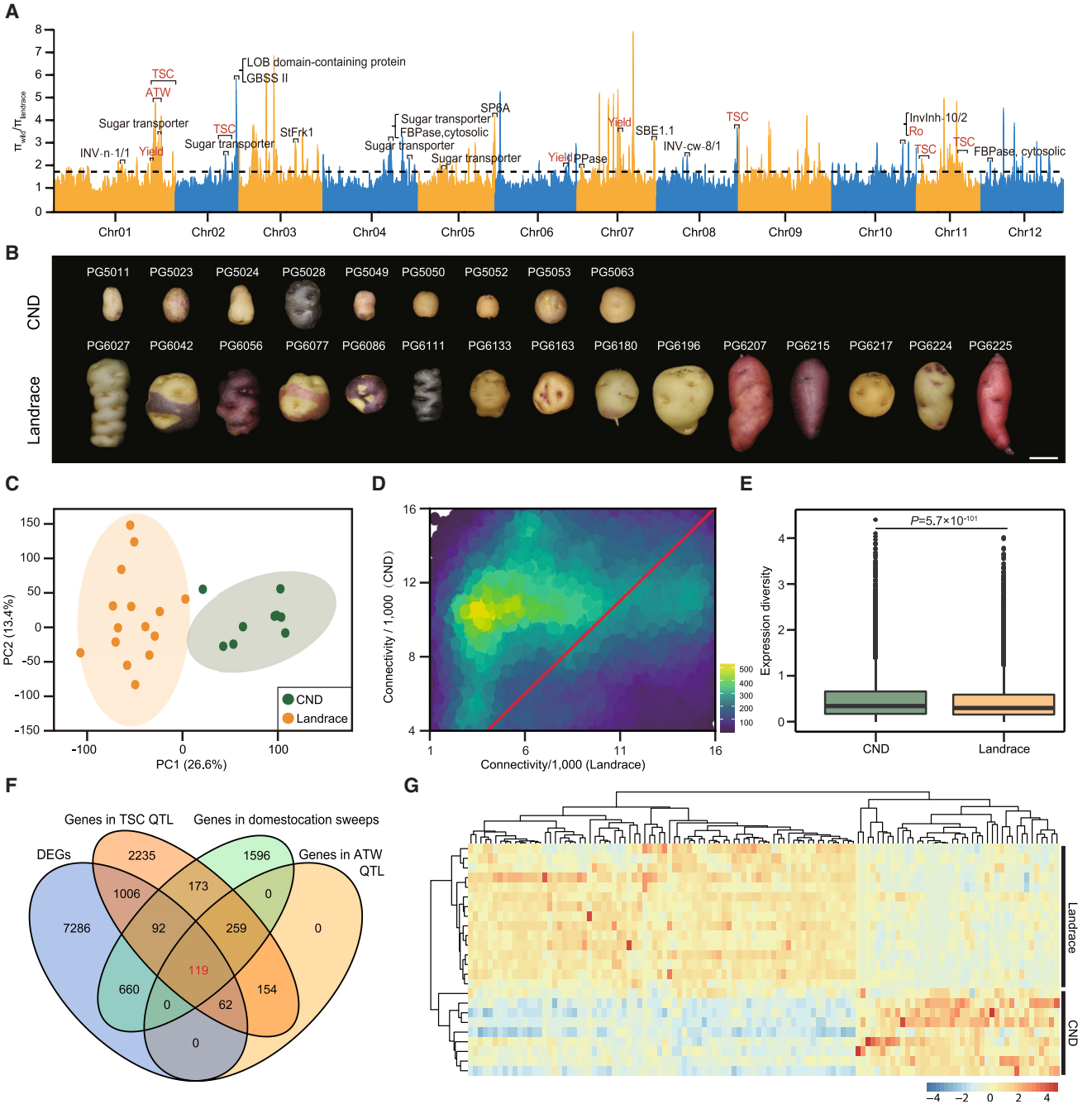

Figure 2. Identification of selection sweeps and reshaping of tuber transcriptome during potato domestication.

(A) Depiction of the 184 domestication sweeps (orange or blue bars above the dashed horizontal threshold line). Known genes (Zrenner et al., 1996; Navarro et al., 2011; Schreiber et al., 2014; Hardigan et al., 2017; Van Harsselaar et al., 2017) or QTLs (Endelman and Jansky, 2016; Śliwka et al., 2016; Braun et al., 2017; Marand et al., 2019; Sołtys-Kalina et al., 2020) located in domestication sweeps are shown in black and red font, respectively. ATW, average tuber weight; TSC, tuber starch content; Ro, tuber length/width.

(B) The phenotypic features of tubers from the CND and landrace groups. One representative tuber per accession was digitally extracted for comparison. Scale bar, 2.0 cm.

(C) PCA of the first two factors based on the FPKM of the 23,509 genes between the nine CND and 15 landrace accessions. The ellipses represent the 95% confidence interval of each group.

(D) Comparison of co-expression connectivity for genes in the CND and landrace groups. Purple indicates a low density of points, and yellow indicates a high density. A biased high connectivity is observed in CND. The red line indicates a slope of 1.

(E) Expression diversity of 23,509 genes for the CND and landrace groups. Significant differences were determined by Wilcoxon's rank-sum test.

(legend continued on next page)

Please cite this article in press as: Lian et al., A genomic variation map provides insights into potato evolution and key agronomic traits, Molecular Plant (2025), https://doi.org/10.1016/j.molp.2025.01.016

comparison to the long-dormant STN (PG6027 and seven STN accessions) (Figure 3D and Supplemental Figure 11). An 11-bp InDel resided in the promoter of *Soltu.DM.03G024570* (−82 bp to −93 bp from start code site) (Figure 3E) and the reference (Ref) allele containing the 11-bp sequence showed much higher frequency in the PHU group than those in STN (Figure 3F), indicative of divergent selection. Moreover, the InDel is predicted to be the binding site for the GATA transcription factor, thereby potentially affecting the transcription activity of *Soltu.DM.03G024570* (Figure 3E). Transient transcription activity assays confirmed that the promoter sequence of the Ref allele had significantly higher transcriptional activity than that of the alternative (Alt) allele (Figure 3G). This is consistent with the significantly higher expression of *Soltu.DM.03G024570*.

The homolog of *Soltu.DM.03G024570* in *Arabidopsis* encodes a peroxisomal acyl-coenzyme A (CoA) synthetase and catalyzes esterification of 12-oxo-phytodienoic acid (OPDA), which is a late and essential step in jasmonic acid (JA) biosynthesis (Kienow et al., 2008). It has been shown that JA and its bioactive jasmonyl-isoleucine (JA-Ile) conjugate acts as a negative regulator of dormancy in apple and pear (Ranjan and Lewak, 1992; Yildiz et al., 2008), and Suttle et al., (2011) found a dramatic increase of JA and JA-Ile during potato dormancy release. Recently, Liu et al., (2024a) reported that 1 mg/L methyljasmonate (MeJA) treatment promoted potato dormancy release and significantly improved tuber sprouting rates. To test whether *Soltu.DM.03G024570* affects tuber dormancy through JA and/or JA-Ile, we measured phytohormones of 11 STN (average dormancy period 12.2 weeks) and nine PHU (average dormancy period one week) accessions and found the short-dormant PHU to contain significantly higher levels of JA-Ile (~4-fold) (Figures 3H and 3I), which may be driven by the higher expression of *Soltu.DM.03G024570* in these accessions. Taken together, our data nominated *Soltu.DM.03G024570* as a plausible candidate of *qDR3.1* for the tuber dormancy period that underwent differential selection between STN and PHU.

Identification of a UDP-glucosyltransferase involved in SGA biosynthesis

Steroidal glycoalkaloids (SGAs) are toxic and bitter-tasting secondary metabolites produced by Solanaceae species (Valkonen et al., 1996; Li et al., 2023a). Potato contains various SGAs with different numbers of glycosidic units in the carbohydrate side chain, based on which these SGAs could be classified into trisaccharide type (α-chaconine, α-solanine, and leptines) or tetrasaccharide type (α-tomatine, demissine, and commersonine) (Distl and Wink, 2009; Zhao et al., 2021). However, the genes contributing to SGA biosynthesis in potato are still not fully understood.

To explore the genetic basis of SGA variation, we performed genome-wide association studies (GWAS) using 115 CND and landrace accessions (Supplemental Table 9). One significant signal (SNP_{SGA}, $p = 1.98 \times 10^{-18}$) was detected on chromosome 7 (Figure 4A), which resided within a previously characterized cluster of *GLYCOALKALOID METABOLISM* (*GAME*) genes of *GAME11*, *GAME6*, and *SGT1* (Itkin et al., 2013) (Figure 4B). Seven CND accessions carrying the "T" allele of peak SNP_{SGA} contained insignificant levels of α-chaconine and α-solanine (0.39 mg/kg on average, compared to 82.61 mg/kg for accessions with the "C" allele) (Supplemental Table 9). This suggests that this target genomic region harbors causal gene(s) that significantly influence SGA biosynthesis. Synteny analysis of this genomic region between potato and tomato showed that a uridine diphosphate (UDP)-glycosyltransferase gene (*SlGAME17*) is present on tomato chromosome 7 but absent from the potato reference genome DM v6.1 (Figure 4C). *De novo* assembly of tuber transcriptome and RT–PCR revealed that a potato gene transcript was specifically expressed in the seven aforementioned CND accessions (Figure 4D). Phylogenetic analysis indicated that this transcript is a homolog of tomato glucotransferase gene *SlGAME17* (referred to hereafter as *StGAME17*) but genetically distant from the previously characterized glucosyltransferase *SGT2* in potato (Figure 4E).

Next, we expressed the StGAME17 recombinant protein in *E. coli* to test its function. When using γ-chaconine (solanidine-1Glu, $m/z = 560.40$), which is the glucosylation product of SGT2 as substrate (McCue et al., 2006), formation of solanidine-2Glu ($m/z = 722.45$) was detected in the reaction containing StGAME17 instead of SGT2, suggesting that StGAME17 is a new glucosyltransferase that can act consecutively with SGT2 (Figure 4F, Supplemental Figure 12). In addition, solanidine-2Glu was also detected in the tuber flesh of transgenic lines (OE-18 and OE-22) overexpressing *StGAME17* (Figure 4G and 4H), consistent with that of the recombinant enzyme assay. As there are only three glycosyltransferases—SGT1 (galactosyltransferase), SGT2 (glucosyltransferase), and SGT3 (rhamnosyltransferase)—working in a "tri-step" manner to catalyze the glycosylation of solanidine into trisaccharide α-chaconine and α-solanine in most potatoes (Cárdenas et al., 2015; Zhao et al., 2021), additional expression of *StGAME17* may lead to the production of various SGAs with two glucose residues. We detected relatively high levels of α-tomatine and demissine (both are lycotetraose SGAs with two glucose residues) in PG5051 and PG5053 (Supplemental Figure 13), providing further support for this hypothesis. Moreover, we found that most randomly selected wild (clade 1+2, clade 3, and clade 4) accessions (97.3%) possessed *StGAME17*, but the proportion of *StGAME17* was reduced to 72.7% (24 in 33) in CND and 2.0% (3 in 151) in landraces (Figure 4I). Using the publicly available haplotype-resolved potato reference genomes (Bao et al., 2022; Hoopes et al., 2022; Sun et al., 2022; Cheng et al., 2025), we identified three haplotypes surrounding *StGAME17* and found that those carrying loss-of-function (*Hap2*) and absence of *StGAME17* (*Hap3*) were fixed in landraces and modern cultivars (Supplemental Figure 14). Our data demonstrated that *StGAME17* is a glucosyltransferase in potato and may contribute to the biosynthesis of various SGAs in potato.

(F) Venn diagram illustrating the identification of 119 candidate genes (in red font) through the intersection of selection sweeps, DEGs, and QTLs (ATW and TSC).

(G) Heatmap depicting the expression of 119 genes assigned to two modules in both species by hierarchical clustering. Expression values were scaled with *Z* score.

Please cite this article in press as: Lian et al., A genomic variation map provides insights into potato evolution and key agronomic traits, Molecular Plant (2025), https://doi.org/10.1016/j.molp.2025.01.016

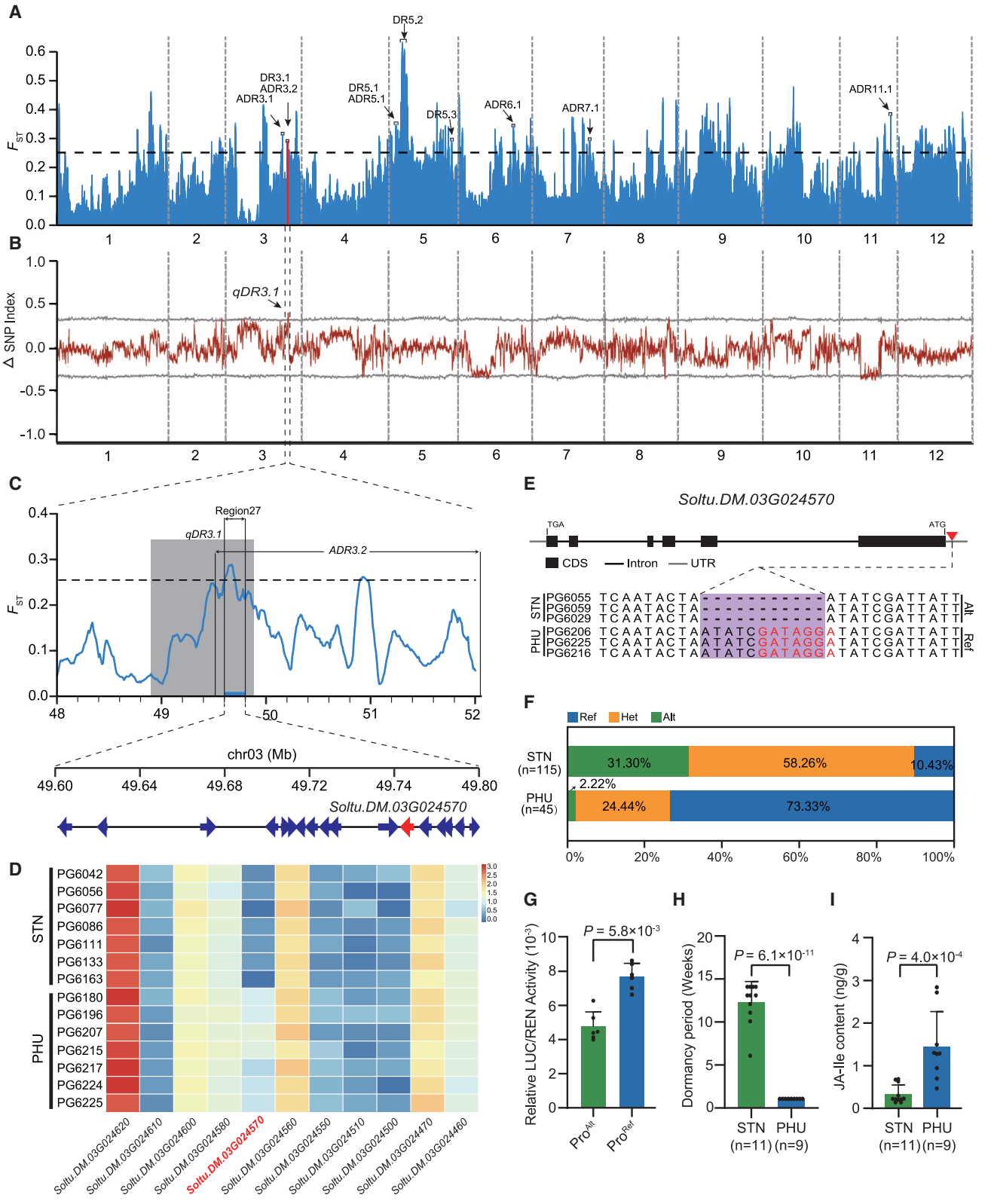

Figure 3. Landrace differentiation and identification of a peroxisomal acyl-CoA synthetase gene underlying short tuber dormancy period in PHU.

(A) Genome-wide F_{ST} values were calculated for the STN and PHU groups. Arrows indicate the 10 QTLs related to dormancy release (DR) or apical dominance release (ADR) (Bisognin et al., 2018) located within the divergent regions.

(legend continued on next page)

Please cite this article in press as: Lian et al., A genomic variation map provides insights into potato evolution and key agronomic traits, Molecular Plant (2025), https://doi.org/10.1016/j.molp.2025.01.016

Discovery of a DREB transcription factor that confers higher average tuber weight in landraces

Tuber weight is critical for potato yield and market value (Knowles and Knowles, 2006; Shubha and Singh, 2018; Aliche et al., 2019). To identify genes underlying tuber weight variation in cultivated potatoes, we performed GWAS analysis for average tuber weight (ATW) with 126 landrace accessions (CLD4-N) (Figures 5A and 5B, Supplemental Table 10). Interestingly, significant signals were detected within a ~370-Kb interval (7.31–7.68 Mb) on Chr03, which contained five annotated genes (*Soltu.DM.03G005750*, *Soltu.DM.03G005760*, *Soltu.DM.03G005770*, *Soltu.DM.03G005790*, and *Soltu.DM.03G005810*), with the most significant SNP_{ATW} residing upstream of *Soltu.DM.03G005790* that encodes a Dehydration-Responsive Element-Binding Protein (DREB) and belongs to the APETALA2/ethylene-responsive element-binding factor (AP2/ERF) transcription factor family (Figures 5C and 5D). In potatoes, the tuber growth rate surges at the bulking stage but drops to nearly zero at maturation (Fernie and Willmitzer, 2001; Kloosterman et al., 2005; Mihovilovich et al., 2014). Among the five candidate genes, only *Soltu.DM.03G005790* showed an obviously higher expression at the bulking stage than that at harvest (Figure 5E), suggesting *Soltu.DM.03G005790* as a plausible candidate for this trait.

To test this hypothesis, we generated *Soltu.DM.03G005790*-overexpressing (OE-13 and OE-23) and knockout (KO-90) lines and measured their average tuber weight and yield in two locations (Hohhot and Dingxi) in 2023 and 2024 (Figures 5F–5I and Supplemental Figure 15). Compared to the wild-type plants, both overexpressing lines OE-13 and OE-23 exhibited significant increases in average tuber weight (22.72%–25.88%) and yield (29.15%–54.27.5%) in 2024 (Figures 5F–5I). In contrast, the knockout line KO-90 showed reductions in average tuber weight (23.3%) and yield (19.8%). These data suggest a potential role for *Soltu.DM.03G005790* in regulating average tuber weight in potato.

Genomic evolution from diploid landraces to tetraploid cultigens

Tetraploid potatoes are cultivated worldwide due to higher vigor and yield compared to their diploid counterparts. However, the route of evolution of tetraploid varieties from their diploid progenitors is poorly understood. Various studies suggested that several diploid species, including *Solanum vernei* and *Solanum brevicaule*, as well as diploid landraces in STN and PHU, might be the progenitors of tetraploid varieties (Hosaka, 1986; Hawkes, 1988; Grun, 1990; Gupta and Tsuchiya, 1991; Sukhotu and Hosaka, 2006; Spooner et al., 2014). To provide insights into the formation of cultivated tetraploids, we inferred the phylogenetic relationships among the nine tetraploid landraces (4X-L) resequenced by Hardigan et al., (2017), together with the 314 diploid accessions examined in this study. Accessions in 4X-L were clustered into one monophyletic group that is sister to diploid landraces, indicating that 4X-L is genetically closer to diploid landraces and CND than to other wild relatives (Supplemental Figure 16). This result was also supported by the markedly higher genome similarity between 4X-L and STN, PHU, and CND compared to other wild species (Supplemental Figure 17).

We further dissected the composition of the tetraploid potato genome by taking advantage of genome-wide group-specific SNPs. To minimize possible bias due to sample size, we randomly selected eight accessions with a genetic background solely from *S. vernei* (only two accessions), *S. brevicaule*, CND, STN, and PHU (Supplemental Figure 18A). We found that the number of group-specific SNPs in 4X-L from STN was 9- to 72-fold greater than that of the other groups (Supplemental Figure 18B). We further genotyped these group-specific SNPs in these 4X-L accessions and found that STN-specific SNPs were predominantly presented (Supplemental Figures 18C and 19). Collectively, our results provide genomic evidence that the STN landrace group might be the closest landrace group that contributes to the formation of tetraploid potatoes.

Introgressions from wild species had been demonstrated to introduce favorable alleles of disease resistance, tuber quality, and/or stress tolerance into tetraploids (Ross, 1966; Barone et al., 2001; Van Der Vossen et al., 2003). To characterize the landscape of wild introgression within the tetraploid potato genome, we performed an ABBA-BABA test and detected 224 candidate introgression regions (size range: 307.09 Kb to 5.05 Mb, with an average of 922.99 Kb) with a length of 206.75 Mb that covered 27.90% of the reference genome (Supplemental Figure 20A and Supplemental Table 11), similar to observations by Hardigan et al., (2017) and Bao et al., (2022). These introgression regions harbored 11,421 genes, among which a cluster of 23 late

(B) Bulked segregant analysis sequencing for tuber dormancy period. The ΔSNP index and its corresponding 99% confidence interval are depicted as red and gray lines, respectively. QTL (*qDR3.1*) on Chr03 is indicated by a black arrow.
(C) Local F_{ST} plot (blue line) surrounding the QTL *qDR3.1* (gray shaded area) for tuber dormancy period. The overlapped region (*qDR3.1*, differentiation region 27, and *ADR3.2*) contained 17 genes.
(D) Different expression of the 11 candidate genes (six other genes without expression in tuber bud are not shown) between STN and PHU. Candidate gene *Soltu.DM.03G024570* is indicated in red. The results are expressed as $\log_{10}$(FPKM + 1).
(E) Gene structure and the 11-bp InDel in the promoter of *Soltu.DM.03G024570*. Upper panel: exon–intron structure of *Soltu.DM.03G024570*. The red triangle indicates the location of the 11-bp InDel. Lower panel: genomic sequence surrounding the 11-bp InDel (−82 bp to −93 bp from start code site) of *Soltu.DM.03G024570* in three STN and three PHU with known genome sequence (Tang et al., 2022); PG6055 and PG6059 are heterozygous (Ref/Alt) for this InDel. The purple-shaded rectangle denotes the position of the 11-bp InDel, and the DNA sequence in red indicates the predicted GATA motif by PlantCare (Lescot et al., 2002).
(F) Distribution of the three genotypes of the 11-bp InDel in 115 STN and 45 PHU accessions.
(G) Validation of function of the 11-bp InDel of *Soltu.DM.03G024570* using transient transcription activity assays. The promoter sequence of the Ref allele (from PG6180) had significantly higher transcriptional activity than that of the Alt allele (from PG6055) (n = 6 biologically independent replicates).
(H and I) Tuber dormancy period **(H)** and quantification of JA-Ile content **(I)** for the 11 STN and nine PHU accessions.
Horizontal dashed line in **(A)** and **(C)** indicates the genome-wide F_{ST} threshold (0.25) between STN and PHU. Data in **(G)–(I)** are presented as means ± SD. Significant differences were determined by a two-tailed Student's t-test.

Please cite this article in press as: Lian et al., A genomic variation map provides insights into potato evolution and key agronomic traits, Molecular Plant (2025), https://doi.org/10.1016/j.molp.2025.01.016

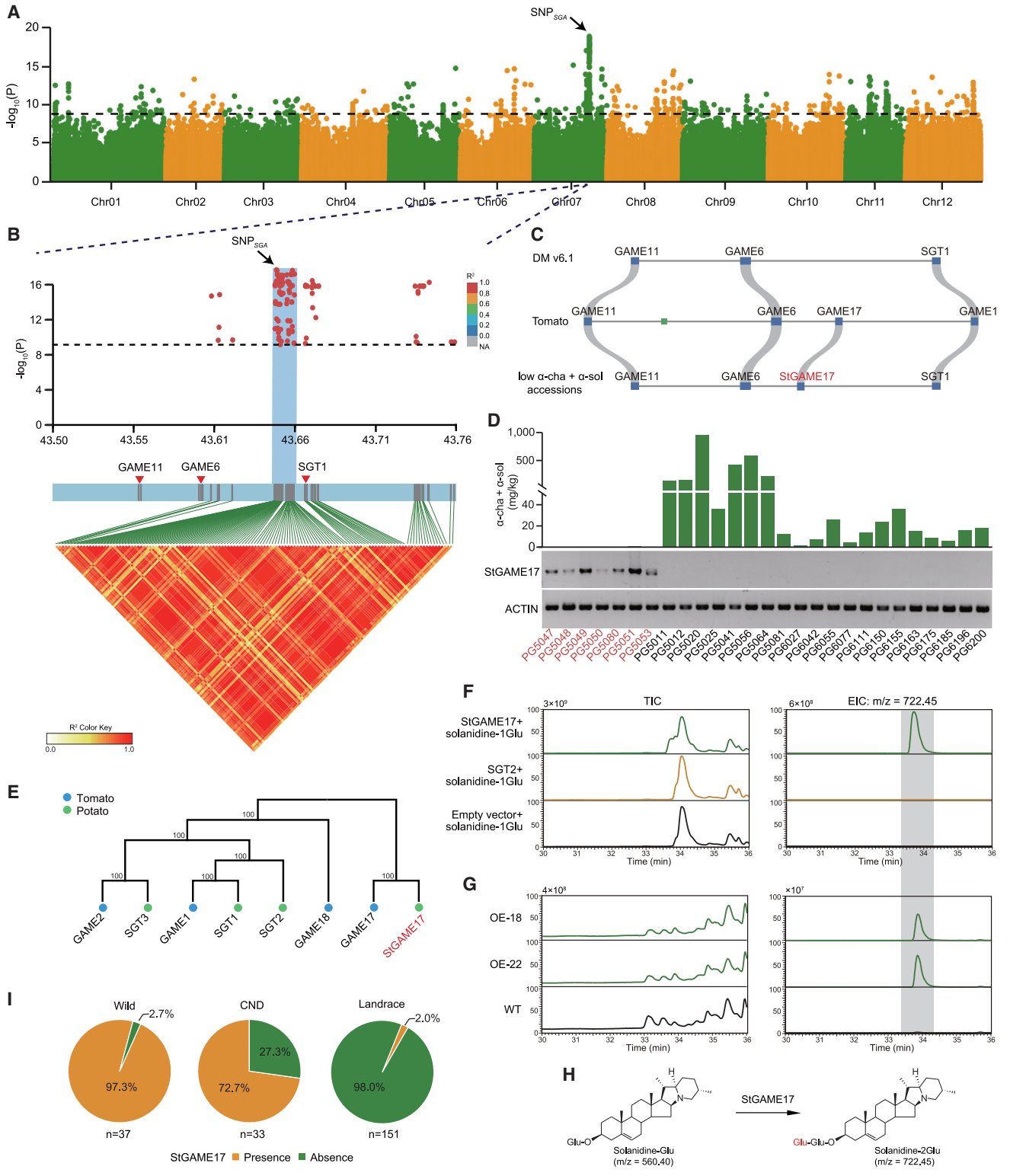

Figure 4. GWAS for tuber flesh SGA content and functional analysis of *StGAME17*.

(A) Significant marker-trait associations were found on chromosome 7 (Chr07: 43,644,193) of the Manhattan plot.

(B) Local Manhattan plot (top, only SNPs beyond the threshold are shown) and LD heatmap (bottom, measured by r^2) surrounding the peak on chromosome 7. The candidate region lies within the light-blue shaded rectangle. In **(A)** and **(B)**, the horizontal dashed line represents the significance threshold ($-\log_{10}(p) = 9.1$).

(legend continued on next page)

Please cite this article in press as: Lian et al., A genomic variation map provides insights into potato evolution and key agronomic traits, Molecular Plant (2025), https://doi.org/10.1016/j.molp.2025.01.016

blight-resistant *R2 Gene Homologs* (*R2GHs*) within a 200-Kb interval (5.76–5.96 Mb) on chromosome 4 was detected (Destefanis et al., 2015) (Supplemental Figure 20A and Supplemental Table 12). A total of 33 representative reliable SNPs (31 of which were verified by HiFi reads, Supplemental Table 13) within this *R2GHs* exhibited contrasting allele compositions between the wild and diploid landrace groups. Notably, the Ref-allele frequencies of tetraploids were more similar to those of wild species or CND than those of diploid landraces, providing strong evidence of introgression (Supplemental Figure 20B). Our data suggest that cultivated tetraploid potatoes primarily originated from the *S. tuberosum* group Stenotomum, with subsequent introgressions from wild species to enhance adaptation to the environment and response to pathogen attack.

DISCUSSION

The resequencing of 314 wild and landrace accessions allowed us to capture genetic variation among the highly diversified tuber-bearing *Solanum* section *Petota*, reconstruct their domestication and differentiation history, and identify genes responsible for important agronomic traits. We revealed transcriptional reshaping during potato domestication and characterized candidate genes controlling important agronomic traits such as tuber dormancy period, SGAs biosynthesis, and average tuber weight. These findings provide new knowledge to our understanding of potato evolution and genome design breeding (Zhang et al., 2021).

Coupling phylogenetic, PCA, and IBD analyses, we grouped the 314 diploid accessions into four major clades, consistent with previously reported classifications (Spooner et al., 2014; Li et al., 2018; Bozan et al., 2023). Additionally, our data support the monophyletic origin of diploid landrace potatoes, with *S. candolleanum* (CND) identified as their immediate wild progenitor, aligning with the recent potato population genetic and pan-genomic studies (Hardigan et al., 2017; Li et al., 2018; Tang et al., 2022; Zhang et al., 2022). Meanwhile, we collected more landrace accessions compared to previous studies (Hardigan et al., 2017; Li et al., 2018; Bozan et al., 2023). This expanded dataset enabled us to distinguish STN and PHU groups successfully, providing a foundation for further exploring the genetic basis underlying their differentiation.

Additionally, we identified 184 candidate domestication sweeps harboring 2,899 genes, only a few of them overlapped with those identified in previous studies, presumably due to the different set of wild ancestors (CND *vs*. Clade 4 wild species) as well as the larger sample size of landrace accessions (160 *vs*. 10–33) used for sweep implementation (Hardigan et al., 2017; Li et al., 2018; Bozan et al., 2023). Moreover, we noticed that tubers in CND and landrace showed altered gene expression landscapes (Figure 2), and some of the DEGs resided within overlapping regions of domestication sweeps and QTLs, serving as potential candidates for these traits. In future studies, further extensive investigations of these genes, i.e., QTL mapping and gene cloning with biparental segregating populations coupled with function verification through transgenic approach, would help us understand how they contributed to the potato domestication process.

Domestication sweeps are genomic regions that show significant reductions of nucleotide diversity in cultivated species compared to their wild ancestors, caused by strong selective pressures acting on favorable genes as well as the hitch-hiking of their closely linked loci (Smith and Haigh, 1974; Meyer and Purugganan, 2013). However, it is important to note that scenarios such as demographic events and background selection acting on deleterious mutations linked to neutral variation may also reduce nucleotide diversity (Kantar et al., 2017). Specifically, nonequilibrium demographic histories such as population bottlenecks and genetic drifts can affect genetic diversity, complicating the identification of regions associated with domestication and differentiation (Beaumont, 2005; Lotterhos and Whitlock, 2014; Schrider et al., 2016). These above factors should be carefully considered to avoid confounding effects. Despite these challenges, our analysis leveraging SweeD analysis, reduction of nucleotide diversity, and F_{ST} provides a preliminary step in identifying candidate genes potentially under selection and differentiation. Future methodologies incorporating demographic modeling will hold promise for improving the accuracy of selection signals.

We leveraged the catalog of genomic variations in this study to identify candidate genes of agronomically important traits. Through the integration of QTL-mapping, transcriptomics, and population differentiation, we nominated a peroxisomal acyl-CoA synthetase gene *Soltu.DM.03G024570* as a candidate gene associated with the tuber dormancy period and provided further biochemical and

(C) Identification of the candidate UDP-glucosyltransferase gene (*StGAME17*) by synteny analysis. α-cha and α-sol denote α-chaconine and α-solanine, respectively.
(D) Expression of *StGAME17* in tuber flesh correlates with the insignificant α-chaconine and α-solanine content phenotype. Upper panel: α-chaconine and α-solanine content in the tuber flesh of seven CND accessions (in red font) with "T" allele for peak SNP and 20 randomly selected CND and landrace accessions (in black font) with "C." Lower panel: Expression analysis of the aforementioned accessions with RT–PCR. *Actin* was used as internal control.
(E) Phylogenetic analysis of StGAME17 and other related steroidal alkaloid glycosyltransferase (SGT)-family proteins from potato and tomato using the maximum-likelihood (ML) method in MEGA6 (Tamura et al., 2013). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) is shown next to the branches. Accession numbers as follows: *SlGAME1* (*Solyc07g043490*), *SlGAME2* (*Solyc07g043410*), *SlGAME17* (*Solyc07g043480*), *SlGAME18* (*Solyc07g043500*), *SGT1* (*Soltu.DM.07G014220*), *SGT2* (*Soltu.DM.08G022920*), and *SGT3* (*Soltu.DM.07G014160*). The sequence of *StGAME17* was obtained from PG5080 (CND) with RT–PCR.
(F and G) *In vitro* and *in vivo* validation of *StGAME17* in potato SGA glucosylation. Product molecule (m/z = 722.45, solanidine-2Glu), indicated by the gray shaded rectangle in extracted ion chromatograms (EIC), was generated when the StGAME17 recombinant protein was incubated with substrate molecule (m/z = 560.40, solanidine-1Glu), compared to empty vector or SGT2 **(F)**. Solanidine-2Glu was also detected in *StGAME17*-overexpression (OE) lines OE-18 and OE-22 **(G)** Products were putatively identified by MS/MS fragmentation spectrum. TIC, total ion chromatogram.
(H) Schematic of biosynthetic pathway from solanidine-1Glu to solanidine-2Glu.
(I) Frequency distribution of *StGAME17* in Wild, CND, and landrace. "Wild" denotes accessions randomly selected from CLD1+2, CLD3, and CLD4.

Please cite this article in press as: Lian et al., A genomic variation map provides insights into potato evolution and key agronomic traits, Molecular Plant (2025), https://doi.org/10.1016/j.molp.2025.01.016

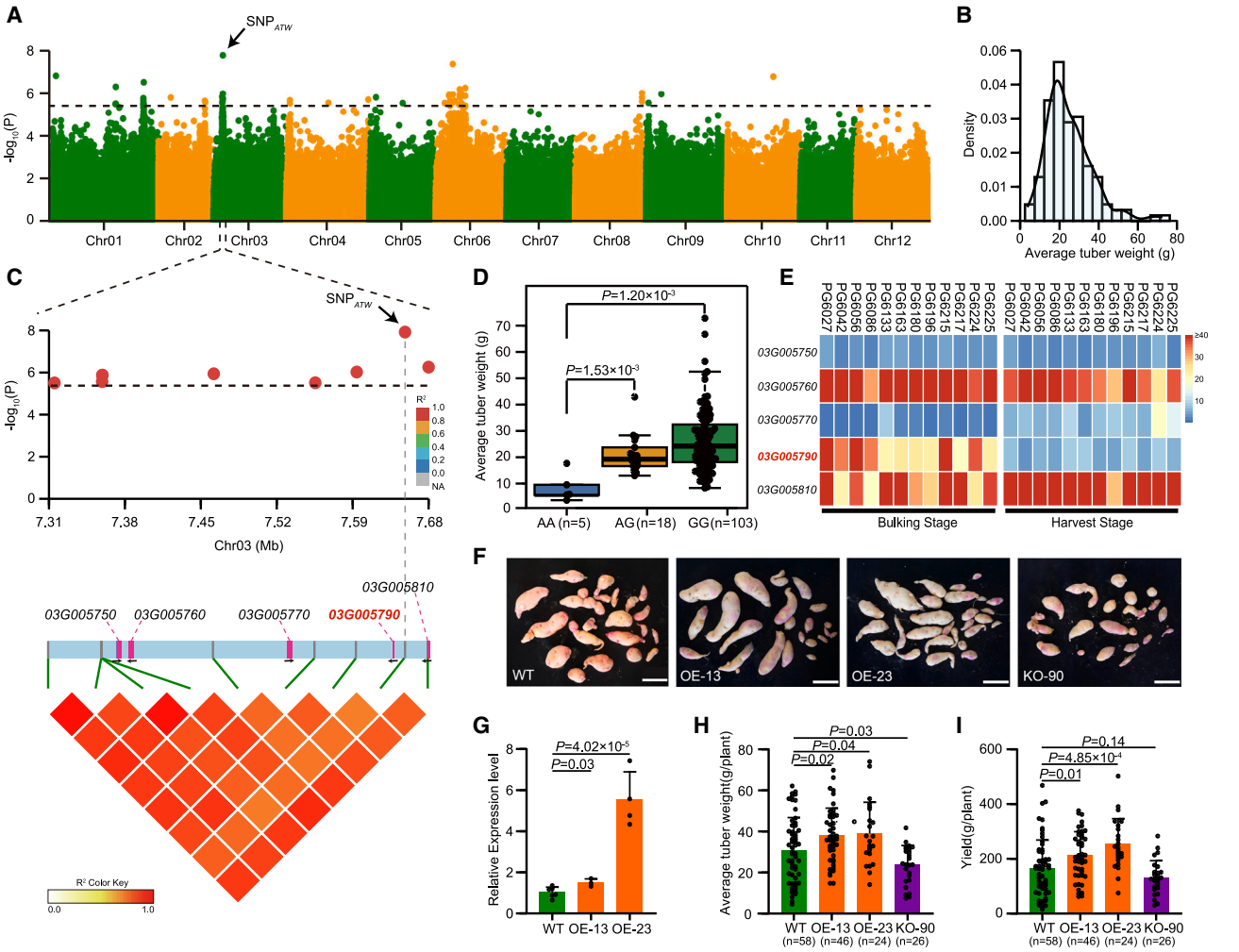

Figure 5. GWAS for average tuber weight in landrace and functional analysis of *Soltu.DM.03G005790*.

(A) Manhattan plot of the GWAS for average tuber weight in potato landrace. Arrow indicates the most significantly associated SNP$_{ATW}$ (Chr03: 7,657,782).

(B) Distribution of average tuber weight for the 126 landrace accessions used for GWAS.

(C) Local Manhattan plot (top, only SNPs beyond the threshold are shown) and LD heatmap (bottom, measured by r^2) surrounding the peak on Chr03. The most significant SNP$_{ATW}$ (Chr03: 7,657,782) in the local Manhattan plot and LD heatmap is connected by a vertical dashed line. The five genes within the candidate region are labeled based on their physical position, with the *03G005790* gene highlighted in red. The orientation of each annotated gene is marked with black arrows under the gene symbols. For conciseness, the prefix "*Soltu.DM.*" for each gene was omitted in this figure. In **(A)** and **(C)**, the horizontal dashed line represents the significance threshold ($-\log_{10}(p)$ = 5.48).

(D) Average tuber weight in accessions carrying distinct alleles of the top SNP. In boxplots, the 25% and 75% quartiles are represented by the lower and upper edges of boxes, respectively, and central lines denote the median. The whiskers extend to 1.5 times the interquartile range. *n* represents the sample size in each group. Significance was determined using two-tailed Student's *t*-tests.

(E) Heatmap showing transcriptional abundance (measured by FPKM) of the five candidate genes in 12 randomly selected landrace tubers at bulking (left) and harvest (right) stages.

(F) Tuber phenotypes of WT and *Soltu.DM.03G005790*-overexpression (OE) and knockout (KO) lines. Scale bars, 5.0 cm.

(G) The expression levels of *Soltu.DM.03G005790* in tubers of WT and OE lines.

(H and I) Average tuber weight **(H)** and yield **(I)** significantly increased in OE lines compared to WT, while they decreased in the KO line. The data were obtained from the field experiment shown in **(F)**. Data are presented as means ± SD (*n* indicated the plants for WT, OE, and KO lines, respectively). Significant differences were determined by a two-tailed Student's *t*-test.

metabolic assays to support our claim (Figure 3). GWAS analysis has been overlooked in previous diploid potato population genomic studies (Hardigan et al., 2017; Li et al., 2018), despite its potential for high-resolution mapping of candidate genes (Lin et al., 2014; Zhao et al., 2019; Liu et al., 2024b). In this research, we performed GWAS and characterized a previously uncharacterized UDP-glycosyltransferase gene *StGAME17* in potato. This gene catalyzed the formation of two glucose residues containing glycosidic units, thus possibly further contributing to the biosynthesis of tetrose-containing SGAs in potato (Figure 4). In contrast to the cultivated potatoes whose major SGAs (~95%) are α-chaconine and α-solanine, wild potato species usually

Please cite this article in press as: Lian et al., A genomic variation map provides insights into potato evolution and key agronomic traits, Molecular Plant (2025), https://doi.org/10.1016/j.molp.2025.01.016

contain various SGAs such as α-tomatine, demissine, and commersonine in their tubers (Kozukue et al., 2008; Distl and Wink, 2009). Although toxic to humans, SGAs have been shown to play crucial protective roles for *Solanaceae* plants by conferring resistance to various pathogens and predators (Osman et al., 1978; Paudel et al., 2017; Wolters et al., 2023; Boccia et al., 2024). For instance, α-tomatine, commersonine, and leptine I showed toxicity to Colorado Potato Beetle (CPB) which is a major pest of *S. tuberosum* and other *Solanum* crops (Sinden et al., 1986; Flanders et al., 1992; Kowalski et al., 2000; Wolters et al., 2023), and α-solanine and α-chaconine inhibited the growth of late blight pathogen *Phytophthora infestans*, *Fusarium oxysporum*, and *Fusarium sulphureum* which causes the potato dry rot (Dahlin et al., 2017; Fagundes et al., 2018; Li et al., 2023b). In recent years, with increasing use of wild *Solanum* germplasm to introgress desirable resistance genes in breeding programs, various SGAs in wild potatoes are usually brought back into the cultivated gene pool, posing considerable concern for food safety (Van Gelder and Scheffer, 1991; Kozukue et al., 1999; Väänänen et al., 2005). The knowledge of *StGAME17* obtained in this study, coupled with its tissue-specific expression in foliage instead of tubers, offers the potential to improve the biotic resistance of leaves to certain pathogens and predators while ensuring tuber safety for human consumption, thereby supporting sustainable and eco-friendly potato production.

We also discovered a DREB transcription factor gene *Soltu.DM.03G005790* conferring increased average tuber weight and verified its function through a transgenic approach (Figure 5). DREB transcription factors are generally regarded to play important roles in plant resistance to abiotic stresses such as drought, salinity, heat, or cold (Lata and Prasad, 2011; Xu et al., 2011). Recent studies also showed overexpression of DREB homologs, including *OsDREB1C* and *GmDREB3* (*GmTDN1*), increased yield in rice and wheat by improving photosynthetic capacity, nitrogen utilization, and flowering time (Wei et al., 2022; Zhou et al., 2022). Knowledge of these candidate genes will add new options to the breeders' toolkit for breeding new varieties with increased yield and quality, especially when wild species or genetic stocks, including landraces, are used.

The missing link between primitive diploid potato landraces and modern commercial tetraploid cultivars is how the tetraploids originated from their diploid progenitors. Based on genome similarity and group-specific SNP analysis, we revealed that the STN landrace group might be the closest landrace group that contributes to the formation of tetraploid potatoes rather than the other diploids. This provides new insights into the long-term dispute on potato polyploidization. Additionally, we identified widespread introgression from various wild species into tetraploid cultivars, supporting earlier findings by Hardigan et al., (2017) and Bao et al., (2022). Notably, we discovered a cluster of 23 late blight-resistant *R2GH* genes within a 200-Kb region on chromosome 4, demonstrating the potential of wild resistance genes in potato breeding.

Our findings shed important light on the evolution history of potatoes and nominated candidate genes of important tuber yield and quality traits. The knowledge and resources generated in this study provide a strong basis for the future use of diploid wild species and landraces in genome-guided hybrid potato breeding.

METHODS

Plants and sequencing

The 314 diploid potato accessions were obtained from CIP (International Potato Center) and USDA ARS (Agricultural Research Service, the US Department of Agriculture). Passport information about these accessions, including their names, origin, classification, and resequencing data, is listed in Supplemental Table 1. All accessions were planted in greenhouses at Inner Mongolia University, Hohhot, China, for genome resequencing. Genomic DNA was extracted from young leaves using the cetyltrimethylammonium bromide (CTAB) method (Gawel and Jarret, 1991). At least 5 μg of DNA per accession was used to construct the sequencing library. Paired-end sequencing (150 bp) of each library was performed on the Illumina HiSeq X Ten platform with an insert size of 300 bp.

Read mapping and variation calling

To identify genetic variations, the clean reads were mapped to the DM v6.1 genome (Pham et al., 2020) using BWA-mem software (v6.0.2) (Li and Durbin, 2009) with default parameters; sam files were converted to bam files using SAMtools (v0.1.18) software (Li et al., 2009). After completing the mapping, the bam file was sorted, and duplicates were marked with Picard tools (v1.119) (http://broadinstitute.github.io/picard/).

Variants were called by GATK (v4.2.3.0) (McKenna et al., 2010) HaplotypeCaller, and the SNPs and InDels were further filtered using the following criteria: SNPs were filtered with "QD < 2.0 || FS > 60.0 || MQ < 40.0 || SOR > 3.0 || MQRankSum < −12.5 || ReadPosRankSum < −8.0" and InDels with "QD < 2.0 || FS > 200.0 || SOR > 10.0 || MQRankSum < −12.5 || ReadPosRankSum < −8.0"; SNPs with a missing rate of >40% or a minor allele frequency (MAF) of <0.01 were removed.

Phylogenetic, population-structure, and IBD analyses

A subset of 126,119 SNPs (MAF ≥ 0.08 and missing rate ≤ 0.4) at 4-fold degenerate sites representing neutral or near-neutral variants were used to construct the maximum-likelihood phylogenetic tree using PhyML (v3.0) (Guindon et al., 2010) with the parameters "-d nt -b 100 -f 0.25,0.25,0.25,0.25 -c 1 –no_memory_check -o tlr -s BEST". The branch length and rate parameters were optimized. For population-structure analysis, to ensure the representativeness of SNPs selected and lighten the computation pressure, we filtered the SNPs in heavy LD with PLINK software (v1.90b3.46) (https://www.cog-genomics.org/plink2) using the following command: plink –file input –indep-pairwise 50 10 0.2 –out output. SNPs with a missing rate of >20% were removed prior to population-structure analysis using Admixture (v1.3.0) (Alexander and Lange, 2011). For IBD analysis, the same SNP dataset as population-structure analysis was used with the command plink –file ld.filterd.example –cluster –matrix –noweb. PCA was performed with EIGENSOFT (v6.0.1) (Patterson et al., 2006) using whole-genome SNPs with missing value ≤40% and MAF ≥ 0.05.

Linkage disequilibrium analysis

LD analysis was performed with PopLDdecay software (v3.40) (Zhang et al., 2019a) using whole-genome SNPs with missing value ≤40%, MAF ≥ 0.05, and the following command: PopLDdecay -InGenotype input.genotype -OutStat result.out -MAF 0.05 -Miss 0.4. LD decay was calculated based on the r^2 value and the corresponding distance between two SNPs.

Genome scanning for selective signals

To detect potential selective sweeps, we evaluated the genome-wide decrease of genetic diversity (π) with VCFtools software (Danecek et al., 2011) using the command vcftools –gzvcf pop.vcf.gz –window-pi 100,000 –window-pi-step 10,000 –out result –keep target.group.list. Evidence for selection across the genome during domestication was evaluated by comparing *S. candolleanum* versus landraces (π_{CND}/$\pi_{landrace}$; both STN and PHU were included). Genomic regions affected by domestication should have substantially lower diversity in the landrace

Please cite this article in press as: Lian et al., A genomic variation map provides insights into potato evolution and key agronomic traits, Molecular Plant (2025), https://doi.org/10.1016/j.molp.2025.01.016

group than CND. Windows with $\pi < 0.001$ in CND were excluded from further analysis, and windows with the top 5% highest ratios of $\pi_{CND}/\pi_{landrace}$ (≥ 1.71) were selected as candidate domestication sweeps. Windows that were ≤ 100 Kb apart were merged into a single selected region. SweeD (v4.0.0) (Pavlidis et al., 2013) analysis was performed using genome-wide SNPs, calculating the likelihood values for each sliding window with a size of 10 Kb. The top 5% of windows with the highest scores were extracted and merged if adjacent. Only windows supported by both reduction of nucleotide diversity and SweeD were retained for subsequent analysis.

RNA-seq analysis

Nine ancestral CND and 15 landrace accessions were randomly selected for RNA-seq and planted in pots in Hohhot, Inner Mongolia, from May to October in 2019 and 2021. Bulking tubers (stages 5–6) (Kloosterman et al., 2005) were collected. Transverse slices were cut from the middle portion of each tuber, immediately frozen in liquid nitrogen, and stored at −80°C until analysis. Total RNA was extracted from tuber tissue using TransZol Plant (TransGen Biotech). Paired-end sequencing (150 bp) of each library was performed on the Illumina HiSeq X Ten platform with an insert size of 300 bp. Clean reads from each accession were mapped to the reference genome DM v6.1 (Pham et al., 2020) using HISAT2 software (Kim et al., 2015), followed by FPKM calculation using StringTie (v1.3.0) (Pertea et al., 2015). A total of 23,509 genes expressed in at least one of the 24 accessions (FPKM $\geq$ 1) were obtained. Differential gene expression analysis between CND and landrace groups was conducted using edgeR (v3.18) (Robinson et al., 2010). Genes identified as differentially expressed had an FDR < 0.05 as determined by edgeR. To gain additional insight into the pattern of gene expression changes between CND and Landrace, we built weighted gene co-expression networks. The connectivity for each gene was computed as the sum of the absolute values of the correlation coefficients of a focal gene with all other genes, as determined by weighted gene co-expression network analysis (WGCNA) (Langfelder and Horvath, 2008). The expression diversity for each gene was determined by calculating the ratio between the standard deviation (SD) (Liu et al., 2019) and the mean of the expression values, calculated separately for the landrace and CND groups.

Field trials and tuber starch content determination

The 39 CND and 160 landrace accessions were planted in greenhouses at Inner Mongolia University, Hohhot, China. Tubers from 172, 151, and 163 accessions were harvested for phenotypic evaluation in 2019, 2020, and 2021, with four replicates per accession. The average tuber weight and tuber yield of each plant were measured according to Braun et al., (2017). The dormancy evaluation was conducted as previously described (Bisognin et al., 2018) with one minor modification: dormancy period was recorded as weeks after harvest to the onset of 75% sprouting tubers with at least one sprout reaching 2 mm.

The tuber starch content was measured using a modified procedure previously reported (Jansen et al., 2001). Harvested tubers were peeled, cut into pieces, and about 20–50 g of potato flesh was grated using a homogenizer with 100 mL of distilled water. The extract was filtered through three layers of 70-μm gauze and pH adjusted to 7.0, incubated for 10 minutes at room temperature before addition of 20 mL 95% ethanol, incubated again, and centrifuged for 10 minutes at 10,000 × *g* to discard the supernatant. The wet starch was dried in an oven at 42°C until a constant weight was achieved. The tuber starch content of each sample was calculated by dry starch weight relative to tuber weight.

Identification of differentiated regions

The population fixation statistics (F_{ST}) were estimated with VCFtools software (Danecek et al., 2011) using 100-Kb sliding windows with a step size of 10 Kb. The average weighted F_{ST} of all sliding windows was regarded as the value at the whole-genome level across different groups. Sliding windows with the top 5% highest F_{ST} values were initially selected. Neighboring windows were then merged into a single fragment. If the distance between two fragments was <100 Kb, the fragments were merged into a single region. The final merged regions were considered to be highly divergent between two different groups.

Bulked segregant analysis

P05 (*S. tuberosum* group Phureja, short-dormant) and PG6027 (*S. tuberosum* group Stenotomum, long-dormant) were crossed to create F_1 progeny. The parents and 127 F_1 segregating progeny were planted in a greenhouse at Inner Mongolia University, Hohhot, China, in 2020 and 2021. Genomic DNA was extracted using the CTAB method from fresh young leaves. For bulked segregant analysis, the bulked pool of long- and short-dormant samples were constructed by mixing equal amounts of DNA from 11 individuals showing extremely long and short dormancy, respectively. Roughly 20× resequenced data for each parent (P05 and PG6027) and bulked pool (long- and short-dormant) were generated.

After removing the short and low-quality reads, clean reads were mapped onto the DM v6.1 reference genome using the Burrows–Wheeler aligner (BWA) (v6.0.2) (Li and Durbin, 2009) with default parameters; SNP were identified using GATK tools software (McKenna et al., 2010). SNPs with base quality value <20 and depth <7× coverage, SNP index <0.3, or SNP index >0.7 in both bulked sequences were filtered out. SNPs that were homozygous in one and heterozygous in the other between two parental genomes were identified for further analysis. SNP index and Δ(SNP index) were calculated using a 500-Kb sliding window with a step size of 10 Kb as described by Takagi et al., (2013).

Extraction and quantification of jasmonate-Ile

Sample extraction and quantification were conducted as described by Fang et al., (2022). Freshly harvested potato tuber buds were ground into a powder in liquid nitrogen, and ~50 mg of the sample was weighed accurately and transferred to 2-mL centrifuge tubes. Before extraction, 10 μL of stable isotope internal standard (10 μg/mL d5-JA, Cayman Chemical) was added to each tube. After adding 1.5 mL of extraction buffer (isopropanol/formic acid at 99.5:0.5, v/v), the sample was vortexed for 2 minutes in a mixer mill at 900 rpm three times. After a 15-minute centrifugation at 14,000 × *g*, the supernatants were dried and resuspended with 1 mL of methanol solvent (85:15, v/v). A Waters Sep-Pak C18 SPE column was used, and 2 mL of SPE column eluent was collected per sample. The eluent was dried and resuspended in 100 μL of methanol solvent (60:40, v/v) for liquid chromatography–mass spectrometry (LC–MS) analysis.

For the determination of jasmonates-Ile, the prepared samples were analyzed with an AB Sciex 4500 QTRAP triple quadrupole mass spectrometer (AB Sciex, MA, USA). All targeted compounds achieved separation on an ACQUITY UPLC BEH C18 column (Waters, Eschborn, Germany) (50 × 2.1 mm, 1.7 μm) based on an ACQUITY UPLC I-Class (Waters). The injection volume was 10 μL, with the flow rate 200 μL/min. Solvents for the mobile phase were 0.1% formic acid in acetonitrile (A) and 0.1% formic acid in acetonitrile (B). The gradient elution was 0–10 minutes, the linear gradient was 50%–100% A, and the column was washed with 100% B and 100% A before the next injection. Calibration standards were used to construct calibration curves of target analytes with the same stable isotope internal standard added to the samples. Mass spectra data and concentration calculations were processed using AB Sciex Analyst v1.6.3 software (AB Sciex).

Transient dual-luciferase assay

For transient transcription activity analysis of *Soltu.DM.03G024570*, the ~2-Kb upstream sequences from PG6055 (STN) and PG6180 (PHU) were cloned and ligated into pGreenII 0800-LUC double-reporter vector (Hellens et al., 2005). The constructed plasmids were co-transformed into tobacco by *Agrobacterium tumefaciens* strain GV3101 psoup-p19.

Please cite this article in press as: Lian et al., A genomic variation map provides insights into potato evolution and key agronomic traits, Molecular Plant (2025), https://doi.org/10.1016/j.molp.2025.01.016

The activities of firefly luciferase (LUC) and Renilla luciferase (REN) were measured using a dual-luciferase assay kit (Vazyme). The results were calculated by the ratio of LUC to REN. At least six transient assay measurements were contained for each assay. All primers used for the following construction are listed in Supplemental Table 14.

Extraction and quantification of SGAs

Freshly harvested potato tuber flesh was ground into a powder in liquid nitrogen, and ~100 mg of the sample was transferred to 1.5-mL centrifuge tubes. After adding 1 mL of extraction solution (80% methanol mixed with 0.1% (v/v) formic acid), the sample was vortexed and subjected to ultrasonic extraction for 60 minutes. The sample was clarified by centrifugation, and the supernatant was diluted to a suitable concentration with 0.1% (v/v) formic acid. A 1.0-mL aliquot was filtered through a 0.22-μm Millipore Millex-GV syringe filter (Millipore, Billerica, MA) prior to detection. To identify the α-solanine, α-chaconine, and other possible types of SGAs, LC–tandem MS (LC–MS/MS) or LC–MS was performed in positive ionization mode. The system was equipped with a SunFire C18 column (100 Å, 5 μm, 2.1 mm × 150 mm; Waters, USA). Eluent A was acetonitrile containing 0.1% (v/v) formic acid, and eluent B was ultra-pure water containing 0.1% (v/v) formic acid. The flow rate was 0.200 mL/min. The gradient began at 10% A and increased linearly to 60% A; the column was washed and equilibrated for 5 minutes before the next injection. The column temperature was 40°C, and the injection volume was 1.00 μL. The calibration curves of α-solanine and α-chaconine (ChromaDex, USA) were constructed with standards. The self-compatible inbred line SVA4 (*Solanum verrucosum*) (Du et al., 2021) with high SGA content was used as a control. To identify tetrasaccharide SGAs, LC–MS/MS was performed in positive ionization mode with some modifications. The system had an ACQUITY UPLC CSH C18 column (130 Å, 1.7 μm, 2.1 mm × 100 mm, 1/pk; Waters, USA). The calibration curves of α-tomatine, tomatidine, and demissidine were constructed with standards.

GWAS

Samples in the CND and/or landrace groups were used for GWAS. Phenotype data were transformed using the formulae $\log_{10}$(SGA content + 1) and $\log_2$(ATW + 1). SNPs were filtered with the criterion MAF $\geq$ 0.05 and missing rate $\leq$0.4. GWAS was then performed using the efficient mixed model association expedited (EMMAX) program (Kang et al., 2010). Population stratification and hidden relatedness were modeled with a kinship (K) matrix in the emmax-kin- intel package of EMMAX. The significance threshold value of the GWAS was evaluated with the formula $p = 0.01/n$ (where n is the total number of SNPs).

Synteny analysis

The software JCVI (Tang et al., 2024) was employed for synteny analysis. We first performed a pairwise synteny search between the tomato genome (v3.2) and the potato genome (DM v6.1) using the following command: python -m jcvi.compara.catalog ortholog potato.input tomato.input –no_strip_names. After obtaining the paired gene file, we visualized the target region using the following command: python -m jcvi.compara.synteny mcscan potato.bed potato.tomato.lifted.anchors –iter = 1 -o potato.tomato.i1.blocks.

Glycosyltransferase assay

For the StGAME17 enzyme activity assay, the full-length coding sequences of *StGAME17* and *SGT2* (*Soltu.DM.08G022920*) were amplified from cDNA (primes are listed in Supplemental Table 14) and inserted between the *Bam*H I and *Kpn* I restriction sites in the vector pMAL-c5x. The plasmids were heat-shock transformed into *E. coli* Rosetta. An empty vector, and SGT2 were used as negative control in the enzyme assays. Fresh overnight cultures were inoculated into 200 mL of fresh Luria broth (LB) with the antibiotics and incubated at 37°C until OD_{600} reached 0.6, following induction by the addition of 0.5 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) and grown continually for 4 hours at 37°C. Cells were harvested and resuspended in column buffer (20 mM Tris–HCl, 200 mM NaCl, 1 mM EDTA, pH 7.4). The cells were disrupted by high pressure, and cell debris was removed by centrifugation (10,000 × *g*, 10 minutes). Amylose resin (NEB) was added to the supernatant containing the target proteins. After incubation for 2 hours, the mixture was washed with 24 column volumes (cv) column buffer and eluted with 3 × 1 cv elution buffer (20 mM Tris–HCl, 200 mM NaCl, 1 mM EDTA, 10 mM maltose, pH 7.4). The solution containing purified protein was concentrated and desalted by protein concentrators. Target proteins were confirmed by SDS–PAGE, and purified recombinant proteins were selected for enzyme assay.

Enzyme assay was performed in Tris–HCl buffer (100 mM, pH 7.4), with 5 μM γ-chaconine, 1.5 mM UDP-glucose, 5 mM $MgCl_2$, 1 mM dithiothreitol (DTT), and 5 μg purified protein in a volume of 200 μL. After 2 hours of incubation at 30°C, the reactions were stopped by adding 1 mL of methanol. Samples were centrifuged for 15 minutes at 12,000 × *g* and analyzed by LC–MS/MS equipped with a BEH C18 column (100 Å, 5 μm, 1.7 mm × 100 mm; Waters, USA). Solvents for the mobile phase were 2 mM ammonium acetate in 100% H_2O (A) and 20% acetonitrile in methanol (B). A 45-min gradient from 5%–98% B was used for extract separation with a flow rate of 300 μL/min. Enzyme assay products were putatively identified by comparing their retention times, isotopic pattern, and MS/MS fragmentation spectrum. The same methods were used to detect the presence of solanidine-2Glu (m/z = 722.45) in the tuber flesh of *StGAME17*-overexpressing (OE-18 and OE-22) lines.

Pangenome gene graph construction

To investigate gene-level variations around *StGAME17* in a straightforward manner, we utilized the Pangene software (Li et al., 2024) to construct a pangenome gene graph. The process involved the following steps.

(1) Protein-to-genome alignment: proteins were aligned to the reference genome using Miniprot (Li, 2023) with the following command: miniprot -G 20 000 –outs = 0.97 -Iu -t 4 –outc = 0.8 DMv6.fa input1.pep.fa > out1.paf.
(2) Pangenome graph construction: the resulting alignments were used to build the pangenome graph with the following command: pangene ./pep.alignment/*paf > graph.gfa.
(3) Graph visualization: the generated graph was visualized using GFATools (https://github.com/lh3/gfatools) for further analysis.

Vector construction and transformation

To generate the expression construct, the full-length cDNA of *Soltu.DM.03G005790* and *StGAME17* were cloned and inserted into *Bam*H I and *Pml* I sites of the binary vector pCAMBIA1305.4 and transformed into the *A. tumefaciens* strain GV3101, respectively. To knock out *Soltu.DM.03G005790*, two 19-nt single-guide RNA sequences were designed: target 1 (GCGTCAGAAGCGGCGATGC) and target 2 (AGCTGTTAGTCGGTGATA). The target sequences were then inserted into the linearized CRISPR/Cas9 vector pKSE402 by digestion with the restriction enzyme *Bsa* I. The presence and correctness of the target sequence in the resulting constructs were confirmed by Sanger sequencing. *Agrobacterium*-mediated transformation of potato internodes was conducted in the background of the diploid *S. tuberosum* group Phureja S15-65 (CIP703541, PHU), as previously described with some modifications (Ye et al., 2018). Mutations in knockout transgenic plants were detected through PCR genotyping, targeting two specific sequences. The polymorphic PCR products were cloned and confirmed by sequencing. The ploidy of the transgenic plants and the non-transformed controls was assessed by flow cytometry. Tetraploid regenerants were excluded from further analysis. Primers are listed in Supplemental Table 14.

Please cite this article in press as: Lian et al., A genomic variation map provides insights into potato evolution and key agronomic traits, Molecular Plant (2025), https://doi.org/10.1016/j.molp.2025.01.016

Evaluation of average tuber weight and yield

Tuber phenotype evaluation was performed in Hohhot, Inner Mongolia (40°45′ N, 111°40′ E, 1040 masl) and Dingxi, Gansu Province (35°44′ N, 104°50′ E, 2094 masl), respectively. At least 30 replicates of *in vitro* WT and *Soltu.DM.03G005790*-overexpression (OE-13 and OE-23) lines or/and knockout (KO-90) lines were grown in the greenhouse or field (between May and October, 2023 and 2024). Phenotype evaluation was scored at harvest by measuring the average tuber weight (the three largest tubers per plant (Braun et al., 2017) and tuber yield of each plant.

Quantitative PCR

Total RNA was extracted from the samples with TRIzol (Invitrogen) or *TransZol* Plant (TransGen Biotech) and reverse transcribed using a PrimeScript RT Reagent Kit with gDNA Eraser (Takara). PCR was performed on the QuantStudio 7 Flex system using SYBR Premix (Takara) according to the manufacturer's instructions. The relative gene expression levels were calculated by the $2^{-\Delta\Delta Ct}$ method. The *ACTIN* gene was used as an internal control. Primers are listed in Supplemental Table 14.

Group-specific SNPs

Model-based clustering was performed for all samples, including the potential progenitor groups (*S. vernei*, *S. brevicaule*, *S. candolleanum*, STN, and PHU) and 4X-L. We randomly selected eight samples possessing a genetic background solely from each group (only two accessions for *S. vernei*) based on the population-structure results obtained as described above. We then reanalyzed the population structure of these selected samples using the same method to guarantee their pure genetic background. The results also indicated that admixture existed in all 4X-L accessions when $K = 3$, while these tetraploids were separated as a distinct group at $K = 4$ and 5 (Supplemental Figure 18A). Finally, a SNP was treated as group-specific if it met the following criteria: (1) the SNP showed only one type of genotype in the test group, while entirely different genotypes were observed in the remaining groups; and (2) the SNP had a missing rate < 20% in the test group.

Introgression analysis

We chose the wild species for introgression analysis by referring to the results of Hardigan et al. (2017) but excluded *S. candolleanum*, as it is the direct ancestor of landraces and may introduce false positives, and retained *Solanum berthaultii*, *Solanum chacoense*, *Solanum kurtzianum*, *Solanum microdontum*, *Solanum raphanifolium*, *S. vernei*, *Solanum chacoense*, *S. brevicaule*, and *Solanum medians*. We then employed Dsuite (v0.3) software (Malinsky et al., 2021) to scan the whole genome and detect the potential introgression regions with the command Dsuite Dtrios-t newick.file–region = start, length-o output pop.chrXX.vcf.gz setfile (start indicates the number of SNPs when windows started, and length indicates the window size; newick.file is the phylogenetic relationship between four taxons [((H1, H2) H3) H4], with H1 representing the diploid landrace group, H2 representing the tetraploid group, H3 representing the aforementioned wild species, and H4 representing the outgroup). To discover gene flow between wild accessions and the tetraploid group, we filtered the windows with the criteria: Z score ≥ 9, $p < 0.01$, D-statistic >0.20, with D-statistic = (nABBA − nBABA)/(nABBA + nBABA). Adjacent windows with strong introgression signals were merged.

Verification of SNPs using HiFi reads

We downloaded the HiFi sequencing reads (Tang et al., 2022). The reads were then aligned to the reference genome using minimap2 (v1) (Li, 2018) with the command: minimap2 -R "@RG\tID:SampleID\tLB:SampleID\tPL:HiFi\tSM:SampleID" -t 4 -ayYL –MD –eqx -x map-hifi ∼/dmv6.fa input.ccs.fastq.gz > out.sam. The resulting sam file was converted into bam format, followed by sorting and indexing using SAMtools (Li et al., 2009). Finally, variant calling was conducted with GATK (v4.2.3.0) (McKenna et al., 2010), producing a VCF file. This VCF file was used to validate the accuracy of SNPs identified through whole-genome sequencing data.

Genome-similarity analysis

To calculate genome similarity, we filtered the SNPs with MAF > 0.01. We defined the similarity score with the following formula: for a biallelic locus with alleles A and B, score (AA, AA) = 1, score (AA, BB) = 0, score (AB, xx) = 0.5, where xx is any other genotype. All nine tetraploid potato accessions were compared with selected accessions to calculate the similarity score throughout the genome.

Statistical analysis

Significant differences between data points were determined by a two-tailed Student's *t*-test.

DATA AND CODE AVAILABILITY

The raw sequencing data reported in this paper have been deposited in the Sequence Read Archive (SRA) under accession number SRA: PRJNA766763 and are publicly accessible. These data were made publicly accessible prior to the manuscript's initial submission in April 2022, and Tang et al., (2022) subsequently referred to and cited PRJNA766763 to select samples for pan-genome analysis. The raw data for the RNA-seq were deposited in the NCBI SRA under BioProject accession numbers NCBI: PRJNA1175824. The variation data reported in this paper have been deposited in the Genome Variation Map (GVM) in the National Genomics Data Center under accession number GVM: GVM000873. All code is available at https://github.com/LianqunBio/potato_pop.

FUNDING

This work was supported by funding from the National Key Research and Development Program of China (2021YFD1201400), the Inner Mongolia Autonomous Region Science and Technology Innovation Guide Award Fund, Special Project for First-Class Discipline Research of Department of Education, Inner Mongolia Autonomous Region (YLXKZX-ND-028), the Central Guidance on Local Science and Technology Development Fund (2022ZY0141), and the Program for Innovative Research Team in Universities of Inner Mongolia Autonomous Region (NMGIRT2410) to J.Q.; the Agricultural Science and Technology Innovation Program (CAAS-ZDXT201804) and Shenzhen Outstanding Talent Training Fund to S.H.; the National Natural Science Foundation of China (32372720) to J.Z.; Yunnan Science Fund (202105AF150028) to Y.S.; the National Key Research and Development Program of China (2023YFF1000100) to T.L.; and China Agriculture Research System of MOF and MARA (CARS-09-P17) and Key Technology Research Project of Inner Mongolia Autonomous Region (2020GG0054) to R.Z.

ACKNOWLEDGMENTS

We thank the US Potato Genebank (Agricultural Research Service, U.S. Department of Agriculture) and CIP (International Potato Center) for providing the potato accessions. We thank Xingtan Zhang, Chunzhi Zhang, Wei Li, Yao Zhou, Xueqing Chen, Yang Zhong (Agricultural Genome Institute at Shenzhen, Chinese Academy of Agricultural Sciences), and Philip Kear (International Potato Center (CIP), China Center for Asia Pacific) for their valuable comments. We gratefully acknowledge Xinxin Zhang, Fujin Zhang (Inner Mongolia Academy of Agricultural & Animal Husbandry Sciences), and Yuan Zhou (Huazhong University of Science and Technology) for quantification and analysis of tuber SGAs. The authors declare no competing interests.

AUTHOR CONTRIBUTIONS

S.H., J.Q., and R.Z. conceived and managed the project. Q.L., Z.Z., and T.L. performed the bioinformatics analysis. Y. Zhang, Z.Y., G. Xin, and Z.F. planted the potato accessions and prepared the samples. J.Q., Y. Zhang, M.D., Z.P., Y.P., G. Xuanyuan, S.B., Q.W., and R.H. designed and

performed the molecular experiments. J.Q., Q.L., S.H., J.Z., T.L., W.W., Z.P., Y. Zhang, L.C., R.D., and M.D. analyzed data, prepared the figures and tables, and wrote the paper. C.W.B.B., S.N., G.J.B., Y.S., X.Z., Y. Zhou, and H.L. contributed to data analysis and revised the manuscript.

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Molecular Plant Online*.

Received: July 24, 2024
Revised: December 7, 2024
Accepted: January 22, 2025

REFERENCES

Alexander, D.H., and Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinf. **12**:246. https://doi.org/10.1186/1471-2105-12-246.

Aliche, E.B., Oortwijn, M., Theeuwen, T.P.J.M., Bachem, C.W.B., van Eck, H.J., Visser, R.G.F., and van der Linden, C.G. (2019). Genetic mapping of tuber size distribution and marketable tuber yield under drought stress in potatoes. Euphytica **215**:186. https://doi.org/10.1007/s10681-019-2508-0.

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. https://doi.org/10.3732/ajb.95.2.252.

Bao, Z., Li, C., Li, G., Wang, P., Peng, Z., Cheng, L., Li, H., Zhang, Z., Li, Y., Huang, W., et al. (2022). Genome architecture and tetrasomic inheritance of autotetraploid potato. Mol. Plant **15**:1211–1226. https://doi.org/10.1016/j.molp.2022.06.009.

Barone, A., Sebastiano, A., Carputo, D., Della Rocca, F., and Frusciante, L. (2001). Molecular marker-assisted introgression of the wild *Solanum commersonii* genome into the cultivated *S. tuberosum* gene pool. Theor. Appl. Genet. **102**:900–907. https://doi.org/10.1007/s001220000498.

Beaumont, M.A. (2005). Adaptation and speciation: what can *F*st tell us? Trends Ecol. Evol. **20**:435–440. https://doi.org/10.1016/j.tree.2005.05.017.

Bisognin, D.A., Manrique-Carpintero, N.C., and Douches, D.S. (2018). QTL analysis of tuber dormancy and sprouting in potato. Am. J. Potato Res. **95**:374–382. https://doi.org/10.1007/s12230-018-9638-0.

Boccia, M., Kessler, D., Seibt, W., Grabe, V., Rodríguez López, C.E., Grzech, D., Heinicke, S., O'Connor, S.E., and Sonawane, P.D. (2024). A scaffold protein manages the biosynthesis of steroidal defense metabolites in plants. Science **386**:1366–1372. https://doi.org/10.1126/science.ado3409.

Bozan, I., Achakkagari, S.R., Anglin, N.L., Ellis, D., Tai, H.H., and Strömvik, M.V. (2023). Pangenome analyses reveal impact of transposable elements and ploidy on the evolution of potato species. Proc. Natl. Acad. Sci. USA **120**:e2211117120. https://doi.org/10.1073/pnas.2211117120.

Braun, S.R., Endelman, J.B., Haynes, K.G., and Jansky, S.H. (2017). Quantitative trait loci for resistance to common scab and cold-induced sweetening in diploid potato. Plant Genome **10**:1–10. https://doi.org/10.3835/plantgenome2016.10.0110.

Cai, D., Rodríguez, F., Teng, Y., Ané, C., Bonierbale, M., Mueller, L.A., and Spooner, D.M. (2012). Single copy nuclear gene analysis of polyploidy in wild potatoes (*Solanum* section *Petota*). BMC Evol. Biol. **12**:70. https://doi.org/10.1186/1471-2148-12-70.

Cárdenas, P.D., Sonawane, P.D., Heinig, U., Bocobza, S.E., Burdman, S., and Aharoni, A. (2015). The bitter side of the nightshades: Genomics drives discovery in *Solanaceae* steroidal alkaloid metabolism. Phytochemistry **113**:24–32. https://doi.org/10.1016/j.phytochem.2014.12.010.

Cheng, L., Wang, N., Bao, Z., et al. (2025). Leveraging a phased pangenome for haplotype design of hybrid potato. Nature. https://doi.org/10.1038/s41586-024-08476-9.

Dahlin, P., Müller, M.C., Ekengren, S., McKee, L.S., and Bulone, V. (2017). The impact of steroidal glycoalkaloids on the physiology of Phytophthora infestans, the causative agent of potato late blight. Mol. Plant Microbe Interact. **30**:531–542. https://doi.org/10.1094/MPMI-09-16-0186-R.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. Bioinformatics **27**:2156–2158. https://doi.org/10.1093/bioinformatics/btr330.

Destefanis, M., Nagy, I., Rigney, B., Bryan, G.J., McLean, K., Hein, I., Griffin, D., and Milbourne, D. (2015). A disease resistance locus on potato and tomato chromosome 4 exhibits a conserved multipartite structure displaying different rates of evolution in different lineages. BMC Plant Biol. **15**:255. https://doi.org/10.1186/s12870-015-0645-8.

Distl, M., and Wink, M. (2009). Identification and quantification of steroidal alkaloids from wild tuber-bearing *Solanum* species by HPLC and LC-ESI-MS. Potato Res. **52**:79–104. https://doi.org/10.1007/s11540-008-9123-0.

Doebley, J.F., Gaut, B.S., and Smith, B.D. (2006). The molecular genetics of crop domestication. Cell **127**:1309–1321. https://doi.org/10.1016/j.cell.2006.12.006.

Du, M., Wang, T., Lian, Q., Zhang, X., Xin, G., Pu, Y., Bryan, G.J., and Qi, J. (2021). Developing a new model system for potato genetics by androgenesis. J. Integr. Plant Biol. **63**:628–633. https://doi.org/10.1111/jipb.13018.

Dudenhoefer, D. (2019). Discovery to Impact Science-Based Solutions for Global Challenges (International Potato Center).

Endelman, J.B., and Jansky, S.H. (2016). Genetic mapping with an inbred line-derived F_2 population in potato. Theor. Appl. Genet. **129**:935–943. https://doi.org/10.1007/s00122-016-2673-7.

Fagundes, M.C.P., Oliveira, A.F.d., Carvalho, V.L.d., Ramos, J.D., Santos, V.A.d., and Rufini, J.C.M. (2018). Alternative control of plant pathogen fungi through ethanolic extracts of Avocado Seeds (Persea Americana Mill.). Braz. Arch. Biol. Technol. **61**:e18180052. https://doi.org/10.1590/1678-4324-2018180052.

Fang, Y., Qin, X., Liao, Q., Du, R., Luo, X., Zhou, Q., Li, Z., Chen, H., Jin, W., Yuan, Y., et al. (2022). The genome of homosporous maidenhair fern sheds light on the euphyllophyte evolution and defences. Nat. Plants **8**:1024–1037. https://doi.org/10.1038/s41477-022-01222-x.

Fernie, A.R., and Willmitzer, L. (2001). Molecular and biochemical triggers of potato tuber development. Plant Physiol. **127**:1459–1465. https://doi.org/10.1104/pp.010764.

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**:83–111. https://doi.org/10.1007/BF00026800.

Gawel, N.J., and Jarret, R.L. (1991). A modified CTAB DNA extraction procedure for *Musa* and *Ipomoea*. Plant Mol. Biol. Rep. **9**:262–266. https://doi.org/10.1007/bf02672076.

Gros-Balthazard, M., Besnard, G., Sarah, G., Holtz, Y., Leclercq, J., Santoni, S., Wegmann, D., Glémin, S., and Khadari, B. (2019). Evolutionary transcriptomics reveals the origins of olives and the genomic changes associated with their domestication. Plant J. **100**:143–157. https://doi.org/10.1111/tpj.14435.

Gross, B.L., and Strasburg, J.L. (2010). Cotton domestication: dramatic changes in a single cell. BMC Biol. **8**:137. https://doi.org/10.1186/1741-7007-8-137.

Grun, P. (1990). The evolution of cultivated potatoes. Econ. Bot. **44**:39–55. https://doi.org/10.1007/bf02860474.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. **59**:307–321. https://doi.org/10.1093/sysbio/syq010.

Gupta, P.K., and Tsuchiya, T. (1991). Chromosome Engineering in Plants: Genetics, Breeding, Evolution (Elsevier). https://doi.org/10.1016/c2009-0-00654-0.

Smith, J.M., and Haigh, J. (1974). The hitch-hiking effect of a favourable gene. Genet. Res. **23**:23–35. https://doi.org/10.1017/s0016672300014634.

Hardigan, M.A., Laimbeer, F.P.E., Newton, L., Crisovan, E., Hamilton, J.P., Vaillancourt, B., Wiegert-Rininger, K., Wood, J.C., Douches, D.S., Farré, E.M., 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. https://doi.org/10.1073/pnas.1714380114.

Hardigan, M.A., Crisovan, E., Hamilton, J.P., Kim, J., Laimbeer, P., Leisner, C.P., Manrique-Carpintero, N.C., Newton, L., Pham, G.M., Vaillancourt, B., et al. (2016). Genome reduction uncovers a large dispensable genome and adaptive role for copy number variation in asexually propagated *Solanum tuberosum*. Plant Cell **28**:388–405. https://doi.org/10.1105/tpc.15.00538.

Hawkes, J.G. (1988). The evolution of cultivated potatoes and their tuber-bearing wild relatives. Kulturpflanze **36**:189–208. https://doi.org/10.1007/bf01986960.

Hawkes, J.G. (1990). The Potato: Evolution, Biodiversity and Genetic Resources (Smithsonian Institution Press).

Hellens, R.P., Allan, A.C., Friel, E.N., Bolitho, K., Grafton, K., Templeton, M.D., Karunairetnam, S., Gleave, A.P., and Laing, W.A. (2005). Transient expression vectors for functional genomics, quantification of promoter activity and RNA silencing in plants. Plant Methods **1**:1–14. https://doi.org/10.1186/1746-4811-1-13.

Hirsch, C.N., Hirsch, C.D., Felcher, K., Coombs, J., Zarka, D., Van Deynze, A., De Jong, W., Veilleux, R.E., Jansky, S., Bethke, P., et al. (2013). Retrospective view of North American potato (*Solanum tuberosum* L.) breeding in the 20th and 21st centuries. G3 (Bethesda). **3**:1003–1013. https://doi.org/10.1534/g3.113.005595.

Hoopes, G., Meng, X., Hamilton, J.P., Achakkagari, S.R., de Alves Freitas Guesdes, F., Bolger, M.E., Coombs, J.J., Esselink, D., Kaiser, N.R., Kodde, L., 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. https://doi.org/10.1016/j.molp.2022.01.003.

Hosaka, K. (1986). Who is the mother of the potato? - restriction endonuclease analysis of chloroplast DNA of cultivated potatoes. Theor. Appl. Genet. **72**:606–618. https://doi.org/10.1007/bf00288998.

Huamán Z, and Spooner, D.M. (2002). Reclassification of landrace populations of cultivated potatoes (*Solanum* sect. *Petota*). Am. J. Bot. **89**:947–965. https://doi.org/10.3732/ajb.89.6.947.

Itkin, M., Heinig, U., Tzfadia, O., Bhide, A.J., Shinde, B., Cardenas, P.D., Bocobza, S.E., Unger, T., Malitsky, S., Finkers, R., et al. (2013). Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes. Science **341**:175–179. https://doi.org/10.1126/science.1240230.

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. https://doi.org/10.1007/bf02410100.

Kang, H.M., Sul, J.H., Service, S.K., Zaitlen, N.A., Kong, S.-y., Freimer, N.B., Sabatti, C., and Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. Nat. Genet. **42**:348–354. https://doi.org/10.1038/ng.548.

Kantar, M.B., Nashoba, A.R., Anderson, J.E., Blackman, B.K., and Rieseberg, L.H. (2017). The genetics and genomics of plant domestication. Bioscience **67**:971–982. https://doi.org/10.1093/biosci/bix114.

Kienow, L., Schneider, K., Bartsch, M., Stuible, H.-P., Weng, H., Miersch, O., Wasternack, C., and Kombrink, E. (2008). Jasmonates meet fatty acids: functional analysis of a new acyl-coenzyme A synthetase family from *Arabidopsis thaliana*. J. Exp. Bot. **59**:403–419. https://doi.org/10.1093/jxb/erm325.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods **12**:357–360. https://doi.org/10.1038/nmeth.3317.

Kloosterman, B., Vorst, O., Hall, R.D., Visser, R.G.F., and Bachem, C.W. (2005). Tuber on a chip: differential gene expression during potato tuber development. Plant Biotechnol. J. **3**:505–519. https://doi.org/10.1111/j.1467-7652.2005.00141.x.

Knowles, N.R., and Knowles, L.O. (2006). Manipulating stem number, tuber set, and yield relationships for northern- and southern-grown potato seed lots. Crop Sci. **46**:284–296. https://doi.org/10.2135/cropsci2005.05-0078.

Koenig, D., Jiménez-Gómez, J.M., Kimura, S., Fulop, D., Chitwood, D.H., Headland, L.R., Kumar, R., Covington, M.F., Devisetty, U.K., Tat, A.V., et al. (2013). Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. Proc. Natl. Acad. Sci. USA **110**:E2655–E2662. https://doi.org/10.1073/pnas.1309606110.

Kowalski, S.P., Domek, J.M., Sanford, L.L., and Deahl, K.L. (2000). Effect of α-tomatine and tomatidine on the growth and development of the Colorado Potato Beetle (Coleoptera: Chrysomelidae): studies using synthetic diets. J. Entomol. Sci. **35**:290–300. https://doi.org/10.18474/0749-8004-35.3.290.

Kozukue, N., Misoo, S., Yamada, T., Kamijima, O., and Friedman, M. (1999). Inheritance of morphological characters and glycoalkaloids in potatoes of somatic hybrids between dihaploid *Solanum acauleand* tetraploid *Solanum tuberosum*. J. Agric. Food Chem. **47**:4478–4483. https://doi.org/10.1021/jf990252n.

Kozukue, N., Yoon, K.-S., Byun, G.-I., Misoo, S., Levin, C.E., and Friedman, M. (2008). Distribution of glycoalkaloids in potato tubers of 59 accessions of two wild and five cultivated *Solanum* Species. J. Agric. Food Chem. **56**:11920–11928. https://doi.org/10.1021/jf802631t.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinf. **9**:559. https://doi.org/10.1186/1471-2105-9-559.

Lata, C., and Prasad, M. (2011). Role of DREBs in regulation of abiotic stress responses in plants. J. Exp. Bot. **62**:4731–4748. https://doi.org/10.1093/jxb/err210.

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., Rouzé, P., and Rombauts, S. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res. **30**:325–327. https://doi.org/10.1093/nar/30.1.325.

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics **34**:3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

Li, H. (2023). Protein-to-genome alignment with miniprot. Bioinformatics **39**:btad014. https://doi.org/10.1093/bioinformatics/btad014.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics **25**:1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

Li, H., Marin, M., and Farhat, M.R. (2024). Exploring gene content with pangene graphs. Bioinformatics **40**:btae456. https://doi.org/10.1093/bioinformatics/btae456.

Li, H., Yang, X., Shang, Y., Zhang, Z., and Huang, S. (2023a). Vegetable biology and breeding in the genomics era. Sci. China Life Sci. **66**:226–250. https://doi.org/10.1007/978-981-19-5367-5_9.

Li, H., Li, M., Fan, Y., Liu, Y., and Qin, S. (2023b). Antifungal activity of potato glycoalkaloids and its potential to control severity of dry rot caused by *Fusarium sulphureum*. Crop Sci. **63**:801–811. https://doi.org/10.1002/csc2.20874.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics **25**:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

Li, Y., Colleoni, C., Zhang, J., Liang, Q., Hu, Y., Ruess, H., Simon, R., Liu, Y., Liu, H., Yu, G., et al. (2018). Genomic analyses yield markers for identifying agronomically important genes in potato. Mol. Plant **11**:473–484. https://doi.org/10.1016/j.molp.2018.01.009.

Lin, T., Zhu, G., Zhang, J., Xu, X., Yu, Q., Zheng, Z., Zhang, Z., Lun, Y., Li, S., Wang, X., et al. (2014). Genomic analyses provide insights into the history of tomato breeding. Nat. Genet. **46**:1220–1226. https://doi.org/10.1038/ng.3117.

Liu, S., Yang, J., Zhang, N., and Si, H. (2024a). Genome-wide analysis of non-coding RNA reveals the role of a novel miR319c for tuber dormancy release process in potato. Horticulture Research **46**:uhae303. https://doi.org/10.1038/ng.3117.

Liu, W., Chen, L., Zhang, S., Hu, F., Wang, Z., Lyu, J., Wang, B., Xiang, H., Zhao, R., Tian, Z., et al. (2019). Decrease of gene expression diversity during domestication of animals and plants. BMC Evol. Biol. **19**:19. https://doi.org/10.1186/s12862-018-1340-9.

Liu, Z., Wang, N., Su, Y., Long, Q., Peng, Y., Shangguan, L., Zhang, F., Cao, S., Wang, X., Ge, M., et al. (2024b). Grapevine pangenome facilitates trait genetics and genomic breeding. Nat. Genet. **56**:2804–2814. https://doi.org/10.1038/s41588-024-01967-5.

Lotterhos, K.E., and Whitlock, M.C. (2014). Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. Mol. Ecol. **23**:2178–2192. https://doi.org/10.1111/mec.12725.

Louderback, L.A., and Pavlik, B.M. (2017). Starch granule evidence for the earliest potato use in North America. Proc. Natl. Acad. Sci. USA **114**:7606–7610. https://doi.org/10.1073/pnas.1705540114.

Lu, X., Li, Q.-T., Xiong, Q., Li, W., Bi, Y.-D., Lai, Y.-C., Liu, X.-L., Man, W.-Q., Zhang, W.-K., Ma, B., et al. (2016). The transcriptomic signature of developing soybean seeds reveals the genetic basis of seed trait adaptation during domestication. Plant J. **86**:530–544. https://doi.org/10.1111/tpj.13181.

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. https://doi.org/10.1111/1755-0998.13265.

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**:317–332. https://doi.org/10.1534/genetics.119.302036.

McCue, K.F., Allen, P.V., Shepherd, L.V.T., Blake, A., Whitworth, J., Maccree, M.M., Rockhold, D.R., Stewart, D., Davies, H.V., and Belknap, W.R. (2006). The primary in vivo steroidal alkaloid glucosyltransferase from potato. Phytochemistry **67**:1590–1597. https://doi.org/10.1016/j.phytochem.2005.09.037.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. **20**:1297–1303. https://doi.org/10.1101/gr.107524.110.

Meyer, R.S., and Purugganan, M.D. (2013). Evolution of crop species: genetics of domestication and diversification. Nat. Rev. Genet. **14**:840–852. https://doi.org/10.1038/nrg3605.

Mihovilovich, E., Carli, C., Mendiburu, F.D., Hualla, V., and Bonierbale, M. (2014). Protocol for tuber bulking maturity assessment of elite and advanced potato clones (International Potato Center).

Navarre, R., and Pavek, M. (2014). The Potato: Botany, Production and Uses (CABI).

Navarro, C., Abelenda, J.A., Cruz-Oró, E., Cuéllar, C.A., Tamaki, S., Silva, J., Shimamoto, K., and Prat, S. (2011). Control of flowering and storage organ formation in potato by FLOWERING LOCUS T. Nature **478**:119–122. https://doi.org/10.1038/nature10431.

Osman, S.F., Herb, S.F., Fitzpatrick, T.J., and Schmiediche, P. (1978). Glycoalkaloid composition of wild and cultivated tuber-bearing Solanum species of potential value in potato breeding programs. J. Agric. Food Chem. **26**:1246–1248. https://doi.org/10.1021/jf60219a024.

Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. PLoS Genet. **2**:e190. https://doi.org/10.1371/journal.pgen.0020190.

Paudel, J.R., Davidson, C., Song, J., Maxim, I., Aharoni, A., and Tai, H.H. (2017). Pathogen and Pest Responses Are Altered Due to RNAi-Mediated Knockdown of GLYCOALKALOID METABOLISM 4 in Solanum tuberosum. Mol. Plant Microbe Interact. **30**:876–885. https://doi.org/10.1094/MPMI-02-17-0033-R.

Pavlidis, P., Živkovic, D., Stamatakis, A., and Alachiotis, N. (2013). SweeD: Likelihood-based detection of selective sweeps in thousands of genomes. Mol. Biol. Evol. **30**:2224–2234. https://doi.org/10.1093/molbev/mst112.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. **33**:290–295. https://doi.org/10.1038/nbt.3122.

Pham, G.M., Hamilton, J.P., Wood, J.C., Burke, J.T., Zhao, H., Vaillancourt, B., Ou, S., Jiang, J., and Buell, C.R. (2020). Construction of a chromosome-scale long-read reference genome assembly for potato. GigaScience **9**:giaa100. https://doi.org/10.1093/gigascience/giaa100.

Ranjan, R., and Lewak, S. (1992). Jasmonic acid promotes germination and lipase activity in non-stratified apple embryos. Physiol. Plantarum **86**:335–339. https://doi.org/10.1034/j.1399-3054.1992.860222.x.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **26**:139–140. https://doi.org/10.1093/bioinformatics/btp616.

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. https://doi.org/10.1007/s00122-010-1422-6.

Ross, H. (1966). The use of wild *solanum* species in German potato breeding of the past and today. Am. Potato J. **43**:63–80. https://doi.org/10.1007/bf02861579.

Schreiber, L., Nader-Nieto, A.C., Schönhals, E.M., Walkemeier, B., and Gebhardt, C. (2014). SNPs in genes functional in starch-sugar interconversion associate with natural variation of tuber starch

Please cite this article in press as: Lian et al., A genomic variation map provides insights into potato evolution and key agronomic traits, Molecular Plant (2025), https://doi.org/10.1016/j.molp.2025.01.016

and sugar content of potato (Solanum tuberosum L.). G3: Genes|Genomes|Genetics **4**:1797–1811. https://doi.org/10.1534/g3.114.012377.

Schrider, D.R., Shanku, A.G., and Kern, A.D. (2016). Effects of linked selective sweeps on demographic inference and model selection. Genetics **204**:1207–1223. https://doi.org/10.1534/genetics.116.190223.

Shubha, K., and Singh, D. (2018). Selection of yield-associated morphological and biochemical traits using correlation and path coefficient analysis in potato (*Solanum tuberosum* L.) in the foothills of north-western himalayas. Potato Res. **61**:273–281. https://doi.org/10.1007/s11540-018-9376-1.

Sinden, S.L., Sanford, L.L., Cantelo, W.W., and Deahl, K.L. (1986). Leptine Glycoalkaloids and Resistance to the Colorado Potato Beetle (Coleoptera: Chrysomelidae) in *Solanum chacoense*. Environ. Entomol. **15**:1057–1062. https://doi.org/10.1093/ee/15.5.1057.

Śliwka, J., Sołtys-Kalina, D., Szajko, K., Wasilewicz-Flis, I., Strzelczyk-Żyta, D., Zimnoch-Guzowska, E., Jakuczun, H., and Marczewski, W. (2016). Mapping of quantitative trait loci for tuber starch and leaf sucrose contents in diploid potato. Theor. Appl. Genet. **129**:131–140. https://doi.org/10.1007/s00122-015-2615-9.

Sołtys-Kalina, D., Szajko, K., Stefańczyk, E., Smyda-Dajmund, P., Śliwka, J., and Marczewski, W. (2020). eQTL mapping of the 12S globulin cruciferin gene *PGCRURSE5* as a novel candidate associated with starch content in potato tubers. Sci. Rep. **10**:17168. https://doi.org/10.1038/s41598-020-74285-5.

Spooner, D.M., and Castillo, T.R. (1997). Reexamination of series relationships of South American wild potatoes (Solanaceae: Solanum sect. Petota): evidence from chloroplast DNA restriction site variation. Am. J. Bot. **84**:671–685. https://doi.org/10.2307/2445904.

Spooner, D.M., McLean, K., Ramsay, G., Waugh, R., and Bryan, G.J. (2005). A single domestication for potato based on multilocus amplified fragment length polymorphism genotyping. Proc. Natl. Acad. Sci. USA **102**:14694–14699. https://doi.org/10.1073/pnas.0507400102.

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

Sukhotu, T., and Hosaka, K. (2006). Origin and evolution of Andigena potatoes revealed by chloroplast and nuclear DNA markers. Genome **49**:636–647. https://doi.org/10.1139/g06-014.

Sun, H., Jiao, W.B., Krause, K., Campoy, J.A., Goel, M., Folz Donahue, K., Kukat, C., Huettel, B., and Schneeberger, K. (2022). Chromosome-scale and haplotype-resolved genome assembly of a tetraploid potato cultivar. Nat. Genet. **54**:342–348. https://doi.org/10.1038/s41588-022-01015-0.

Suttle, J.C., Huckle, L.L., and Lulai, E.C. (2011). The effects of dormancy status on the endogenous contents and biological activities of jasmonic acid, N-(jasmonoyl)-isoleucine, and tuberonic acid in potato tubers. Am. J. Potato Res. **88**:283–293. https://doi.org/10.1007/s12230-011-9192-5.

Swanson-Wagner, R., Briskine, R., Schaefer, R., Hufford, M.B., Ross-Ibarra, J., Myers, C.L., Tiffin, P., and Springer, N.M. (2012). Reshaping of the maize transcriptome by domestication. Proc. Natl. Acad. Sci. USA **109**:11878–11883. https://doi.org/10.1073/pnas.1201961109.

Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoka, C., Uemura, A., Utsushi, H., Tamiru, M., Takuno, S., et al. (2013). QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. Plant J. **74**:174–183. https://doi.org/10.1111/tpj.12105.

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. https://doi.org/10.1093/molbev/mst197.

Tang, D., Jia, Y., Zhang, J., Li, H., Cheng, L., Wang, P., Bao, Z., Liu, Z., Feng, S., Zhu, X., et al. (2022). Genome evolution and diversity of wild and cultivated potatoes. Nature **606**:535–541. https://doi.org/10.1038/s41586-022-04822-x.

Tang, H., Krishnakumar, V., Zeng, X., Xu, Z., Taranto, A., Lomas, J.S., Zhang, Y., Huang, Y., Wang, Y., Yim, W.C., et al. (2024). JCVI: A versatile toolkit for comparative genomics analysis. iMeta **3**:e211. https://doi.org/10.1002/imt2.211.

Ugent, D., Dillehay, T., and Ramirez, C. (1987). Potato remains from a late pleistocene settlement in southcentral Chile. Econ. Bot. **41**:17–27. https://doi.org/10.1007/bf02859340.

Väänänen, T., Ikonen, T., Rokka, V.-M., Kuronen, P., Serimaa, R., and Ollilainen, V. (2005). Influence of incorporated wild *Solanum* genomes on potato properties in terms of starch nanostructure and glycoalkaloid content. J. Agric. Food Chem. **53**:5313–5325. https://doi.org/10.1021/jf0501342.

Valkonen, J.P.T., Keskitalo, M., Vasara, T., Pietilä, L., and Raman, K.V. (1996). Potato Glycoalkaloids: A Burden or a Blessing? Crit. Rev. Plant Sci. **15**:1–20. https://doi.org/10.1080/07352689609701934.

Van Der Vossen, E., Sikkema, A., Hekkert, B.t.L., Gros, J., Stevens, P., Muskens, M., Wouters, D., Pereira, A., Stiekema, W., and Allefs, S. (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. https://doi.org/10.1046/j.1365-313x.2003.01934.x.

Van Gelder, W.M.J., and Scheffer, J.J.C. (1991). Transmission of steroidal glycoalkaloids from *Solanum vernei* to the cultivated potato. Phytochemistry **30**:165–168. https://doi.org/10.1016/0031-9422(91)84118-c.

Van Harsselaar, J.K., Lorenz, J., Senning, M., Sonnewald, U., and Sonnewald, S. (2017). Genome-wide analysis of starch metabolism genes in potato (*Solanum tuberosum* L.). BMC Genom. **18**:37. https://doi.org/10.1186/s12864-016-3381-z.

Watanabe, K. (2015). Potato genetics, genomics, and applications. Breed Sci. **65**:53–68. https://doi.org/10.1270/jsbbs.65.53.

Wei, S., Li, X., Lu, Z., Zhang, H., Ye, X., Zhou, Y., Li, J., Yan, Y., Pei, H., Duan, F., et al. (2022). A transcriptional regulator that boosts grain yields and shortens the growth duration of rice. Science **377**:eabi8455. https://doi.org/10.1126/science.abi8455.

Wolters, P.J., Wouters, D., Tikunov, Y.M., Ayilalath, S., Kodde, L.P., Strijker, M.F., Caarls, L., Visser, R.G.F., and Vleeshouwers, V.G.A.A. (2023). Tetraose steroidal glycoalkaloids from potato provide resistance against *Alternaria solani* and Colorado potato beetle. Elife **12**:RP87135. https://doi.org/10.7554/eLife.87135.

Xu, Z.-S., Chen, M., Li, L.-C., and Ma, Y.-Z. (2011). Functions and application of the AP2/ERF transcription factor family in crop improvement. J. Integr. Plant Biol. **53**:570–585. https://doi.org/10.1111/j.1744-7909.2011.01062.x.

Ye, M., Peng, Z., Tang, D., Yang, Z., Li, D., Xu, Y., Zhang, C., and Huang, S. (2018). Generation of self-compatible diploid potato by knockout of *S-RNase*. Nat. Plants **4**:651–654. https://doi.org/10.1038/s41477-018-0218-6.

Yildiz, K., Muradoğlu, F., and Yilmaz, H.A. (2008). The effect of jasmonic acid on germination of dormant and nondormant pear (*Pyrus communis* L.) seeds. Seed Sci. Technol. **36**:569–574. https://doi.org/10.15258/sst.2008.36.3.06.

Zhang, C., Dong, S.S., Xu, J.Y., He, W.M., and Yang, T.L. (2019a). PopLDdecay: a fast and effective tool for linkage disequilibrium

decay analysis based on variant call format files. Bioinformatics **35**:1786–1788. https://doi.org/10.1093/bioinformatics/bty875.

Zhang, C., Wang, P., Tang, D., Yang, Z., Lu, F., Qi, J., Tawari, N.R., Shang, Y., Li, C., and Huang, S. (2019b). The genetic basis of inbreeding depression in potato. Nat. Genet. **51**:374–378. https://doi.org/10.1038/s41588-018-0319-1.

Zhang, C., Yang, Z., Tang, D., Zhu, Y., Wang, P., Li, D., Zhu, G., Xiong, X., Shang, Y., Li, C., and Huang, S. (2021). Genome design of hybrid potato. Cell **184**:3873–3883.e12. https://doi.org/10.1016/j.cell.2021.06.006.

Zhang, F., Qu, L., Gu, Y., Xu, Z.H., and Xue, H.W. (2022). Resequencing and genome-wide association studies of autotetraploid potato. Mol. Hortic. **2**:6. https://doi.org/10.1186/s43897-022-00027-y.

Zhao, D.K., Zhao, Y., Chen, S.Y., and Kennelly, E.J. (2021). *Solanum steroidal glycoalkaloids*: *structural diversity*, *biological activities*, *and biosynthesis*. Nat. Prod. Rep. **38**:1423–1444. https://doi.org/10.1039/d1np00001b.

Zhao, G., Lian, Q., Zhang, Z., Fu, Q., He, Y., Ma, S., Ruggieri, V., Monforte, A.J., Wang, P., Julca, I., et al. (2019). A comprehensive genome variation map of melon identifies multiple domestication events and loci influencing agronomic traits. Nat. Genet. **51**:1607–1615. https://doi.org/10.1038/s41588-019-0522-8.

Zhou, Y., Liu, J., Guo, J., Wang, Y., Ji, H., Chu, X., Xiao, K., Qi, X., Hu, L., Li, H., et al. (2022). *GmTDN1* improves wheat yields by inducing dual tolerance to both drought and low-N stress. Plant Biotechnol. J. **20**:1606–1621. https://doi.org/10.1111/pbi.13836.

Zrenner, R., Krause, K.P., Apel, P., and Sonnewald, U. (1996). Reduction of the cytosolic fructose-1, 6-bisphosphatase in transgenic potato plants limits photosynthetic sucrose biosynthesis with no impact on plant growth and tuber yield. Plant J. **9**:671–681. https://doi.org/10.1046/j.1365-313x.1996.9050671.x.