



OPEN ACCESS

EDITED BY

Yi-Hong Wang,
University of Louisiana at Lafayette,
United States

REVIEWED BY

Laura M. Shannon,
University of Minnesota Twin Cities,
United States
Jinping Zhao,
Texas A and M University, United States

*CORRESPONDENCE

Julien Leuenberger
✉ Julien.leuenberger@inrae.fr

RECEIVED 09 February 2024

ACCEPTED 17 April 2024

PUBLISHED 17 May 2024

CITATION

Leuenberger J, Sharma SK, McLean K, Pellé R, Bérard A, Lesage M-L, Porhel D, Dantec M-A, Chauvin J-E, Bryan GJ, Pilet-Nayel M-L, Kerlan M-C and Esnault F (2024) A genomic dataset integrating genotyping-by-sequencing, SolCAP array and PCR marker data on tetraploid potato advanced breeding lines. *Front. Plant Sci.* 15:1384401. doi: 10.3389/fpls.2024.1384401

COPYRIGHT

© 2024 Leuenberger, Sharma, McLean, Pellé, Bérard, Lesage, Porhel, Dantec, Chauvin, Bryan, Pilet-Nayel, Kerlan and Esnault. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A genomic dataset integrating genotyping-by-sequencing, SolCAP array and PCR marker data on tetraploid potato advanced breeding lines

Julien Leuenberger^{1,2*}, Sanjeev Kumar Sharma³, Karen McLean³, Roland Pellé¹, Aurélie Bérard⁴, Marie-Laure Lesage¹, Danièle Porhel¹, Marie-Ange Dantec¹, Jean-Eric Chauvin¹, Glenn J. Bryan³, Marie-Laure Pilet-Nayel¹, Marie-Claire Kerlan¹ and Florence Esnault¹

¹Institut de Génétique, Environnement et Protection des Plantes (IGEPP), INRAE, Institut Agro, Univ Rennes, Ploudaniel, France, ²Association des Créateurs de Variétés Nouvelle de Pomme de Terre (ACVNPT), Hanvec, France, ³Cell & Molecular Science Department, The James Hutton Institute, Dundee, United Kingdom, ⁴Univ Paris Saclay, INRAE, Evry, France

KEYWORDS

potato, pre-breeding clones, genomic dataset, genotyping by sequencing, SolCAP DNA array, analysis pipeline

1 Introduction

Developing new potato varieties with enhanced resistance to diseases and pests is essential for ensuring food security and promoting sustainable agricultural production. INRAE has been creating potato advanced breeding lines for over 50 years by crossing different wild potato species with *Solanum tuberosum* to improve various traits, such as resistance to major pests and diseases or quality traits. These breeding lines have been distributed to French potato breeders, contributing significantly to the development of more resilient potato varieties. However, further research progress is needed to breed durable multi-resistant potato varieties adapted to low-input (mainly pesticides and fertilizer) agriculture. Molecular markers can effectively assist in pyramiding multiple loci for resistance (Pilet-Nayel et al., 2017) to major diseases and pests of potato, such as late blight disease, cyst nematodes, and viruses, while preserving the agronomic and quality values of varieties.

The development of genomics and the evolution of sequencing methods in the last decade have facilitated the identification of a large number of SNPs (Single Nucleotide polymorphism) from sequencing data. SNPs discovered using EST (Expressed Sequence Tag) or transcriptome data from six potato varieties were used to construct the widely used Infinium 8K SNP SolCAP array (Hamilton et al., 2011; Felcher et al., 2012). However, this genotyping method may display ascertainment bias especially for germplasm containing

significant introgressions from wild species genomes. More recently, genotyping-by-sequencing (GBS) approaches were developed, utilizing genome complexity reduction methods. GBS allows the detection of unique SNPs, specific for the variants present within the studied panel, thus offering advantages over the SolCAP array. However, GBS also has some limitations, such as potential biases in genome representation, lower accuracy in detecting rare alleles, and challenges in handling missing data (Negro et al., 2019). Both genotyping approaches are therefore complementary.

Genome-wide association studies (GWAS) have emerged as a powerful tool for identifying genes and genomic regions underlying complex agronomic traits. In potato, Marhadour and Prodhomme (2023) reviewed GWAS studies for the identification of genes and genomic regions involved in resistance to different diseases and pests, or in diverse quality traits. The panel used in this study gathers 285 INRAE tetraploid potato advanced clones originating from different breeding programs mainly focusing on disease and pest resistance. It is therefore a unique set of plant material for performing joint genetic analysis of resistance to multiple diseases and pests and studying its genetic relationships with quality traits. The panel, hereafter, will be referred to as potato MRQ (Multi-Resistance and Quality) panel.

In this study, we developed a potato genomic dataset obtained on the potato MRQ panel using three genotyping techniques: GBS, Infinium 8K SNP SolCAP array and PCR-based markers. This dataset provides a valuable resource for potato researchers and breeders to carry out GWAS analysis on multiple traits, with a particular focus on resistance to multiple diseases and pests.

2 Materials and methods

2.1 Plant material and DNA extraction

The potato MRQ panel contains 258 advanced breeding clones, 16 commercial varieties and 11 late blight R-gene differentials of Black (originating from *S. demissum*) (Black et al., 1953). The breeding clones were created in various INRAE research programs aimed at (1) enhancing resistance to different pathogens or diseases, including *Phytophthora infestans*, cyst nematodes *Globodera pallida* and *G. rostochiensis*, root knot nematode *Meloidogyne incognita*, *Fusarium* spp, *Pectobacterium*, Potato Leaf Roll Virus (PLRV), and Potato Virus Y (PVY) and (2) improving quality traits such as cold-induced sweetening or after cooking darkening.

Total genomic DNA was extracted from frozen leaves using a protocol derived from Doyle and Doyle (1990).

2.2 Genotyping-by-sequencing

2.2.1 GBS library preparation and sequencing

DNA samples were quantified using the Quant-iT™ PicoGreen® dsDNA Assay Kit (Invitrogen, San Diego, CA). GBS libraries were prepared according to the procedure described by Sharma et al. (2024). Briefly, DNA samples (100 ng each) from the studied panel were processed for double restriction enzyme

digestion (*Pst*I-*Mse*I) followed by adapter ligation in a 96-plex (*i.e.*, 96 samples per GBS library) multiplexing format. A single *Mse*I adapter was used as common adapter for all samples whereas *Pst*I adapters were barcoded and unique for each sample per library. Processed samples in each GBS library were pooled, PCR-amplified and size-selected (300bp-500bp) using BluePippin™ (Sage Science Inc., Beverly, MA). Quality checks and fragment size analysis on pooled library samples were performed using Bioanalyzer High-Sensitivity DNA chip (Agilent technologies). Each GBS library was sequenced on a single lane of Illumina HiSeq 4k platform to generate 150 bp paired-end sequence reads.

2.2.2 Quality control of raw sequencing data

All FASTQ files were inspected using FastQC to identify any anomalies. The FastQC analysis results revealed no residual sequencing adapters and displayed patterns consistent with libraries generated from DNA subjected to restriction digestion.

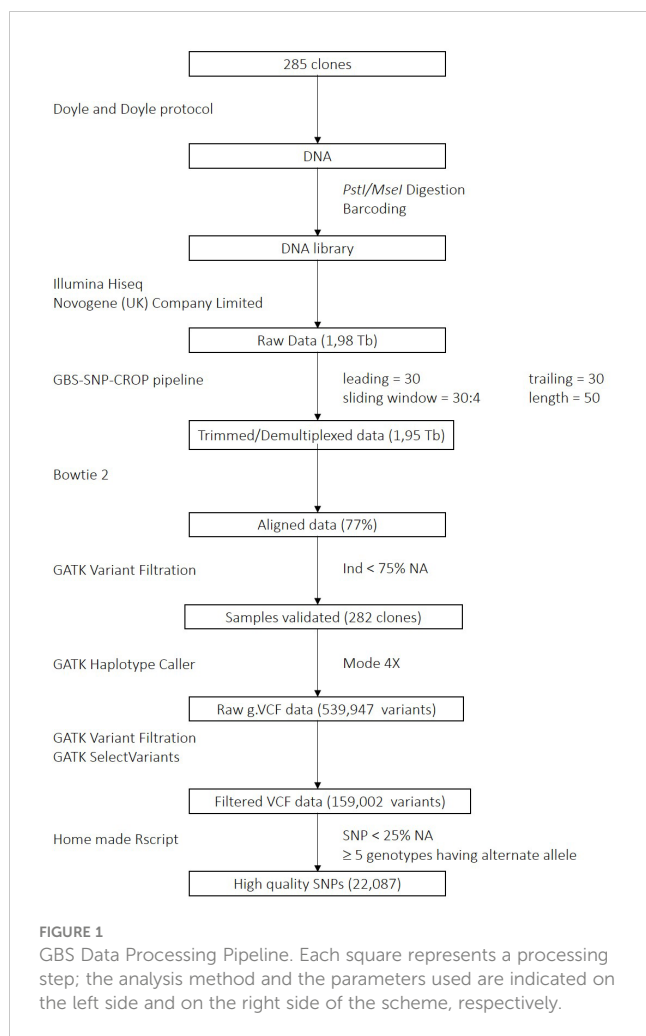
2.2.3 Demultiplexing and SNP calling

We demultiplexed the raw reads and trimmed the adapters using the GBS SNP crop pipeline (Melo et al., 2016) with the following parameters: leading = 30, trailing = 30, sliding window = 30:4, and minimum length = 50. Then, sequences were aligned in paired-end mode to the reference genome *Solanum tuberosum* group *Phureja* DM 1-3 516 R44 v6.1 using bowtie2 (Langmead and Salzberg, 2012) v2.4.5 with very sensitive mode. Next, we called variants using GATK v4.2.6.1 HaplotypeCaller in mode 4X (Van der Auwera and O'Connor, 2020). Samples with more than 75% of missing data were removed. We then applied filters for biallelic SNPs, QUAL (variant quality) > 30, QD (depth adjusted quality) > 2, GQ (genotype quality) > 20, and DP (coverage depth) > 20. SNPs were then extracted as a matrix with genotypes coded from 0 to 4 based on the dosage of the alternate allele. To further improve the usability of the variants, we filtered for a minimum of 5 genotypes having the alternate allele and removed markers with more than 25% of missing data. The workflow is presented in Figure 1.

2.3 SolCAP array genotyping and data analysis

A comprehensive quality control (QC) process was conducted for all DNA samples, by examining absorbance and electrophoretic profiles on agarose gels. Additionally, the concentration of each sample was determined using fluorometric measurements with Quant-iT™ PicoGreen® (Invitrogen). Following these assessments, samples were normalized to 50 ng/μl in 96-well plates.

The Infinium 8K SNP array (Felcher et al., 2012) was used following the Infinium HD Assay Ultra Protocol (Illumina, Inc). For each SNP, the assignment of one of the five possible genotypes (AAAA, AAAB, AABB, ABBB, BBBB) to each clone was performed using the FitTetra software (Voorrips et al., 2013) with the Theta value (Carley et al., 2017). To process the genotyping data, default parameters were used (maxn.bin = 200, nbin = 200, sd.threshold = 0.1, p.threshold = 0.99, call.threshold = 0.6, peak.threshold = 0.85).



Genotypes were coded as numerical values from 0 to 4 according to the dosage of the alternate allele.

The position of the SNPs on the reference genome DM 1-3 516 R44 v6.1 was attributed according to information available on the Spud DB Potato Genomics Resources website (<http://spuddb.uga.edu>; DM 1-3 516 R44 - SolCAP 69K and PotVar GBS SNPs - v6.1). Each SNP could be positioned on one of the 12 chromosomes of the reference genome if the sequence had less than 2 mismatches or did not match to multiple positions. The SNP that did not fulfill these conditions were assigned to a fictive chromosome (chromosome 0).

2.4 Comparison between GBS and SolCAP array SNP calls

We conducted a thorough evaluation of the consistency and complementarity of SNPs identified using GBS and SolCAP 8K DNA array technologies. The complementarity was checked by evaluating the number of specific variants identified from each technology. To determine the level of concordance between these technologies, we compared, for each common variant, the assignment of one of the five possible genotypes (AAAA, AAAB,

AABB, ABBB, BBBB) to each clone, and calculated the percentage of discrepancies.

2.5 PCR-based markers

We genotyped the panel using eight dominant PCR-based markers which are commonly used in INRAE marker-assisted selection programs to screen for the presence of gene or QTL for resistance to cyst nematodes *G. pallida* or *G. rostochiensis*. These markers are 57R (Finkers-Tomczak et al., 2011), TG689 (Milczarek et al., 2011; Schultz et al., 2012), ASC151, MS137 (unpublished data), Gpa2-2 (Asano et al., 2012), Gro1-4 (Gebhardt et al., 2006), HC (Sattarzadeh et al., 2006) and Spud1636 (Bryan et al., 2002). The genotyping was performed using protocols adapted from original publications with DNA concentration of about 50 ng/μl. All PCR products were separated by electrophoresis on agarose gel and revealed with ethidium bromide.

We aligned the PCR probe sequences to the reference genome *Solanum tuberosum* group Phureja DM 1-3 516 R44 v6.1 using bowtie2 version 2.4.5, with the very sensitive mode and retained the best alignment position for each marker.

3 Results

3.1 Genotyping-by-sequencing

The GBS raw data of the potato MRQ panel represented 1.98 terabases. After demultiplexing and trimming, 98% of the reads remained (1.95 terabases). The alignment rate of the reads to the reference genome *Solanum tuberosum* Group Phureja DM 1-3 516 R44 v6.1 was 77%. Three clones showing more than 75% missing data were removed. Variant calling resulted in the identification of 539,947 variants. After applying filters, 159,002 variants remained. Finally, last filters on alternate allele count and missing data produced a dataset consisting in 22,087 high quality SNPs.

3.2 SolCAP array genotyping

Out of 2,366,355 total genotyping points, we were able to assign 1,301,718 points. A total of 5053 markers met the criteria set by fitTetra parameters. However, for these markers, some genotypes could not be assigned. Therefore the dataset contains 4.7% of missing data.

From these 5053 SNPs, 764 could not be positioned on any of the 12 potato chromosomes of the reference genome DM 1-3 516 R44 v6.1, leading to their assignment to chromosome 0.

3.3 Comparison of GBS and SolCAP array SNP calls

We identified 57 SNPs in common between both methods. Therefore, the majority of these SNPs were unique to each method

However, the markers were not evenly distributed along the whole genome (Figure 2B). None or few markers were identified in the centromeric regions, which can be explained by the use of the restriction enzyme *PstI* which is sensitive to methylation. The genomic distribution of the different types of markers was also examined and no visible patterns were detected.

To verify the consistency of the genomic dataset, we conducted a linkage disequilibrium (LD) analysis using the methods described by Vos et al. (2017). The LD values ranged from 0.1 to 1.4 Mb for LD_{1/2 90} and from 2.0 to 11.7 Mb for LD_{1/10 90}, depending on the chromosome, which was consistent with those reported by Sharma et al. (2018). The average interval between two SNPs being lower than the estimated LD values, our dataset guarantees extensive genome coverage for the discovery of marker-trait associations across the genome and the building of haplotype blocks at targeted loci.

The extensive SNP distribution of this dataset will enable breeders to perform GWAS for a diversity of traits as the ones mentioned in Marhadour & Prodhomme (2023), especially those relating to disease and pests resistance and quality. It will also support marker-assisted selection efforts (Islam et al., 2024). Breeders could also use this dataset to perform genomic predictions of breeding values for traits of interest (Gemenet et al., 2020) and identify good genitors among the genotypes of the potato MRQ panel for crossings in genomic selection schemes (Pandey et al., 2023).

The scientific community will also benefit from this dataset as a great tool for identifying ancestral haploblocks (Wang et al., 2022), which are notably difficult to discern in autotetraploid and highly heterozygous species like potato. The high number and diversity of genetic markers in this panel provide a robust groundwork for the creation and refinement of computational strategies and techniques for precise phasing in tetraploid potatoes.

5 Conclusion

The data reported in this study represents a comprehensive genomic dataset for 282 tetraploid potato clones, predominantly comprising INRAE advanced breeding lines for resistance to multiple pests and diseases, by integrating genotyping data derived from GBS, SolCAP 8k DNA array, and PCR-based markers. The final dataset includes 27,148 high-quality SNPs, providing a valuable resource for potato researchers and breeders especially working on resistance to biotic stresses.

Data availability statement

The data presented in the study are deposited in the Omics Dataverse repository accession number doi: [10.57745/DPHPOR](https://doi.org/10.57745/DPHPOR).

Author contributions

JL: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft. SS: Methodology, Supervision, Writing – review & editing. KM: Investigation, Writing – review & editing. RP: Resources, Writing – review & editing. AB: Investigation, Writing – review & editing. ML: Investigation, Writing – review & editing. DP: Investigation, Writing – review & editing. MD: Resources, Writing – review & editing. JC: Funding acquisition, Writing – review & editing. GB: Supervision, Writing – review & editing. MP: Supervision, Writing – review & editing, Project administration, Conceptualization, Funding acquisition. MK: Supervision, Writing – review & editing, Project administration, Conceptualization, Funding acquisition, Resources. FE: Funding acquisition, Project administration, Supervision, Writing – review & editing, Conceptualization, Resources.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was funded by ACVNPT (Association des Créateurs de Variétés Nouvelles de Pomme de terre), the European project G2PSol (Grant Agreement number: 677379 -G2P-Sol -H2020-SFS-2014-2015/H2020-SFS-2015-2), INRAE Metaprogramme SELGEN and CASDAR project GECONEM (2020 -C2018-07).

Acknowledgments

We thank Genouest BioInformatics platform for hosting sequencing data and supporting bioinformatic work. We acknowledge the BrACySol BRC and UE RGCO (INRAE Ploudaniel, France) for providing the plants that were used in this study. We are grateful to Sylvie Marhadour and Charlotte Prodhomme for helpful discussions about this work.

Conflict of interest

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

Publisher's note

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

References

- Asano, K., Kobayashi, A., Tsuda, S., Nishinaka, M., and Tamiya, S. (2012). DNA marker-assisted evaluation of potato genotypes for potential resistance to potato cyst nematode pathotypes not yet invading into Japan. *Breed Sci.* 62, 142–150. doi: 10.1270/jsbbs.62.142
- Bastien, M., Boudhrioua, C., Fortin, G., and Belzile, F. (2018). Exploring the potential and limitations of genotyping-by-sequencing for SNP discovery and genotyping in tetraploid potato. *Genome* 61, 449–456. doi: 10.1139/gen-2017-0236
- Black, W., Mastenbroek, C., Mills, W. R., and Peterson, L. C. (1953). A proposal for an international nomenclature of races of *Phytophthora infestans* and of genes controlling immunity in *Solanum demissum* derivatives. *Euphytica* 2, 173–178.
- Bryan, G., McLean, K., Bradshaw, J., De Jong, W., Phillips, M., Castelli, L., et al. (2002). Mapping QTLs for resistance to the cyst nematode *Globodera pallida* derived from the wild potato species *Solanum vernei*. *Theor. Appl. Genet.* 105, 68–77. doi: 10.1007/s00122-002-0873-9
- Carley, C. A. S., Coombs, J. J., Douches, D. S., Bethke, P. C., Palta, J. P., Novy, R. G., et al. (2017). Automated tetraploid genotype calling by hierarchical clustering. *Theor. Appl. Genet.* 130, 717–726. doi: 10.1007/s00122-016-2845-5
- J. J. Doyle and J. J. Doyle. Eds. (1990). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* pp. 13–15.
- Felcher, K. J., Coombs, J. J., Massa, A. N., Hansey, C. N., Hamilton, J. P., Veilleux, R. E., et al. (2012). Integration of two diploid potato linkage maps with the potato genome sequence. *PLoS One* 7, e36347. doi: 10.1371/journal.pone.0036347
- Finkers-Tomczak, A., Bakker, E., de Boer, J., van der Vossen, E., Achenbach, U., Golas, T., et al. (2011). Comparative sequence analysis of the potato cyst nematode resistance locus H1 reveals a major lack of co-linearity between three haplotypes in potato (*Solanum tuberosum* spp.). *Theor. Appl. Genet.* 122, 595–608. doi: 10.1007/s00122-010-1472-9
- Gebhardt, C., Bellin, D., Henselewski, H., Lehmann, W., Schwarzfischer, J., and Valkonen, J. P. T. (2006). Marker-assisted combination of major genes for pathogen resistance in potato. *Theor. Appl. Genet.* 112, 1458–1464. doi: 10.1007/s00122-006-0248-8
- Gemenet, D. C., Lindqvist-Kreuzer, H., De Boeck, B., Pereira, G., de, S., Mollinari, M., et al. (2020). Sequencing depth and genotype quality: accuracy and breeding operation considerations for genomic selection applications in autopolyploid crops. *Theor. Appl. Genet.* 133, 3345–3363. doi: 10.1007/s00122-020-03673-2
- Hamilton, J. P., Hansey, C. N., Whitty, B. R., Stoffel, K., Massa, A. N., Van Deynze, A., et al. (2011). Single nucleotide polymorphism discovery in elite north american potato germplasm. *BMC Genomics* 12, 302. doi: 10.1186/1471-2164-12-302
- Islam, S., Li, J., Rahman, M. A., Xie, F., Song, B., and Nie, B. (2024). Resistance to biotic and abiotic stress in potato: the origin of the genes and corresponding molecular markers. *Phytopathol. Res.* 6, 4. doi: 10.1186/s42483-023-00222-9
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–U54. doi: 10.1038/nmeth.1923
- Marhadour, S., and Prodhomme, C. (2023). “Chapter 15 - Recent trends in genetics studies and molecular breeding of potato,” in *Potato Production Worldwide*. Eds. M. E. Çalişkan, A. Bakhsh and K. Jabran (Academic Press), 273–301. Available at: <https://shop.elsevier.com/books/potato-production-worldwide/caliskan/978-0-12-822925-5>.
- Melo, A. T. O., Bartaula, R., and Hale, I. (2016). GBS-SNP-CROP: a reference-optional pipeline for SNP discovery and plant germplasm characterization using variable length, paired-end genotyping-by-sequencing data. *BMC Bioinf.* 17, 29. doi: 10.1186/s12859-016-0879-y
- Milczarek, D., Flis, B., and Przetakiewicz, A. (2011). Suitability of molecular markers for selection of potatoes resistant to *Globodera* spp. *Am. J. Pot. Res.* 88, 245–255. doi: 10.1007/s12230-011-9189-0
- Negro, S. S., Millet, E. J., Madur, D., Bauland, C., Combes, V., Welcker, C., et al. (2019). Genotyping-by-sequencing and SNP-arrays are complementary for detecting quantitative trait loci by tagging different haplotypes in association studies. *BMC Plant Biol.* 19, 318. doi: 10.1186/s12870-019-1926-4
- Pandey, J., Scheuring, D. C., Koym, J. W., Endelman, J. B., and Vales, M. I. (2023). Genomic selection and genome-wide association studies in tetraploid chipping potatoes. *Plant Genome* 16. doi: 10.1002/tpg2.20297
- Pham, G. M., Hamilton, J. P., Wood, J. C., Burke, J. T., Zhao, H., Vaillancourt, B., et al. (2020). Construction of a chromosome-scale long-read reference genome assembly for potato. *GigaScience* 9, gaa100. doi: 10.1093/gigascience/giaa100
- Pilet-Nayel, M.-L., Moury, B., Caffier, V., Montarry, J., Kerlan, M.-C., Fournet, S., et al. (2017). Quantitative resistance to plant pathogens in pyramiding strategies for durable crop protection. *Front. Plant Sci.* 8, 1838. doi: 10.3389/fpls.2017.01838
- Sattarzadeh, A., Achenbach, U., Lübeck, J., Strahwald, J., Tacke, E., Hofferbert, H.-R., et al. (2006). Single nucleotide polymorphism (SNP) genotyping as basis for developing a PCR-based marker highly diagnostic for potato varieties with high resistance to *Globodera pallida* pathotype Pa2/3. *Mol. Breed.* 18, 301–312. doi: 10.1007/s11032-006-9026-1
- Schultz, L., Cogan, N. O. I., McLean, K., Dale, M. F. B., Bryan, G. J., Forster, H. W., et al. (2012). Evaluation and implementation of a potential diagnostic molecular marker for H1-conferred potato cyst nematode resistance in potato (*Solanum tuberosum* L.). *Plant Breed.* 131, 315–321. doi: 10.1111/j.1439-0523.2012.01949.x
- Sharma, S. K., MacKenzie, K., McLean, K., Dale, F., Daniels, S., and Bryan, G. J. (2018). Linkage disequilibrium and evaluation of genome-wide association mapping models in tetraploid potato. *G3 Genes|Genomes|Genetics* 8, 3185–3202. doi: 10.1534/g3.118.200377
- Sharma, S. K., McLean, K., Hedley, P. E., Dale, F., Daniels, S., and Bryan, G. J. (2024). Genotyping-by-sequencing targets genic regions and improves resolution of genome-wide association studies in autotetraploid potato. *Theor. Appl. Genet.* (accepted).
- Van der Auwera, G., and O'Connor, B. (2020). *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*. 1st edition (Sebastopol, CA: O'Reilly Media).
- Voorrips, R. E., Gort, G., and Vosman, B. (2013). Genotype calling in tetraploid species from bi-allelic marker data using mixture models. *BMC Bioinf.* 12, 172. doi: 10.1186/1471-2105-12-172
- Vos, P. G., Paulo, M. J., Voorrips, R. E., Visser, R. G. F., van Eck, H. J., and van Eeuwijk, F. A. (2017). Evaluation of LD decay and various LD-decay estimators in simulated and SNP-array data of tetraploid potato. *Theor. Appl. Genet.* 130, 123–135. doi: 10.1007/s00122-016-2798-8
- Wang, Z., Wang, W., Xie, X., Wang, Y., Yang, Z., Peng, H., et al. (2022). Dispersed emergence and protracted domestication of polyploid wheat uncovered by mosaic ancestral haplotype inference. *Nat. Commun.* 13, 3891. doi: 10.1038/s41467-022-31581-0