Deploying High-throughput Phenotyping for Genomic assisted Breeding: Pathway to Accelerated Genetic Gain and Efficiency in Potato

A DISSERTATION SUBMITTED TO THE FACULTY OF THE GRADUATE
SCHOOL OF THE UNIVERSITY OF MINNESOTA
BY

Muyideen Yusuf

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Dr. Laura M. Shannon

Advisor

December 2024

Acknowledgments

I extend my deepest gratitude to my advisor, Dr. Laura Shannon, for her unwavering support, insightful guidance, and the invaluable opportunity to pursue these research projects in her lab. Her mentorship has been a cornerstone of my academic journey.

I am profoundly grateful to my supervisory committee, Drs. Rex Bernado, Zhanyou Xu, and Yaniv Brandvain, for their thoughtful input, critical suggestions, and encouragement throughout the course of my research. Their expertise has been instrumental in shaping my work.

Special thanks go to the past and present members of the Shannon lab for their friendship, support, and assistance. I also wish to acknowledge the contributions and influence of Drs. Jeffery Endelman and Osval, whose insights have had a meaningful impact on my research. My sincere appreciation extends to Drs. Andreas Gisel, Ismail Rabbi, Guillaume Bauchet, Lukas Mueller, Mohammed Somo, and Sara Tirado, who have provided exceptional mentorship both before and during my PhD journey.

I am deeply thankful for the financial support I received from the Department of Horticultural Science, the Applied Plant Science program, and the UMN Plant Breeding Center, without which this research would not have been possible. I am also indebted to Dr. Cindy Tong for her academic support and encouragement. My heartfelt thanks to Serena Jones and Kelsey Bogdanovich for their invaluable academic assistance and advice.

To my friends and family: your unwavering support has been my anchor. I am grateful to numerous friends, including Adenike Ige, Emmanuel Adeyemo, Cedric Ndinga, Lucas Roberts, Mjay Espina, and all my APS Club friends, for being a part of this journey. My appreciation also extends to my friends in the US and Nigeria, particularly Bakare Moshood, ThankGod Ogwuche, Ruth Uwugiaren, Gideon Oyebode, and Shittu Haruna, for their encouragement and belief in me. Lastly, I am eternally grateful to my parents for their unceasing prayers, encouragement, and support throughout my PhD program. To Farida and my siblings: thank you for your prayers, humor, and timely distractions—they meant more than words can express.

Dedication

To the glory of God, my family, friends and well-wishers.

Abstract

Potato (*Solanum tuberosum*) is one of the most widely cultivated non-grain food crops in the world, playing a critical role in global food security. It is utilized as a vegetable, livestock feed, and a source of raw materials for various industrial products. However, yield improvement through traditional breeding approaches has stagnated. This is largely attributed to the crop's genome complexity, characterized by its autotetraploid, clonal, and outcrossing nature. Despite recent advancements in ploidy theory and the growing availability of genomic resources, challenges in accurate phenotyping and the comprehensive evaluation of numerous traits persist. This research investigates innovative phenomic technologies to enhance genomic-assisted breeding in two potato market classes: chips and fresh market. The studies employ high-throughput phenotyping approaches, integrating digital imaging, drone-derived multispectral data, and machine learning techniques to address key challenges.

The first study explored the use of quantitative phenotypic scores based on digital images for improving genomic selection of quality related traits. Analysis revealed significant contributions of additive and non-additive genetic effects, offering critical insights into the genetic mechanisms underlying these traits and their implications for potato breeding. The second study explored leveraging drone derived multispectral image data for phenomic selection and its potential to augment genomic selection. Highlight of this work shows phenomic selection surpassed genomic selection for yield prediction, with a combined phenomic-genomic approach improving prediction accuracy by over 30%. The third study assessed drone-derived multispectral variables to predict total plant nitrogen (N). Machine learning models achieved moderate-to-high accuracy in N prediction, with varying important features selected for improved prediction. The findings underscore the transformative potential of integrating phenomics with genomics to overcome breeding barriers in potato, improving the precision and efficiency of genetic enhancement strategies.

Table of Contents

List of Tables

List of Figures

Chapter 1

Literature review

1.1 Introduction

The cultivated potato is the most popular non grain food crop contributing significantly to food security in the world (FAO, 2023). It has a vast wealth of genetic resources due to ease of crossing ability with several of its related species (Hardigan et al., 2017; Jansky and Spooner, 2018). Potatoes are believed to have been domesticated thousands of years ago from wild relatives informally grouped as the Solanum brevicaule complex, native to the Andes in southern Peru (Spooner et al., 2007; Hardigan et al., 2017). Over time, potato cultivation spread to highland savannahs in Chile, where plants adapted to longer summer days. During this process, potatoes transitioned into an autotetraploid form (2n = 4x = 48) due to the frequent occurrence of 2n gametes in early diploid landrace species, such as *S. tuberosum* groups *Stenotomum* and *Phureja*. This transition led to the emergence of Andean 4x cultivated species (Hardigan et al., 2017). The potato's ancient origin and its adaptation to varying day-length requirements have supported its global cultivation. Its genetic diversity enables the selection of cultivars tailored to local environmental conditions, making it a versatile and resilient crop (Hardigan et al., 2017; Slater et al., 2017).

Despite its global importance, the potato faces significant biological and economic challenges that stagnates its improvement and agronomic performance. Its clonal propagation increases susceptibility to diseases, leading to poor tuber quality and reduced yields. The complexity of its autotetraploid genome, combined with high heterozygosity and outbreeding characteristics, further complicates efforts to improve the crop for essential quality traits.

Additionally, breeding efforts require the development of large populations to identify a small number of elite cultivars with specific traits tailored to distinct market demands and environmental conditions. This necessitates the evaluation of an extensive array of traits, making the process labor-intensive and resource-demanding (Jansky, 2009).

Traditional potato breeding is a demanding, multigenerational process that begins with the crossing of parents selected for complementary traits of interest, producing F1 progeny. This is followed by several cycles of clonal phenotypic recurrent selection to achieve desirable trait combinations, a process that typically spans nearly a decade (Jansky, 2009). The identification of superior cultivars requires simultaneous evaluation of multiple traits, necessitating in-depth genetic knowledge and understanding of environmental influences. While traits like disease resistance are often straightforward to select due to their qualitative nature, complex traits such as yield demand advanced genetic analysis and precision tools, given their quantitative inheritance (Slater et al., 2014; Bradshaw, 2017).

The challenges associated with traditional potato breeding, characterized by its slow pace and complexity, underscore the urgent need for innovative approaches to accelerate the process and enhance precision in selecting both simple and complex traits. This is particularly important to bring potato breeding advancements in line with the progress achieved in grain crops like maize (Ortiz, 2020). Several ongoing strategies present opportunities to overcome these obstacles and improve breeding efficiency. The application of advanced genomic tools and resources, such as marker-assisted selection (MAS) and genomic selection (GS), allows breeders to more effectively target desirable traits at the molecular level (Bradshaw, 2017; Slater et al., 2017). Additionally, high-throughput phenotyping and imaging technologies offer the capability to rapidly and accurately capture phenotypic data. These technologies enable the measurement of

critical traits such as growth rates, disease resistance, heat tolerance, and nutrient use efficiency (Slater et al., 2017).

These innovations hold promise for revolutionizing potato breeding through the integration these modern tools enabling the development of superior cultivars tailored for diverse environments and end-use applications. By leveraging these tools, potato breeding can overcome historical constraints and meet the growing demands of a changing world.

1.2 Development of genomic tools and resources

The development of new tools and genomic resources not initially available for analysis of autopolyploid genome have now enabled the use of DNA technology for genomic assisted breeding of complex traits in potato. The genotyping of DNA sequence variants such as SNPs is challenging for autopolyploids like potato compared to diploids. This is because a gene is represented by up to four alleles per locus per genotype, requiring a specialized approach for sequencing and genotyping that can distinguish among alleles copy numbers (Uitdewilligen et al., 2013; Bourke et al., 2018).

However, high throughput sequencing technologies have evolved from use of double monoploids for development of genome assemblies (Xu et al., 2011; Felcher et al., 2012; Freire et al., 2021) to use of phased genomes and pangenomes (Pham et al., 2020; Hoopes et al., 2022). These technologies have improved development of low density marker assays (Vos et al., 2015; Asano and Endelman, 2023; Endelman et al., 2024) used as high throughput markers for downstream analysis such as genetic diversity studies, QTL mapping, genome wide association studies (GWAS), genomic selection (GS). Also, specialized tools modified to handle complex trait analysis for autopolyploid genome have also been developed providing seamless solution to

ploidy complexity such as double reduction (Rosyara et al., 2016; Bourke et al., 2018; Endelman, 2023). This significant stride provides breeders with tools for early generation selection to speed up the breeding process through the use of marker assisted selection or genomic selection.

1.3 Marker assisted selection

QTL discovery from association studies or QTL mapping are utilized by breeders for markers assisted selection (MAS) which can speed the breeding process (Slater et al., 2016; Ortiz, 2020). Applying genetic markers as early as the first or second field year offers significantly greater efficiency compared to traditional visual screening methods (Bradshaw, 2017). While marker assisted selection (MAS) can be highly efficient, its effectiveness is generally limited to traits with simple genetic architecture and few large effect genes such as self-compatibility (Clot et al., 2020; Kaiser et al., 2021), maturity (Kloosterman et al., 2013; Gutaker et al., 2019), oblong vs. circular shape (Eck et al., 2017; Endelman et al., 2024), and disease resistance (Herrera et al., 2018; Prodhomme et al., 2019; Meade et al., 2020; Asano and Endelman, 2023).

Although many markers linked to genes for critical traits have been identified, their application in commercial potato breeding programs remains limited. For MAS to be viable in practice, its cost-effectiveness must surpass that of conventional screening methods. This is particularly evident in the use of disease resistance markers, where significant time savings and efficiency gains have been demonstrated (Slater et al., 2016; Asano and Endelman, 2023). Expanding the practical use of MAS for complex traits will require innovations to reduce costs and enhance applicability across diverse breeding objectives.

1.4 Genomic selection

Genomic selection (GS) is a powerful tool in modern breeding that estimates genome-wide

marker effects for a given trait by using a training population that has been both genotyped and

phenotyped. This information is then used to predict the genetic values of untested selection

candidates (Meuwissen et al., 2001). Unlike marker-assisted selection (MAS), which relies on a

limited set of markers linked to QTLs, GS leverages data from all available markers to capture

the full genetic variance of a trait. This comprehensive approach reduces the risk of

overestimating marker effects, a common issue with GWAS or QTL mapping (Beavis and van

Heerwaarden, 2017; Crossa et al., 2017).

Genomic selection (GS) has proven to be a superior tool compared to marker-assisted

selection (MAS), particularly for traits with complex genetic architectures. This is due to GS's

ability to harness the cumulative effects of numerous small genetic contributions across the

genome (Crossa et al., 2017; Ortiz, 2020). However, the effective application of GS in potato

breeding requires addressing several critical factors, including marker density, linkage

disequilibrium (LD), population size, and heritability (Slater et al., 2016).

1.4.1 Marker density

High-density genetic markers are essential for achieving a comprehensive spread across

the potato genome, ensuring that genetic variations are captured and that at least one marker is in

LD with every QTL. Empirical and simulation studies have validated the number of markers

needed for potato breeding programs to achieve these objectives effectively (Slater et al., 2016).

Advances in marker development have resulted in tools that fulfill these requirements, supported

by diverse versions and genotyping platforms tailored to potato breeding. With the continued

decrease in marker costs, these tools have become increasingly accessible, enhancing the efficiency and precision of genomic selection (Vos et al., 2015; Endelman et al., 2024).

1.4.2 Linkage disequilibrium (LD)

LD refers to the non-random association of alleles at different loci and is estimated by the correlation between markers, with values ranging from 0 to 1. In potato, LD is a critical determinant for the number of markers required to capture the majority of QTLs (Vos et al., 2017). Research has shown that in potato breeding populations, high LD exists at distances less than 1 cM but decays rapidly to values below 0.2 at greater distances (D'hoop et al., 2010; Slater et al., 2016).

1.4.3 Population size and genetic diversity

A large population size is essential for maintaining genetic diversity, which is crucial for capturing a broad range of alleles and enhancing heritability estimates. However, effective population size (Ne) is often more critical than the actual population size. Ne contributes to stable heritability estimates across generations and can inform the number of markers required for GS (Slater et al., 2016).

Despite decreasing sequencing costs, genotyping the tens of thousands of unique genetic individuals involved in early potato breeding generations remains expensive. However, for complex traits, GS offers the potential to increase genetic gains and shorten breeding cycles, making it an invaluable tool in modern potato breeding. Studies have demonstrated its effectiveness in improving genetic gain for complex traits, even in the context of the challenges associated with marker density, LD, and population size (Habyarimana et al., 2017; Stich and Van Inghelandt, 2018; Endelman et al., 2018; Martins et al., 2023).

1.5 High throughput phenotyping

Advances in high-throughput phenotyping (HTP) are expected to enhance the efficiency and accuracy of MAS and GS. HTP can enhance morphological trait scoring in plant breeding, accelerate the phenotyping processes, and increase precision (Shakoor et al., 2017; Hickey et al., 2019). Image-based phenotyping of tubers has been implemented to develop quantitative measures for quality traits, such as shape and color, in potatoes (Caraza-Harter and Endelman, 2020; Neilson et al., 2021; Miller et al., 2022; Feldman et al., 2024). The phenotypic data derived from these images allows for genomic prediction of quality traits (Yusuf et al., 2024).

Advances in sensor technology, machine learning, and data analytics are further enhancing the efficiency and accuracy of HTP in plant breeding (Hickey et al., 2019). Both ground-based and aerial HTP platforms have been deployed for phenotyping plants, either by directly capturing specific traits or indirectly examining phenotypic differences among plants using spectral reflectance data (Song et al., 2021). Aerial phenotyping offers the significant advantage of covering a wider expanse of field trials within a short timeframe. One such aerial platform is the unmanned aerial vehicle (UAV), which is equipped with multi-sensor cameras capable of capturing wavelengths near and beyond the visible spectrum. These wavelengths reflect different physiochemical properties related to plant structure, health, and nutrient status based on canopy leaf absorption and reflectance (Alkhaled et al., 2023). Studies in several crops have reported significant improvements in prediction accuracy using HTP temporal data, both for direct predictions and within genomic selection models (Silva et al., 2022; Montesinos-López et al., 2023; Xu et al., 2023; Abdelhakim et al., 2024).

References

Abdelhakim, L.O.A., B. Pleskačová, N.Y. Rodriguez-Granados, R. Sasidharan, L.S. Perez-Borroto, et al. 2024. High Throughput Image-Based Phenotyping for Determining Morphological and Physiological Responses to Single and Combined Stresses in Potato. J Vis Exp (208). doi: 10.3791/66255.

Alkhaled, A., P.A. Townsend, and Y. Wang. 2023. Remote Sensing for Monitoring Potato Nitrogen Status. Am. J. Potato Res. 100(1): 1–14. doi: 10.1007/s12230-022-09898-9.

Asano, K., and J.B. Endelman. 2023. Development of KASP markers for the potato virus Y resistance gene Rychc using whole-genome resequencing data. : 2023.12.20.572658. doi: 10.1101/2023.12.20.572658.

Beavis, W.D., and J.P.M. van Heerwaarden. 2017. Mapping QTLs: A comparison of methods. Molecular Genetics of Extracellular Plant Pathogens. Springer. p. xx–xx

Bourke, P.M., R.E. Voorrips, R.G.F. Visser, and C. Maliepaard. 2018. Tools for Genetic Studies in Experimental Populations of Polyploids. Front Plant Sci 9. doi: 10.3389/fpls.2018.00513.

Bradshaw, J.E. 2017. Review and Analysis of Limitations in Ways to Improve Conventional Potato Breeding. Potato Res. 60(2): 171–193. doi: 10.1007/s11540-017-9346-z.

Caraza-Harter, M.V., and J.B. Endelman. 2020. Image-based phenotyping and genetic analysis of potato skin set and color. Crop Science 60(1): 202–210. doi: 10.1002/csc2.20093.

Clot, C.R., C. Polzer, C. Prodhomme, C. Schuit, C.J.M. Engelen, et al. 2020. The origin and widespread occurrence of Sli-based self-compatibility in potato. Theor Appl Genet. doi: 10.1007/s00122-020-03627-8.

Crossa, J., P. Pérez-Rodríguez, J. Cuevas, O. Montesinos-López, D. Jarquín, et al. 2017. Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. Trends in Plant Science 22(11): 961–975. doi: 10.1016/j.tplants.2017.08.011.

D'hoop, B.B., M.J. Paulo, K. Kowitwanich, M. Sengers, R.G.F. Visser, et al. 2010. Population structure and linkage disequilibrium unravelled in tetraploid potato. Theoretical and Applied Genetics 121: 1151–1170. doi: 10.1007/s00122-010-1379-5.

Eck, H.J. van, P.G. Vos, J.P.T. Valkonen, J.G.A.M.L. Uitdewilligen, H. Lensing, et al. 2017. Graphical genotyping as a method to map Ny (o,n)sto and Gpa5 using a reference panel of tetraploid potato cultivars. Theor Appl Genet 130(3): 515–528. doi: 10.1007/s00122-016-2831-y.

Endelman, J.B. 2023. Fully efficient, two-stage analysis of multi-environment trials with directional dominance and multi-trait genomic selection. Theor Appl Genet 136(4): 65. doi: 10.1007/s00122-023-04298-x.

Endelman, J.B., C.A.S. Carley, P.C. Bethke, J.J. Coombs, M.E. Clough, et al. 2018. Genetic Variance Partitioning and Genome-Wide Prediction with Allele Dosage Information in Autotetraploid Potato. Genetics 209(1): 77–87. doi: 10.1534/genetics.118.300685.

Endelman, J.B., M. Kante, H. Lindqvist-Kreuze, A. Kilian, L.M. Shannon, et al. 2024. Targeted genotyping-by-sequencing of potato and software for imputation. : 2024.02.12.579978. doi: 10.1101/2024.02.12.579978.

Felcher, K.J., J.J. Coombs, A.N. Massa, C.N. Hansey, J.P. Hamilton, et al. 2012. Integration of Two Diploid Potato Linkage Maps with the Potato Genome Sequence. PLOS ONE 7(4): e36347. doi: 10.1371/journal.pone.0036347.

Feldman, M.J., J. Park, N. Miller, C. Wakholi, K. Greene, et al. 2024. A scalable, low-cost phenotyping strategy to assess tuber size, shape, and the colorimetric features of tuber skin and flesh in potato breeding populations. The Plant Phenome Journal 7(1): e20099. doi: 10.1002/ppj2.20099.

Freire, R., M. Weisweiler, R. Guerreiro, N. Baig, B. Hüttel, et al. 2021. Chromosome-scale reference genome assembly of a diploid potato clone derived from an elite variety. G3 (Bethesda): jkab330. doi: 10.1093/g3journal/jkab330.

Gutaker, R.M., C.L. Weiß, D. Ellis, N.L. Anglin, S. Knapp, et al. 2019. The origins and adaptation of European potatoes reconstructed from historical genomes. Nat Ecol Evol 3(7): 1093–1101. doi: 10.1038/s41559-019-0921-3.

Habyarimana, E., B. Parisi, and G. Mandolino. 2017. Genomic prediction for yields, processing and nutritional quality traits in cultivated potato (Solanum tuberosum L.). Plant Breeding 136(2): 245–252. doi: 10.1111/pbr.12461.

Hardigan, M.A., F.P.E. Laimbeer, L. Newton, E. Crisovan, J.P. Hamilton, 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 U S A 114(46): E9999–E10008. doi: 10.1073/pnas.1714380114.

Herrera, M. del R., L.J. Vidalon, J.D. Montenegro, C. Riccio, F. Guzman, et al. 2018. Molecular and genetic characterization of the Ryadg locus on chromosome XI from Andigena potatoes conferring extreme resistance to potato virus Y. Theor Appl Genet 131(9): 1925–1938. doi: 10.1007/s00122-018-3123-5.

Hoopes, G., X. Meng, J.P. Hamilton, S.R. Achakkagari, F. de Alves Freitas Guesdes, et al. 2022. Phased, chromosome-scale genome assemblies of tetraploid potato reveal a complex genome, transcriptome, and predicted proteome landscape underpinning genetic diversity. Molecular Plant 15(3): 520–536. doi: 10.1016/j.molp.2022.01.003.

Jansky, S. 2009. Chapter 2 - Breeding, Genetics, and Cultivar Development. In: Singh, J. and Kaur, L., editors, Advances in Potato Chemistry and Technology. Academic Press, San Diego. p. 27–62

Jansky, S.H., and D.M. Spooner. 2018. The Evolution of Potato Breeding. Plant Breeding Reviews. John Wiley & Sons, Ltd. p. 169–214

Kaiser, N.R., S. Jansky, J.J. Coombs, P. Collins, M. Alsahlany, et al. 2021. Assessing the Contribution of Sli to Self-Compatibility in North American Diploid Potato Germplasm Using KASP^{TM} Markers. Am. J. Potato Res. doi: 10.1007/s12230-021-09821-8.

Kloosterman, B., J.A. Abelenda, M. del M.C. Gomez, M. Oortwijn, J.M. de Boer, et al. 2013. Naturally occurring allele diversity allows potato cultivation in northern latitudes. Nature 495(7440): 246–250. doi: 10.1038/nature11912.

Martins, V.S., M.H.M.L. Andrade, L.N. Padua, L.A. Miguel, C.C. Fernandes Filho, et al. 2023. Evaluating the impact of modeling the family effect for clonal selection in potato-breeding programs. Front Plant Sci 14: 1253706. doi: 10.3389/fpls.2023.1253706.

Meade, F., S. Byrne, D. Griffin, C. Kennedy, F. Mesiti, et al. 2020. Rapid Development of KASP Markers for Disease Resistance Genes Using Pooled Whole-Genome Resequencing. Potato Res. 63(1): 57–73. doi: 10.1007/s11540-019-09428-x.

Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics 157(4): 1819–1829. doi: 10.1093/genetics/157.4.1819.

Miller, M.D., C.A. Schmitz Carley, R.A. Figueroa, M.J. Feldman, D. Haagenson, et al. 2022. TubAR: an R Package for Quantifying Tuber Shape and Skin Traits from Images. Am. J. Potato Res. doi: 10.1007/s12230-022-09894-z.

Montesinos-López, O.A., S. Ramos-Pulido, C.M. Hernández-Suárez, B.A. Mosqueda González, F.A. Valladares-Anguiano, et al. 2023. A novel method for genomic-enabled prediction of cultivars in new environments. Frontiers in Plant Science 14. https://www.frontiersin.org/articles/10.3389/fpls.2023.1218151 (accessed 25 July 2023).

Neilson, J.A.D., A.M. Smith, L. Mesina, R. Vivian, S. Smienk, et al. 2021. Potato Tuber Shape Phenotyping Using RGB Imaging. Agronomy 11(9): 1781. doi: 10.3390/agronomy11091781.

Ortiz, R. 2020. Genomic-Led Potato Breeding for Increasing Genetic Gains: Achievements and Outlook. Crop Breeding, Genetics and Genomics 2(2). doi: https://doi.org/10.20900/cbgg20200010.

Pham, G.M., J.P. Hamilton, J.C. Wood, J.T. Burke, H. Zhao, et al. 2020. Construction of a chromosome-scale long-read reference genome assembly for potato. GigaScience 9(9): giaa100. doi: 10.1093/gigascience/giaa100.

Prodhomme, C., D. Esselink, T. Borm, R.G.F. Visser, H.J. van Eck, et al. 2019. Comparative Subsequence Sets Analysis (CoSSA) is a robust approach to identify haplotype specific

SNPs; mapping and pedigree analysis of a potato wart disease resistance gene Sen3. Plant Methods 15(1): 60. doi: 10.1186/s13007-019-0445-5.

Rosyara, U.R., W.S.D. Jong, D.S. Douches, and J.B. Endelman. 2016. Software for Genome-Wide Association Studies in Autopolyploids and Its Application to Potato. The Plant Genome 9(2): plantgenome2015.08.0073. doi: 10.3835/plantgenome2015.08.0073.

Silva, M.F. e, G.M. Maciel, R.B. Gallis, R.L. Barbosa, V.Q. Carneiro, et al. 2022. High-throughput phenotyping by RGB and multispectral imaging analysis of genotypes in sweet corn. Hortic. Bras. 40: 92–98. doi: 10.1590/s0102-0536-2022012.

Slater, A.T., N.O.I. Cogan, J.W. Forster, B.J. Hayes, and H.D. Daetwyler. 2016. Improving Genetic Gain with Genomic Selection in Autotetraploid Potato. The Plant Genome 9(3): plantgenome2016.02.0021. doi: https://doi.org/10.3835/plantgenome2016.02.0021.

Slater, A.T., N.O.I. Cogan, B.C. Rodoni, H.D. Daetwyler, B.J. Hayes, et al. 2017. Breeding Differently—the Digital Revolution: High-Throughput Phenotyping and Genotyping. Potato Res. 60(3): 337–352. doi: 10.1007/s11540-018-9388-x.

Slater, A.T., G.M. Wilson, N.O.I. Cogan, J.W. Forster, and B.J. Hayes. 2014. Improving the analysis of low heritability complex traits for enhanced genetic gain in potato. Theor Appl Genet 127(4): 809–820. doi: 10.1007/s00122-013-2258-7.

Song, P., J. Wang, X. Guo, W. Yang, and C. Zhao. 2021. High-throughput phenotyping: Breaking through the bottleneck in future crop breeding. The Crop Journal 9(3): 633–645. doi: 10.1016/j.cj.2021.03.015.

Spooner, D.M., J. Núñez, G. Trujillo, M. del R. Herrera, F. Guzmán, et al. 2007. Extensive simple sequence repeat genotyping of potato landraces supports a major reevaluation of their gene pool structure and classification. PNAS 104(49): 19398–19403. doi: 10.1073/pnas.0709796104.

Stich, B., and D. Van Inghelandt. 2018. Prospects and Potential Uses of Genomic Prediction of Key Performance Traits in Tetraploid Potato. Front. Plant Sci. 9. doi: 10.3389/fpls.2018.00159.

Uitdewilligen, J.G.A.M.L., A.-M.A. Wolters, B.B. D'hoop, T.J.A. Borm, R.G.F. Visser, et al. 2013. A Next-Generation Sequencing Method for Genotyping-by-Sequencing of Highly Heterozygous Autotetraploid Potato. PLOS ONE 8(5): e62355. doi: 10.1371/journal.pone.0062355.

Vos, P.G., M.J. Paulo, R.E. Voorrips, R.G.F. Visser, and H.J. van Eck. 2017. Evaluation of LD decay and various LD-decay estimators in simulated and SNP-array data of tetraploid potato. Theoretical and Applied Genetics 130: 123–135. doi: 10.1007/s00122-016-2798-8.

Vos, P.G., J.G.A.M.L. Uitdewilligen, R.E. Voorrips, R.G.F. Visser, and H.J. van Eck. 2015. Development and analysis of a 20K SNP array for potato (Solanum tuberosum): an insight into the breeding history. Theor Appl Genet 128(12): 2387–2401. doi: ros.

Xu, X., S. Pan, S. Cheng, B. Zhang, D. Mu, et al. 2011. Genome sequence and analysis of the tuber crop potato. Nature 475(7355): 189–195. doi: 10.1038/nature10158.

Xu, S., X. Xu, Q. Zhu, Y. Meng, G. Yang, et al. 2023. Monitoring leaf nitrogen content in rice based on information fusion of multi-sensor imagery from UAV. Precision Agric. doi: 10.1007/s11119-023-10042-8.

Yusuf, M., M.D. Miller, T.R. Stefaniak, D. Haagenson, J.B. Endelman, et al. 2024. Genomic prediction for potato (Solanum tuberosum) quality traits improved through image analysis. The Plant Genome n/a(n/a): e20507. doi: 10.1002/tpg2.20507.

Chapter 2

Genomic prediction for potato (*Solanum tuberosum*) quality traits improved through image analysis

Overview

Consumers and processors evaluate potatoes based on quality traits such as shape and skin color, making these traits important targets for breeders. Achieving and evaluating genetic gain is facilitated by precise and accurate trait measures. Historically, quality traits have been measured using visual rating scales which are subject to human error and necessarily lump individuals with distinct traits into categories. Image analysis offers a method of generating quantitative measures of quality traits. In this study, we use TubAR, an image-analysis R package, to generate quantitative measures of shape and skin color traits for use in genomic prediction. We developed and compared different genomic models based on additive and additive plus non-additive kernels for two aspects of skin color, redness, and lightness, and two aspects of shape, roundness, and length-to-width ratio, for fresh market red and yellow potatoes grown in Minnesota between 2020 and 2022. Similarly, we used the much larger chipping potato population grown during the same time to develop a multi-trait selection index including roundness, specific gravity, and yield. Traits ranged in heritability with shape traits falling between 0.23 and 0.85, and color traits falling between 0.34 and 0.91. Genetic effects were primarily additive with color traits showing the strongest effect (0.47), while shape traits varied based on market class. Modeling non-additive effects did not significantly improve prediction models for quality traits. The

combination of image analysis and genomic prediction presents a promising avenue for

improving potato quality traits.

2.1 Introduction

Potato (*Solanum tuberosum* L.) is the most widely grown vegetable crop in the world (FAO, 2023). However, the adoption of new potato varieties is slow compared to many other staple crops (Douches et al., 1996). This can be partially attributed to grower, industry, and consumer expectations for specific sets of quality and processing traits that define market. A new potato variety needs not just high yield, processing potential, and disease resistance, but to meet a series of tuber appearance standards including shape and color to gain acceptance (Carputo et al., 2004; Bradshaw, 2017).

Many tuber appearance traits including color (Caraza-Harter and Endelman, 2020) and shape (Meijer et al., 2018) are quantitative. Although major effect loci have been characterized for both tuber shape (Wu et al., 2018; van Ek et al., 2022; Jo et al., 2023) and skin color (Bonar et al., 2018; Jung et al., 2009; De Jong et al., 2003), these loci do not explain the full range of phenotypes. Even among round red potatoes, there is a strong genetic component for both skin color and shape (Jones et al., 2021; Stefaniak et al., 2021). These differences can determine grower and consumer preference; for example, somatic mutants of fresh market red potatoes resulting in darker red skin color are often marketed as new varieties e.g. Dark Red Norland or Dark Red Chieftain. While marker-assisted selection could be used for major effect genes (Endelman et al., 2024), selection for the crucial smaller effect genes will require an alternate approach. Genomic selection (GS) provides a more robust approach to clonal and parental selection (Slater et al., 2016).

Genomic selection for yield and yield components has been successfully implemented in tetraploid potatoes (Endelman et al., 2018; Gemenet et al., 2020; Habyarimana et al., 2017; Selga et al., 2021; Sood et al., 2020; Stitch & Van Inghelandt, 2018; Martins et al, 2023). Disease

resistance (Gemenet et al., 2020; Selga et al., 2021; Sood et al., 2020; Stitch & Van Inghelandt, 2018; Sverrisdottir et al., 2018; Enciso-Rodriguez et al., 2018), nutritional value (Pandey et al., 2023) and crucial processing traits like specific gravity and chip color (Endelman et al., 2018; Sverrisdottir et al., 2017; Habyarimana et al., 2017; Sood et al., 2020; Stitch & Van Inghelandt, 2018; Caruana et al., 2019; Sverrisdottir et al., 2018; Martins et al, 2023) have also been targets for prediction. Selection for quality traits has been more challenging. Models for tuber shape based on visual ratings exhibit much lower prediction accuracy than models for quantitatively measured traits in the same population (Sood et al., 2020).

Heritability is a crucial determinant of prediction accuracy (Kaler et al., 2022), and heritability depends on the precision and accuracy of phenotyping methods (Singh et al., 2019). Tuber quality traits are traditionally measured on ordinal visual scales (Reeves 1988; Van ek et al. 1994; Prashar et al. 2014; Buhrig et al., 2015). Such measures are often subjective, imprecise, and prone to errors due to variations in scoring and experience levels of individual raters (Parker et al., 1995; Poland and Nelson, 2011). One alternative to visual scales is digital imaging, which can produce precise and accurate quantitative measures of tuber quality traits including skin color and skin retention (Miller et al., 2023; Caraza-Harter & Endelman, 2020), tuber shape (Su et al., 2017; Miller et al., 2023; Feldman et al., 2024), and tuber blemishes (Leiva et al., 2024). The use of digital imaging methods can improve heritability estimates, and thus potentially prediction accuracy, over visual ratings (Miller et al., 2023). Furthermore, digital imaging can break down phenotypes into component parts such as lightness, hue, and chroma values for skin color (Caraza-Harter & Endelman, 2020). Across crops, measuring components of phenotypes often improves prediction accuracy over assessing the phenotype as a whole (Singh et al., 2019; Yonis et al., 2020; Yu et al., 2020).

In this study, we combine genotyping data with quantitative phenotypic scores for tuber shape and skin color generated from the R/TubAR package (Miller et al., 2023) to develop genomic selection models for chips and fresh market potatoes. Fresh market potatoes are generally marketed by skin color (red or yellow) and expected to be round. Vivid color makes a variety more marketable. Perceived color is determined not only by how red a tuber is (on a scale of green to red) but also how light it is (on a scale of black to white) (Caraza-Harter & Endelman, 2020). While there is some flexibility on tuber shape for fresh market tubers, chipping potatoes must be round in order to be compatible with processing equipment (Kirkman, 2007). We estimated and characterized the genetic variance components for each of these quality traits. Then, we evaluated our ability to predict shape and skin color using genomic prediction and assessed how each is affected by nonadditive genetic variation. Lastly, we developed selection indices for the chipping potatoes by combining our genomic estimated genotypic value (GEGV) for shape with estimates for yield and specific gravity and selected superior clones.

2.2 Materials and Methods

2.2.1 Plant materials

We evaluated advanced clones for 134 chippers and 81 fresh market red and yellow clones between 2020 and 2022 at the Sand Plains Research Farm (SPRF) in Becker, MN. The 77 chipping and 38 fresh market clones were from crosses first grown in 2018 and 38 chipping and 24 fresh market clones were from crosses first grown in 2019. Commonly grown varieties were used as checks, including Atlantic (Webb et al., 1978), Superior (Rieman, 1962), Lamoka (De Jong et al., 2017), Snowden, and Cascade for the chip processing market class, and Red Norland (Johansen et al., 1959), Red Pontiac, Chieftain (Weigle et al., 1968), and Yukon Gold (Johnston

and Rowberry, 1981) for the fresh market class. In 2022, fourteen clones from the University of Wisconsin (UW) Potato Breeding Program and seven clones from the North Dakota State University (NDSU) Potato Breeding Program were included in this study. We evaluated these clones in an augmented block design with row/column information and repeated checks. All plots contained 15 seed pieces with 30 cm in-row spacing and 91 cm between rows.

2.2.4 Phenotyping

We collected phenotype data for yield, specific gravity, roundness, redness, lightness, and length to width ratio (L/W). Yield is reported as the total plot weight for all harvested tubers in Mg ha^{-1}. In 2020 and 2021 yield was measured on a tuber basis and summed over plots using a modified AgRay™ potato grader. In 2022 yield was measured on a tuber basis and summed over plots using an AccuVision Dual™ View L.E.D. Grader System from Exeter. Specific gravity was determined by weighing a ten-tuber sample from each plot in air and then weighing them again when submerged in water. The specific gravity was then calculated by dividing the weight in air by the difference between the weight in air and the weight in water (Kleinschmidt et al., 1984).

All other traits were measured using an image-based analysis tool. Image samples of ten harvested and washed tubers from each plot were taken using an Ortery Photosimile 200 Lightbox with a Cannon Rebel™ T6i camera following the methods from Caraza-Harter and Endelman (2020). The trait measurement was performed using the potato image analysis R package, TubAR (Miller et al., 2023). Length width ratio and roundness were calculated for each tuber, then the median tuber value for each trait in each image was used as the trait value for each plot. Tuber skin color was quantified in terms of redness and lightness. The median tuber value for each image is used for analysis.

To exclude outliers for each trait, we removed values that were two standard deviations from the mean to obtain normally distributed data for all traits (Figure S1). We excluded two to thirteen data points per trait for the chips, and seventeen to forty for the fresh market potatoes.

2.2.5 Genotyping

The clones were genotyped with different potato SNP array platforms (Table S1). A total of 97 University of Minnesota (UMN) clones were genotyped using the Gene-seek genomic profiler (GGP) 31 k V4 Potato Array based on the Illumina Infinium array technology and GeneSeek custom content (Felcher et al., 2012; Vos et al., 2015). The clones from UW and NDSU were genotyped using the 12k array V2 and the remaining 89 UMN clones were genotyped using a 2K targeted GBS platform, DArTag (Endelman et al. 2024; Table S1). The fitPoly R package (Voorrips et al., 2011; Zych et al., 2019) was used to call genotypes. The F1 genotypes from the Premier Russet × Rio Grande mapping population (Douches et al., 2014; Schmitz Carley et al., 2017) were used to check for shifted markers using the *checkF1* and *correctDosages* function in fitPolyTools. Marker genotypic classes were described using dosage (0-4). The genotypic datasets were merged and imputed using the *merge_impute* function of the R package polyBreedR (Endelman, 2023b)(https://github.com/jendelman/polyBreedR). The merging and imputation of all the dataset retained a total of 21,531 markers which we used to carry out further analysis. The data were thereafter separated into different market classes and two covariance relationship matrices for additive (G) and dominance (D) genetic effects were constructed from the markers in the StageWise package (Endelman, 2023a).

2.2.6 Variance partitioning and genomic prediction

A two-stage analysis (Damesa et al., 2017) with the StageWise (Endelman, 2023a) R package was used to partition variance and develop a genomic selection model for all the studied traits. The analysis in stage one computes best linear unbiased estimates (BLUE) for each genotype in each year as well as heritability on a plot basis. We fit a spatial model using a 2D spline, implemented with SpATS (Rodríguez-Álvarez et al., 2018), within the StageWise package in R (R Core Team, 2022) based on field position information. In stage two, we used the BLUEs from stage one as response variable to develop the linear mixed model:

$$BLUE[g_{ij}] = y_{ij} = E_j + g_i + gE_{ij} + s_{ij} \qquad (1)$$

Where g_{ij} is the genotypic value for individual i in year j, E_j is the fixed effect for year j, g_i is the random effect for genotype i and the GxE effect, gE_{ij} is the model residual. The s_{ij} random effect accounts for the error in Stage 1 and follows a multivariate normal distribution based on the variance-covariance matrix of the Stage 1 BLUEs.

After modeling g_i effects as independent, marker information was incorporated. This enabled the partitioning of genetic variance into additive effect and nonadditive effect. The additive effect is $\mathbf{a} \sim N(0, \mathbf{G}\sigma_A^2)$, where $\mathbf{G}$ is the genomic additive relationship matrix and σ_A^2 is the genomic additive variance. Nonadditive effects were modeled in two ways, first as the genetic residual (R) and then as directional dominance (D). The genetic residual (R) is independently identically distributed (i.i.d), and the dominance effect is $\mathbf{d} \sim N(-b\mathbf{F}, \mathbf{D}\sigma_D^2)$, where $\mathbf{D}$ is the dominance relationship matrix, σ_D^2 is the dominance variance, and $\mathbf{F}$ are genomic inbreeding coefficients (Endelman, 2023a). We thereafter tested the different genetic models based on additive covariance structure (G), additive plus dominance covariance structure (G+D) or additive plus genetic residual (G+R). StageWise (Endelman, 2023a) estimates the proportion

of variance explained (PVE) by each effect without the main effect E_j, which is the variance of each effect divided by the sum of all variance effects.

Empirical BLUPs for total genotypic value were computed as implemented in the StageWise package (Endelman, 2023a) for each of the three genetic models. The marker based and non-marker-based estimates for total genetic values were compared based on their reliability scores (r^2), estimated from prediction error variance (PEV) in StageWise. For each trait in both market classes, we determined the best fit model based on the Akaike Information Criterion (AIC).

Thereafter, we implemented a ten-fold cross-validation scheme replicated five times to estimate prediction ability for each genetic model. In this scheme, and for each replication, we used each fold as validation data by masking phenotypes from the training set and then predicting phenotypes for the masked individuals. The prediction for each trait was then estimated as the correlation between the genomic predictions and the validation data, calculated by BLUP assuming independent clone effects.

2.2.7 Multi program model

To investigate the role of population size in prediction ability, we combined yield and genotype data from 984 chipping clones from the UW breeding program and the National Chip Processing Trial evaluated in Wisconsin between 2015 and 2020 (Endelman, 2023a), with the UMN chipping data described above. We hypothesized that these populations were sufficiently related to be informative, because there is minimal population structure in US chipping potatoes (Love et al., 1999; Agha et al., in review) and many of the UW and UMN clones share parents and grandparents. The variance partitioning and genomic prediction steps from section 2.4 were repeated on the combined data set.

2.2.8 Selection index

We created a selection index to summarize the genetic merits of each clone across traits. Due to population size, we were only able to create a selection index for chipping potatoes. We identified top performing clones based on roundness, yield, and specific gravity. We used weighted multi-trait index selection within StageWise to identify top performing clones. The merit for each clone was determined based on the following expression:

$$SI = w_1 \text{ x } GEGV_{yield} + w_2 \text{ x } GEGV_{sg} + w_3 \text{ x } GEGV_{roundness}$$

Where SI is the selection index, w_1, w_2 and w_3 were the assigned weights for each trait, and GEGV is the standardized total genotypic value for the trait indicated in the subscript. We considered two scenarios for calculating the selection index for each clone. For scenario one, we assigned equal weights (+3) across the three traits, and for scenario two, we assigned equal weight for yield and specific gravity and a lesser weight (+2) for roundness.

2.3 RESULTS

2.3.1 Phenotypic and genotypic analysis

For each trait, in each year, we calculated BLUEs for individuals and broad sense heritabilities for breeding populations in each market class. The broad sense heritability within the fresh market clones for redness was between 76% and 91%, lightness ranged between 34% and 86%, roundness ranged between 45% and 85%, and length-width ratio (LW) was between 50% and 67% (Figure 1A). In general, heritabilities were highest in 2021, although for color traits they were equally high in 2020. Within the chips, broad sense heritability for roundness, ranged between 23% and 29%, specific gravity was between 59% and 83%, and yield between

22

23% and 87% (Figure 1B). The BLUEs and the corresponding covariance obtained were utilized for further analysis.

We removed SNPs with minor allele frequencies lower than 5% which resulted in 21,471 markers for the chipping clones and 20,753 makers for fresh market clones. These were used to calculate covariance matrices for additive (G) and dominance (D) relationships.

2.3.2 Variance components and genomic prediction

The genotype effects were first modeled without markers to estimate total genetic variance (V_g) and variance of genotype by year (V_{gy}). As shown in (Table 1), the relationship between estimates for V_g and V_{gy} depends on the trait in question.

We considered three models for incorporating genetic data: one using just the additive covariance matrix (G), one using the additive and dominance covariance matrices (G+D), and one using the additive covariance matrix and the genetic residual (G+R). For each model, we performed variance component decomposition to portion the genetic variance for each trait into additive and non-additive (dominance or genetic residual) components. The results are reported here as the proportion of variance explained for each effect (Figure 2). For fresh market clones, all traits except length to width ratio exhibited large genetic components. While, for chipping clones, GxE was the primary contributor to variance for most traits. For quality traits measured through image analysis, G and G+R produced very similar results. However, for yield and specific gravity in chipping clones G+R introduced a genetic residual component larger than the additive component. The use of G+R lowered AIC for all traits in both market classes except lightness where G+D produced a better fit (Table 2). Using the G+D model increased the additive genetic variance for lightness from 46% to 47.2% (Figure 2).

We estimated a total genetic value and a breeding value for each clone-trait combination. Genetic values were calculated first based on phenotypic data alone without markers and then with marker assisted estimation. We compared reliability scores for each method (r^2). Estimates of total genetic value made using marker data (Table S2) were more reliable than those made without marker data for all quality traits (Figure 3). Notably, the reliability for phenotype-only predictions for roundness in chipping clones was 0 for all individuals. Reliability for marker assisted predictions for the same trait was as high as 0.5, suggesting that marker data is required to make shape predictions for chipping potatoes. However, the total genetic value calculated with and without marker data showed similar reliability scores for yield and specific gravity for the chipping potatoes (Figure 3B).

We used cross-validation to determine our ability to predict total genetic value for each trait-market class combination. G and G+R were equally effective at predicting quality traits in both market classes (Figure 4). For yield and specific gravity in the chips, G+R reduced our prediction ability. Including dominance in the G+D model improved our ability to predict roundness in the chip processing clones, had no effect on predictions of lightness, and reduced our ability to predict other quality traits. In chipping potatoes, G and G+D were equally effective predictors of yield and G+D and G+R were equally ineffective predictors of specific gravity.

2.3.3 Multi program model

In order to investigate the effect of population size, we combined our chipping potato data with a published dataset from the UW potato breeding program (Endelman, 2023a). The combined data set included 1051 individuals with phenotypic and genotypic information. Including this data more than doubled our estimates for the additive variance component for all three models (figure 5A) and enhanced our ability to predict yield (figure 5B).

2.3.4 Selection index

A successful potato cultivar must combine a series of yield and quality traits, and selecting for some traits can inadvertently select against others. Therefore, we created a selection index with yield, specific gravity, and roundness for chipping potatoes. We considered two indices, one with equal weights for all traits along with one that emphasized yield and specific gravity over roundness. The top performing clones for both indices overlapped (Table S3 and Table S4). Of the ten top performing clones, eight were either clones from the UW or were the result of crosses originally done at the UW selected in Minnesota.

2.4 Discussion

2.4.1 Image based phenotyping allows us to build genomic selection models for potato quality traits

High throughput phenotypic data and genomic assisted selection can improve selection accuracy and accelerate genetic gain in a breeding program (Slater et al., 2016; Bradshaw, 2017; Bykova et al., 2017). Although genomic selection has been successfully implemented in potato for yield and yield components (Endelman et al., 2018; Gemenet et al., 2020; Habyarimana et al., 2017; Selga et al., 2021; Sood et al., 2020; Stich & Van Inghelandt, 2018; Martins et al, 2023), disease resistance (Gemenet et al., 2020; Selga et al., 2021; Sood et al., 2020; Stich & Van Inghelandt, 2018; Sverrisdottir et al., 2018; Enciso-Rodriguez et al., 2018), nutritional value (Pandey et al., 2023), and processing traits (Endelman et al., 2018; Sverrisdottir et al., 2017; Habyarimana et al., 2017; Sood et al., 2020; Stich & Van Inghelandt, 2018; Caruana et al., 2019; Sverrisdottir et al., 2018; Martins et al, 2023), prediction of tuber appearance has been less successful (Sood et al.,

25

2020). However, tuber appearance is crucial for marketability, especially for fresh market

potatoes. In this study we combined quantitative phenotypes generated from image analysis and

genomic data to characterize variance components and develop genomic selection models for

tuber shape and skin color.

Measurement of quality traits in potatoes has historically relied on visual scores which are

subjective and prone to errors. The image-based tool known as TubAR, was developed as an R

package (Miller et al., 2023) to improve the precision and accuracy of these phenotypes.

Between 2020 and 2022 we photographed subsamples of fresh market and chipping clones from

our breeding program yield trials and measured tuber color and shape using TubAR. These

phenotypes exhibited a moderately normal distribution across all years evaluated (Figure S1).

The heritability values for traits differ widely across years except for redness which ranged

between 76% and 91% (Figure 1). This is consistent with previous studies in similar populations

which show high GxE for roundness (Stefaniak et al., 2021; Jones et al., 2021), lightness

(Stefaniak et al., 2021; Jones et al., 2021), and yield (Agha et al., In review; Schmitz Carley et

al., 2019). The stability of redness across years contributes to higher prediction ability (Figure 4).

Similarly, the narrow sense heritability was highest for redness within the fresh market class,

followed by roundness and lightness, which suggests potential for the improvement of these

traits (Figure 2).

We considered models with additive only and both additive and non-additive covariance

structures. For yield and specific gravity, allowing for only additive covariance increased the

additive component over models where non-additive covariance was included. The additive

values increased from 9% to 16.7% and 15.9% to 22.1% respectively. This suggests that the

additive effect from additive only covariance models actually captures some of the non-additive

genetic variance for most traits considered. This effect has been observed in other highly heterozygous clonal crops (Wolfe et al., 2016; de Andrade et al., 2022).

2.4.2 Color

We were most accurate in our predictions for color (Figure 4), which exhibited high broad (Figure 1) and narrow sense (Figure 2) heritability. Heritability estimates were lowest in 2022, most likely due to incomplete drying of tubers before photography, which affects color in general and lightness in particular.

We separated color into two component traits, redness and lightness. These traits were highly correlated (R^2=-0.8, Figure 6), with darker tubers also generally being redder. This is consistent across other attempts to implement multi-dimensional analysis of color (Caraza-Harter & Endelman, 2020). However, the correlation is not perfect, and in particular we observe extensive variation in lightness among the reddest tubers. It is also worth noting that heritability for lightness varied more than heritability for redness, possibly because lightness is more influenced by environmental traits like low nitrogen than redness (Stefaniak et al., 2021). As a result, the clones identified as the best parents (highest GEBV) for redness were not the same as the clones identified as the best parents for lightness, suggesting the possibility of improving appearance by crossing strong parents for each of the two components of color. Using image analysis allows us to break down appearance, a crucial trait for consumers, into component parts with different genetic basis, thus improving our ability to choose crosses most likely to produce attractive offspring.

Potatoes from both the red-skin and yellow-skin germplasm groups were included in our fresh market trials, leading to a bimodal distribution for redness (Figure S1 A), but nonetheless a unimodal distribution for lightness (Figure S1 B). Some of the variation in redness is most likely

accounted for by known large effect color genes, StAN2 (Jung et al., 2009) and StDFR (De Jong et al., 2003), and this accounts for some of the high heritability, but we observe continuous variation in GEBVs consistent with a multigenic genetic architecture.

2.4.3 Shape

Roundness was the only trait for which we built a model in each market class. The additive genetic variance for roundness was much higher in the fresh market clones (41.9%) than in the chips (10.3%) (Figure 2). This is most likely due to higher total variation for tuber shape in fresh market potatoes (0.93-0.99) as compared to chipping potatoes (0.98-0.99). While there is a consumer preference for round fresh market red and yellow potatoes, the shape requirements in chipping potatoes are determined by processing equipment and therefore must be more strictly adhered to.

Despite this difference in additive variance, our prediction ability was similar for both market classes. The exception being that including dominance covariance in the model decreased our prediction ability for chips but improved it for fresh market clones (Figure 4). We also found evidence for dominance variation and heterosis for roundness in fresh market potatoes that we did not see in chipping potatoes (Figure 2). This could be explained by OFP20, a known large effect shape gene (Wu et al., 2018; Van Ek et al., 2022) that exhibits partial dominance (Endelman & Jansky, 2016). Variation at OFP20 could explain the positive effect of including dominance covariance on prediction accuracy for fresh market potatoes.

Although the shape of chipping potatoes exhibits little variation, and a small proportion of that variance is additive, there is still some variation to be selected upon and chipping potato shape could be improved. This fine tuning of shape is impossible with phenotypic selection (Figure 3) and would be even more so were we relying on visual ratings. The combination of

28

high throughput phenotyping and genomic selection allows us to make progress on otherwise hard to tackle quality traits.

2.4.4 Yield and specific gravity

Compared to the quality traits, we found low prediction ability for yield and specific gravity (Figure 4). Our prediction abilities for specific gravity were slightly lower than those reported by Endelman et al. (2018) and much lower than those reported by Sood et al. (2020). Similarly, our prediction accuracies for yield, were lower than those of previous studies, which ranged from 0.2-0.6 (Endelman et al., 2018; Gemenet et al., 2020; Habyarimana et al., 2017; Selga et al., 2021; Sood et al., 2020; Stich & Van Inghelandt, 2018). Combining data from our 134 chipping clones with data from the 984 from Endelman (2023a) substantially improved our ability to predict yield (Figure 5), consistent with the general observation that increased sample size increases prediction ability (Heffner et al., 2009). This suggests our prediction models will improve as we add more years of data, and this pattern is likely to hold for traits beyond yield. We developed models for yield and specific gravity to facilitate index selection, which allows breeders to balance the required combination of traits for a successful potato variety. The top chipping potatoes we identified, which combine yield, specific gravity, and roundness, and were all developed in the UW program where genomic selection for yield and specific gravity in chipping potatoes has been in practice for several years (Endelman et al., 2018; Endelman 2023a). This result demonstrates the success of multiple rounds of genomic selection. Additionally, the fact that clones developed at UW using genomic selection also perform well at UMN, suggests the potential for joint multi-environment models across the two states, increasing the population size. The efficacy of combining published data from UW with UMN breeding program data (Figure 5), further supports this possibility. Models which incorporate multiple

state breeding programs present a promising path to improving selection and identifying broadly

adapted varieties.
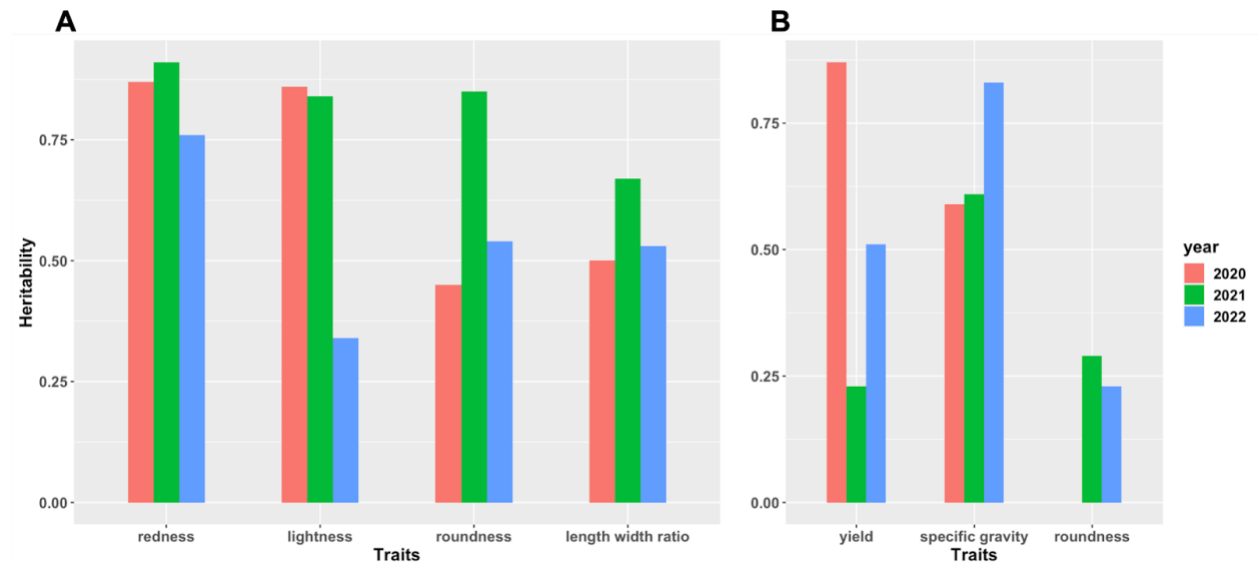
Figure 1. Broad sense heritability for each trait in each year measured in A) fresh market clones and B) chipping clones.

Table 1. Variance parameter estimates from baseline independent genotype effects.

	Fresh market				Chips		
Variance components	Redness	Lightness	Roundness	Length/width	Yield (Mg^2ha^{-2})	Specific gravity	roundness
V_g	6.3	14.7	6.4×10^{-5}	3.3×10^{-3}	80	3.2×10^{-5}	0
V_{gy}	17.3	31.5	1.63×10^{-4}	4.9×10^{-3}	77	4.75×10^{-5}	4.1×10^{-5}
Stage1.error	4.3	28.9	1.46×10^{-4}	1.23×10^{-2}	111	1.78×10^{-5}	2.98×10^{-5}

Table 2. Comparison of AIC for nonadditive models

	Fresh market				**Chips**		
	Redness	Lightness	Roundness	Length/width	Yield	Specific gravity	Roundness
G	428.94	581.76	-809.31	-391.17	1271.68	-1673.72	-1583.75
G+D	430.94	578.12	-806.67	-386.36	1268.18	-1663.71	-1575.53
G+R	425.89	583.76	-807.30	-389.17	1265.72	-1673.71	-1581.75
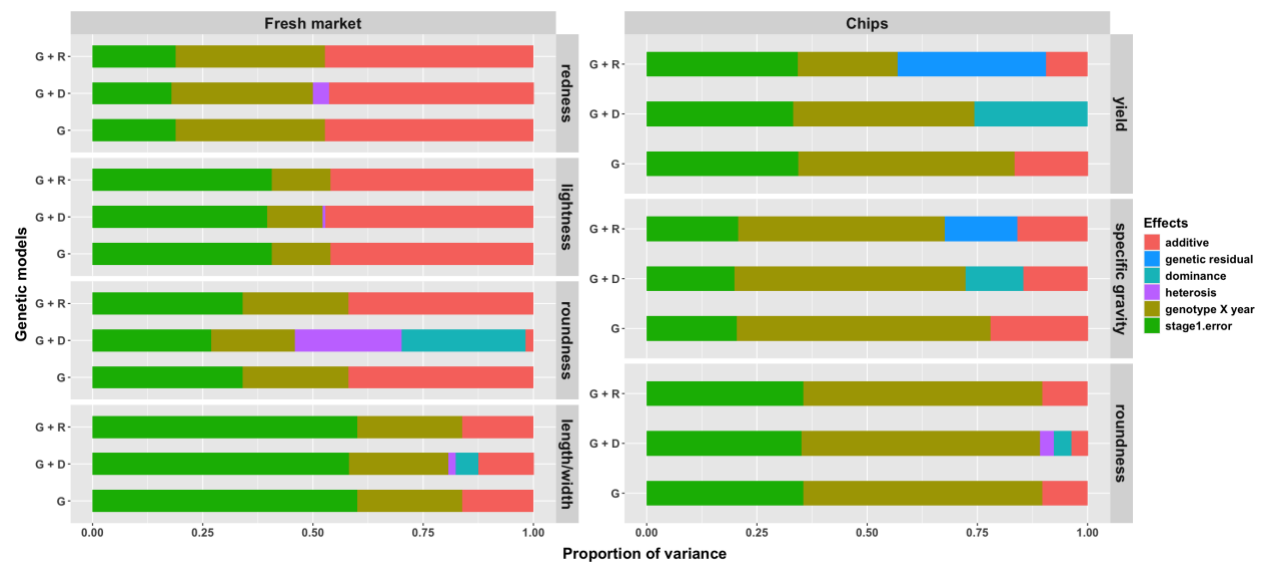
Figure 2. Proportion of variance explained for each trait based on different genetic models (G, G+D and G+R) for each market class.
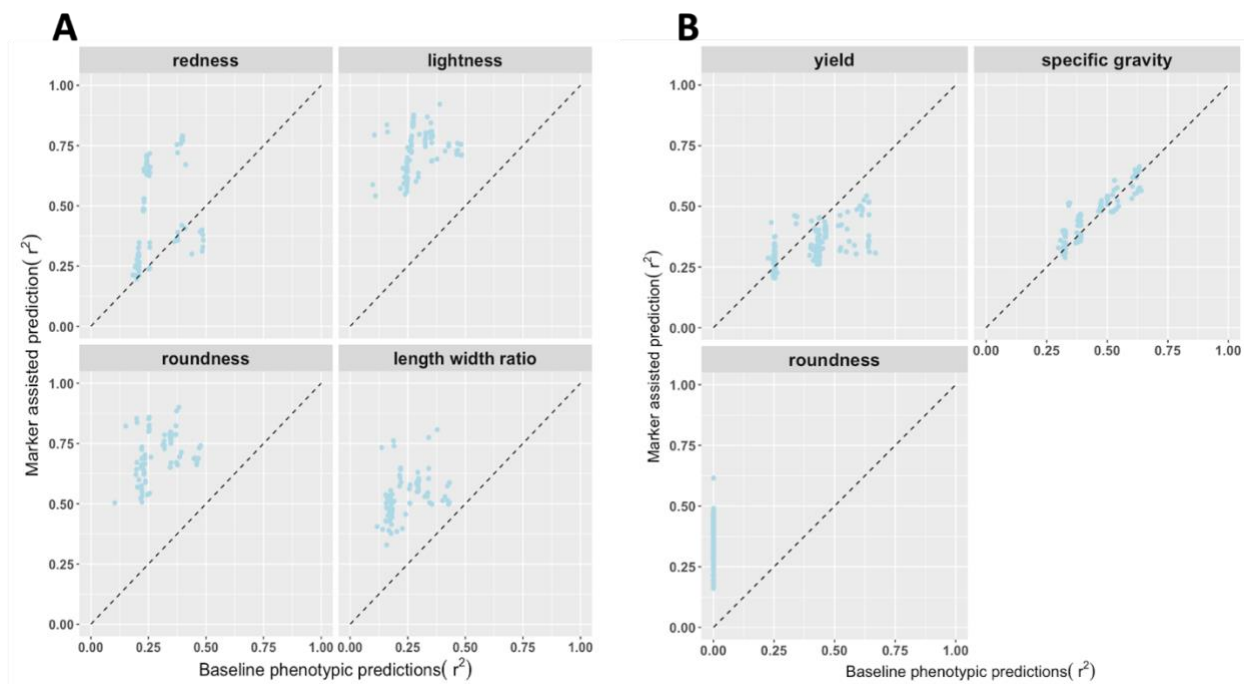
Figure 3. Correlation between reliability (r2) of predictions using all data and of predictions made using only phenotypic data for each trait in A) fresh market clones and B) chips clones.
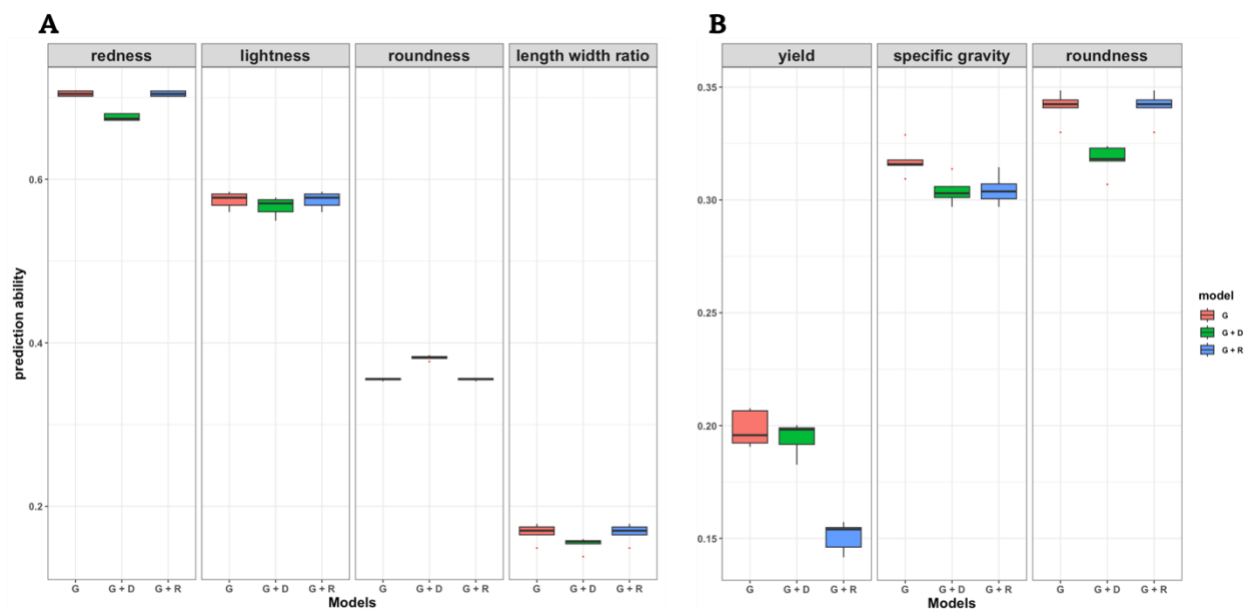
Figure 4. Comparison of each model's ability to predict total genetic value for each trait in chips (A) and fresh market clones (B) as determined by 10-fold cross validation.
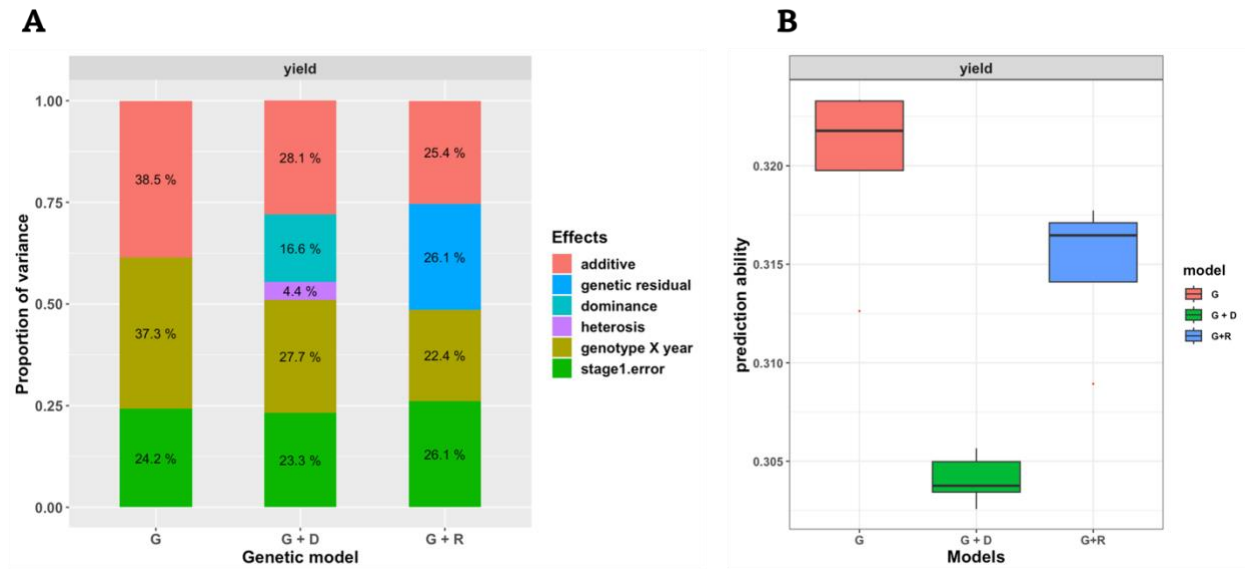
Figure 5. Increasing population size improved A) genetic variance components estimation as compared to Figure 2 and B) prediction ability for yield in chipping potatoes as compared to Figure 4B.
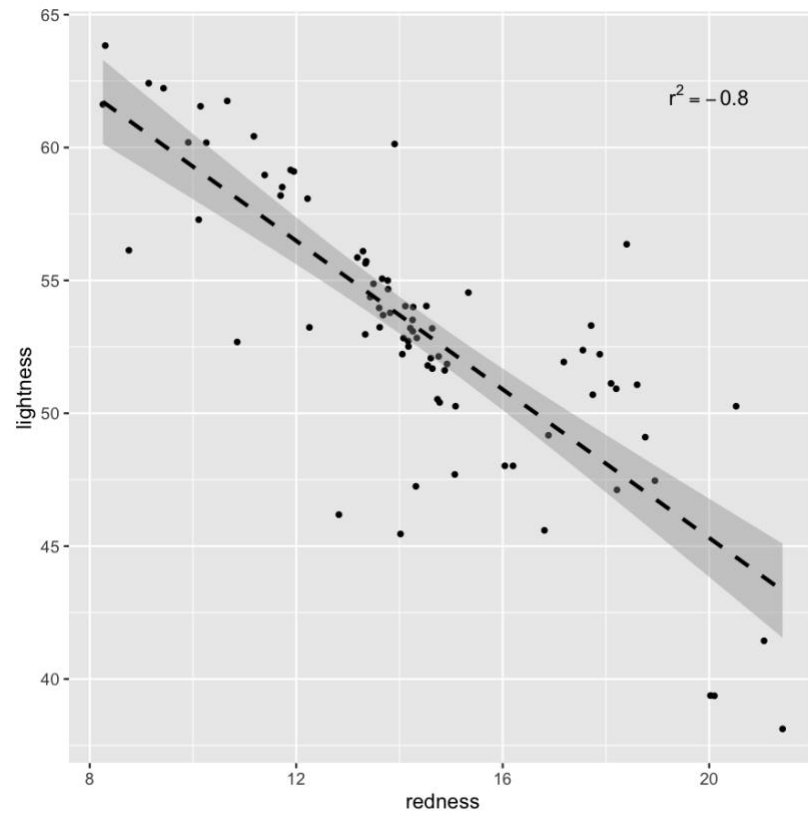
Figure 6. The correlation between GEBVs for redness and lightness. The dotted line indicates the linear regression line and shaded area indicates the 95% confidence interval.

References

Agha, H. I., Endelman, J. B., Chitwood-Brown, J., Clough, M., Coombs, J., de Jong, W. S., Douches, D. S., Higgins, C., Holm, D., Novy, R., Resende, M. F. R., Sathuvalli, V., Thompson, A. L., Yencho, G. C., Zoterelli, L., Shannon, L.M. (In Review). Genotype-by-Environment interactions and local adaptation shape selection in the United States National Chip Processing Trial.

de Andrade, L. R. B., Sousa, M. B. e, Wolfe, M., Jannink, J.-L., de Resende, M. D. V., Azevedo, C. F., & de Oliveira, E. J. (2022). Increasing cassava root yield: Additive-dominant genetic models for selection of parents and clones. Frontiers in Plant Science, 13. 10.3389/fpls.2022.1071156

Bonar, N., Liney, M., Zhang, R., Austin, C., Dessoly, J., Davidson, D., Stephens, J., McDougall, G., Taylor, M., Bryan, G.J., & Hornyik, C. (2018). Potato miR828 is associated with purple tuber skin and flesh color. Frotiers in Plant Science. 9. https://doi.org/10.3389/fpls.2018.01742

Bradshaw, J. E. (2017). Review and Analysis of Limitations in Ways to Improve Conventional Potato Breeding. Potato Research, 60(2), 171–193. doi: 10.1007/s11540-017-9346-z.

Buhrig, W., Thorton, M. K., Olsen, N., Morishita, D., & McIntosh, C. (2015). The influence of ethepon application timing and rate on plant growth, yield, tuber size distribution, and skin color of Red LaSoda potatoes. American Journal of Potato Research, 92, 100-108.

Bykova, I. V., Shmakov, N. A., Afonnikov, D. A., Kochetov, A. V., & Khlestkina, E. K. (2017). Achievements and prospects of applying high-throughput sequencing techniques to potato genetics and breeding. Russian Journal of Genetics: Applied Research, 7(7), 736–743. doi: 10.1134/S2079059717070036.

Caraza-Harter, M. V., & Endelman, J. B. (2020). Image-based phenotyping and genetic analysis of potato skin set and color. Crop Science, 60(1), 202-210. https://doi.org/10.1002/csc2.20093

Carputo, D., Aversano, R., & Frusciante L. (2004). Breeding potato for quality traits. Meeting of the Physiology Section of the European Association for Potato Research 684 :55-64. https://doi.org/10.17660/ActaHortic.2005.684.7

Caruana, B. M., Pembleton L. W., Constable F., Rodoni B., Slater A. T., Cogan N. O. I. (2019). Validation of Genotyping by Sequencing Using Transcriptomics for Diversity and Application of Genomic Selection in Tetraploid Potato. Frontiers in Plant Science, 10. https://doi.org/10.3389/fpls.2019.00670

Damesa, T. M., Möhring, J., Worku, M., & Piepho, H.-P. (2017). One Step at a Time: Stage-Wise Analysis of a Series of Experiments. Agronomy Journal, 109(3), 845–857. doi: 10.2134/agronj2016.07.0395.

De Jong, W.S., De Jong, D.M., De Jong, H., Kalazich, J., & Bodis, M. (2003). An allele of dihyrdoflavinol 4-reductase associated with the ability to produce anthocyanin pigments in potato. Theoretical and Applied Genetics, 107, 1375-1383. https://doi.org/10.1007/s00122-003-1395-9

De Jong, W.S., Halseth, D.E., Plaisted, R.L., Wang, X., Perry, K.L., Qu, X., Paddock, K.M., Falise, M., Christ, B.J., & Porter, G.A. (2017). Lamoka, a variety with excellent chip color out of cold storage and resistance to golden cyst nematode. American Journal of Potato Research. 94, 148-152. https://doi.org/10.1007/s12230-016-9557-x

Douches, D.S., Mass, D., Jastrzebski, R., & Chase, W. (1996). Assessment of potato breeding progress in the USA over the last century. Crop Science. 36: 1544-1552. https://doi.org/10.2135/cropsci1996.0011183X003600060024x

Douches, D., Hirsch, C. N., Manrique-Carpintero, N. C., Massa, A. N., Coombs, J., Hardigan, M., Bisognin, D., De Jong, W., & Buell, C. R. (2014). The Contribution of the Solanaceae Coordinated Agricultural Project to Potato Breeding. Potato Research, 57(3-4), 215-224. https://doi.org/10.1007/s11540-014-9267-z

Enciso-Rodriguez, F., Douches, D., Lopez-Cruz, M., Coombs, J., & de los Campos, G. (2018). Genomic Selection for Late Blight and Common Scab Resistance in Tetraploid Potato (Solanum tuberosum). G3 Genes|Genomes|Genetics, 8(7), 2471–2481. https://doi.org/10.1534/g3.118.200273

Endelman, J. B. (2023a). Fully efficient, two-stage analysis of multi-environment trials with directional dominance and multi-trait genomic selection. Theoretical and Applied Genetics, 136(4), 65. doi: 10.1007/s00122-023-04298-x.

Endelman, J. (2023b). polyBreedR. Retrieved May 24, 2023 from https://github.com/jendelman/polyBreedR.

Endelman, J. B., & Jansky, S. H. (2016). Genetic mapping with an inbred line-derived F2 population in potato. Theoretical and Applied Genetics, 129(5), 935-943. doi:10.1007/s00122-016-2673-7

Endelman, J. B., Kante, M., Lindqvist-Kreuze, H., Kilian, A., Shannon, L. M., Vaillancourt, B., & Buell, C. R. (2024). Targeted genotyping-by-sequencing of potato and software for imputation. bioRxiv 2024.02.12.579978. https://doi.org/10.1101/2024.02.12.579978

Endelman, J. B., Schmitz Carley, C. A., Bethke, P. C., Coombs, J. J., Clough, M. E., da Silva, W. L., De Jong, W. S., Douches, D. S., Frederick, C. M., Haynes, K. G., Holm, D. G., Miller, J. C., Muñoz, P. R., Navarro, F. M., Novy, R. G., Palta, J. P., Porter, G. A., Rak, K. T., Sathuvalli, V. R., Thompson, A. L., & Yencho, G. C. (2018). Genetic Variance Partitioning and Genome-Wide Prediction with Allele Dosage Information in Autotetraploid Potato. Genetics, 209(1), 77–87. https://doi.org/10.1534/genetics.118.300685

FAO. Crop and Livestock products. License: CC BY-NC-SA 3.0 IGO. Extracted from: https://www.fao.org/faostat/en/#data/QCL. Data of Access: 01-10-24

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

Feldman, M. J., Park, J., Miller, N., Wakholi, C., Greene, K., Abbasi, A., Rippner, D., Navarre, D., Schmitz Carley, C., Shannon, L. M., Novy, R. (2024). A scalable, low-cost phenotyping strategy to assess tuber size, shape, and the colorimetric features of tuber skin and flesh in potato breeding populations. The Plant Phenome Journal, 7(1), e29900. https://doi.org/10.1002/ppj2.20099

Gemenet, D. C., Lindqvist-Kreuze, H., De Boeck, B., da Silva Pereira, G., Mollinari, M., Zeng, Z.-B., Yencho, G. C., & Campos, H. (2020). Sequencing depth and genotype quality: accuracy and breeding operation considerations for genomic selection applications in autopolyploid crops. Theoretical and Applied Genetics, 133, 3345–3363. https://doi.org/10.1007/s00122-020-03673-2

Habyarimana, E., Parisi, B., & Mandolino, G. (2017). Genomic prediction for yields, processing and nutritional quality traits in cultivated potato (Solanum tuberosum L.). Plant Breeding, 136, 245–252. https://doi.org/10.1111/pbr.12461

Heffner, E. L., Sorrells, M. E., & Jannink, J.-L. (2009). Genomic Selection for Crop Improvement. Crop Science, 49(1), 1–12. https://doi.org/10.2135/cropsci2008.08.0512

Jo, K. R., Choi, J.-G., Kwon, D.-H., Park, Y.-E., & Kim, S.-J. (2023). Revealing Genetic Variations Associated with Chip-Processing Properties in Potato (Solanum tuberosum L.). Agronomy, 13(3), 642. https://doi.org/10.3390/agronomy13030642

Johansen, R.H., Sandar, N., Hoyman, W.G., & Lana, E.P. (1959). Norland a new red-skinned potato variety with early maturity and moderate resistance to common scab. American Potato Journal, 36. 12-15. https://doi.org/10.1007/BF02877209

Johnston, G.R. & Rowberry, R.G. (1981). Yukon gold: A new yellow-fleshed medium-early, high quality table and French-fry cultivar. American Journal of Potato Research. 58, 241-244. https://doi.org/10.1007/BF02853905

Jones, C. R., Michales, T. E., Schmitz Carley, C., Rosen, C. J., & Shannon, L. M. (2021). Nitrogen uptake and utilization in advanced fresh-market red potato breeding lines. Crop Science, 61(2), 878–895. https://doi.org/10.1002/csc2.20297

Jung, C., Griffiths, H., De Jong, D., Cheng, S., Bodis, M., et al. (2009). The potato developer (D) locus encodes an R2R3 MYB transcription factor that regulates expression of multiple anthocyanin structural genes in tuber skin. Theoretical and Applied Genetics, 120, 45–57. https://doi.org/10.1007/s00122-009-1158-3

Kaler, A. S., Purcell, L. C., Beissinger, T., & Gillman, J. D. (2022). Genomic prediction models for traits with differing heritability in soybean, rice, and maize. BMC Plant Biology, 22, 87. https://doi.org/10.1186/s12870-022-03479-y

Kirkman, M.A. (2007). Global markets for processed potato products in *Potato Biology and Biotechnology*. https://doi.org/10.1016/B978-044451018-1/50044-0

Kleinschmidt, G. D., Kleinkopf, G. E., Westermann, D. T. and Zalewski, J. C. (1984). *Specific gravity of potatoes*. (Current Information Series No. 609). University of Idaho.

Leiva, F., Abdelghafour, F., Alsheikh, M., Nagy, N. E., Davik, J., & Chawade, A. (2024). ScabyNet, a user-friendly application for detecting common scab in potato tubers using deep learning and morphological traits. Scientific Reports, 14, 1277. https://doi.org/10.1038/s41598-023-51074-4f

Love, S.L. (1999). Founding clones, major contributing ancestors and exotic progenitors of prominent North American potato cultivars. American Journal of Potato Research, 76:263-272. https://doi.org/10.1007/BF02853624

Martins, V. S., Andrade, M. H. M. L., Padua, L. N., Miguel, L. A., Filho, C. C. F., et al. (2023). Evaluating the impact of modeling the family effect for clonal selection in potato breeding programs. Frontiers in Plant Science, 14, 1253706. https://doi.org/10.3389/fpls.2023.1253706

Meijer, D., Viquez-Zamora, M., van Eck, H.J., Hutten, R.C.B., Su, Y., Rothengatter, R., Visser, R.G.F., Lindhout, W.H., & van Heudsen, A.W. (2018). QTL mapping in diploid potato by using selfed progenies of the cross S. tuberosum x S. chacoense. Euphytica, 214, 121. https://doi.org/10.1007/s10681-018-2191-6

Miller, M. D., Schmitz Carley, C. A., Figueroa, R. A., Feldman, M. J., Haagenson, D., & Shannon, L. M. (2023). TubAR: An R package for quantifying tuber shape and skin traits from images. American Journal of Potato Research. https://doi.org/10.1007/s12230-022-09894-z

Pandey, J., Scheuring, D.C., Koym, J.W., Endelman, J.B., & Vales, M.I. (2023). Genomic selection and genome-wide association studies in tetraploid chipping potatoes. The Plant Genome, 16(1), e20297. https://doi.org/10.1002/tpg2.20297

Parker, S.R., Shaw, M.W., & Royle, D.J. (1995). The reliability of visual estimates of disease severity on cereal leaves. Plant Pathology. 44(5):856-864. https://doi.org/10.1111/j.1365-3059.1995.tb02745.x

Poland, J. A., & Nelson, R. J. (2011). In the eye of the beholder: the effect of rater variability and different rating scales on QTL mapping. Phytopathology 101(2): 290-298. 10.1094/PHYTO-03-10-0087

Prashar, A.C., Hornyik, C., Young, V., McLean, K., Sharma, S.K., & Dale, M.F.B., Bryan, G.J. (2014). Construction of a dense SNP map of a highly heterozygous diploid potato population and QTL analysis of tuber shape and eye depth. Theoretical and Applied Genetics, 127, 2159-2171.

R Core Team. (2022). CRAN mirrors. Retrieved May 25, 2023, from https://cran.r-project.org/mirrors.html

Reeves, A.F. (1988). Varietal differences in potato tuber greening. American Potato Journal, 65, 651-658.

Rieman, G.H. (1962). Superior: A new white, medium-maturing, scab-resistant potato variety with high chipping quality. American Potato Journal, 39, 19-28.

Rodríguez-Álvarez, M.X., Boer, M.P., van Eeuwijk, F.A., & Eilers, P.H.C. (2018). Correcting for spatial heterogeneity in plant breeding experiments with P-splines. Spatial Statistics, 23, 52–71. https://doi.org/10.1016/j.spasta.2017.10.003

Schmitz Carley, C. A., Coombs, J. J., Douches, D. S., Bethke, P. C., Palta, J. P., Novy, R. G., & Endelman, J. B. (2017). Automated tetraploid genotype calling by hierarchical clustering. Theoretical and Applied Genetics, 130(4), 717–726. https://doi.org/10.1007/s00122-016-2845-5

Schmitz Carley, C.A., Coombs, J.J., Clough, M.E., De Jong, W.S., Douches, D.S., et al. (2019). Genetic covariance of environments in the potato national chip processing trial. Crop Science, 59(1), 107-114. https://doi.org/10.2135/cropsci2018.05.0314

Selga, C., Koc, A., Chawade, A., & Ortiz, R. (2021). A bioinformatics pipeline to identify a subset of SNPs for genomics-assisted potato breeding. Plants, 10, 30. https://doi.org/10.3390/plants10010030

Singh, D., Wang, X., Kumar, U., Gao, L., Noor, M., et al. (2019). High-throughput phenotyping enabled genetic dissection of crop lodging in wheat. Frontiers in Plant Science, 10. https://doi.org/10.3389/fpls.2019.00394

Slater, A.T., Cogan, N.O.I., Forster, J.W., Hayes, B.J., & Daetwyler, H.D. (2016). Improving genetic gain with genomic selection in autotetraploid potato. The Plant Genome, 9(3), plantgenome2016.02.0021. https://doi.org/10.3835/plantgenome2016.02.0021

Sood, S., Lin, Z., Caruana, B., Slater, A.T., & Daetwyler, H.D. (2020). Making the most of all data: Combining non-genotyped and genotyped potato individuals with HBLUP. The Plant Genome, 13(3), e20056. doi: 10.1002/tpg2.20056

Stefaniak, T.R., Fitzcollins, S., Figueroa, R., Thompson, A.L., Schmitz Carley, C., & Shannon, L.M. (2021). Genotype and variable nitrogen effects on tuber yield and quality for red fresh market potatoes in Minnesota. Agronomy, 11(2), 255. https://doi.org/10.3390/agronomy11020255

Stich, B., & Van Inghelandt, D. (2018). Prospects and potential uses of genomic prediction of key performance traits in tetraploid potato. Frontiers in Plant Science, 9. https://doi.org/10.3389/fpls.2018.00159

Su, Q., Kondo, N., Li, M., Sun, H., & Al Riza, D.F. (2017). Potato feature prediction based on machine vision and 3D model rebuilding. Computers and Electronics in Agriculture, 137, 41-51. https://doi.org/10.1016/j.compag.2017.03.020

Sverrisdóttir, E., Byrne, S., Sundmark, E. H. R., Johnsen, H. Ø., Kirk, H. G., Asp, T., Janss, L., & Nielsen, K. L. (2017). Genomic prediction of starch content and chipping quality in tetraploid potato using genotyping-by-sequencing. Theoretical and Applied Genetics, 130, 2091–2108. https://doi.org/10.1007/s00122-017-2944-y

Sverrisdóttir, E., Sundmark, E. H. R., Johnsen, H. Ø., Kirk, H. G., Asp, T., Janss, L., Bryan, G., & Nielsen, K. L. (2018). The value of expanding the training population to improve genomic selection models in tetraploid potato. Frontiers in Plant Science, 9. https://doi.org/10.3389/fpls.2018.01118

Van Eck, H.J., Jacobs, J.M., Stam, P., Ton, J., Stiekema, W.J., & Jacobsen, E. (1994). Multiple alleles for tuber shape in diploid potato detected by qualitative and quantitative genetic analysis using RFLPs. Genetics, 137, 303-309.

van Eck, H.J., Oortwijn, M.E.P., Terpstra, I.R., van Lieshout, N.H.M., van der Knaap, E., Willemsen, J.H., & Bachem, C.W.B. (2022). Engineering of tuber shape in potato (Solanum tuberosum) with marker assisted breeding or genetic modification using StOFP20. Research Square, Preprint. https://doi.org/10.21203/rs.3.rs-1807189/v1

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

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

Webb, R.E., Wilson, D.R., Shumaker, J.R., Graves, B., Henninger, M.R., Watts, J., Frank, A.J., & Murphy, H.J. (1978). Atlantic: A new potato variety with high solids, good processing quality, and resistance to pests. American Journal of Potato Research. 55, 141-145. https://doi.org/10.1007/BF02852087

Weigle, J.L., Kehr, A.E., Akeley, R.V., & Horton, J.C. (1968). Chieftain: A red-skinned potato with attractive appearance and broad adaptability. American Journal of Potato Research. 45, 293-296. https://doi.org/10.1007/BF02850285

Wolfe, M.D., Kulakow, P., Rabbi, I.Y., & Jannink, J.-L. (2016). Marker-based estimates reveal significant nonadditive effects in clonally propagated cassava (Manihot esculenta): Implications for the prediction of total genetic value and the selection of varieties. G3 Genes|Genomes|Genetics, 6(11), 3497–3506. https://doi.org/10.1534/g3.116.033332

Wu, S., Zhang, B., Keyhaninejad, N., Rodríguez, G.R., Kim, H.J., et al. (2018). A common genetic mechanism underlies morphological diversity in fruits and other plant organs. Nature Communications, 9. https://doi.org/10.1038/s41467-018-07216-8

Yonis, B.O., del Carpio, D.P., Wolfe, M., Jannink, J.L., Kulalow, P., & Rabbi, I. (2020). Improving root characterization for genomic prediction in cassava. Scientific Reports, 10, 8003. https://doi.org/10.1038/s41598-020-64963-9

Yu, X., Lieboff, S., Li, X., Guo, T., Ronning, N., et al. (2020). Genomic prediction of maize microphenotypes provides insights for optimizing selection and mining diversity. 18(12), 2456-2465. https://doi.org/10.1111/pbi.13420

Zych, K., Gort, G., Maliepaard, C.A., Jansen, R.C., & Voorrips, R.E. (2019). FitTetra 2.0 – improved genotype calling for tetraploids with multiple population and parental data support. BMC Bioinformatics, 20(1), 148. https://doi.org/10.1186/s12859-019-2703-y.

Chapter 3

Leveraging unmanned aerial vehicle derived multispectral data for improved genomic prediction in potato (*Solanum tuberosum*)

Overview

Multispectral leaf canopy reflectance as measured by unmanned aerial vehicles (UAV) is the result of genetic and environmental interactions driving plant physio-chemical processes. These measures can then be used to construct relationship matrices for modelling genetic main effects. This type of phenotypic prediction is particularly relevant for trials with many entries, such as those used in early generation potato (*Solanum tuberosum*) breeding. We compared three methods for making predictions in our potato breeding program: first, using multispectral derived relationship matrices; second, using the traditional approach based on genomic derived relationships; and third using a combination of both. Multispectral bands were collected at five different time points for two market classes of potato, chipping and fresh market. We modeled genetic main effects for yield and quality traits at each time point and all stages combined. Models with multispectral relationship matrices exhibited better prediction accuracy for yield and roundness than genomic only models and models featuring spectra plus genomic kernels outperformed both single kernel predictions in terms of accuracy for most traits. Time points were variably informative depending on the trait measured, however for all traits combining across time points performed as well or better than single time point models. Similarly, using

feature selection to limit our models to important variables did not improve prediction accuracy significantly. This work highlights two potential uses for spectral data in genomic prediction, first, as an alternative to genetic data and second in combination with genetic data to increase precision of selection.

3.1 Introduction

Cultivated potato (*Solanum tuberosum* L.) is the most consumed non grain crop in the world (FAO, 2024). However, breeding progress, particularly for yield has been slow in comparison to other staple crops (Douches et al., 1996; Jansky, 2009). This is in part due to US cultivated potato's clonal nature and its highly heterozygous (Hardigan et al., 2017; Hoopes et al., 2022; Tuttle et al., 2024) autotetraploid genome. Historically potatoes have been bred using F1 crosses followed by approximately a decade of winnowing through selection to identify potential varieties (Jansky, 2009; Bradshaw, 2017). In many programs F1 seed is generated in green houses and then ramped up in small seed increase trials for two years before there is sufficient seed for a preliminary yield trial. When breeders rely on traditional phenotypic selection this results in five-to-seven-year breeding cycles.

Marker assisted selection can speed the breeding process (Slater et al., 2016; Ortiz, 2020). Genetic markers applied within the first or second field year offer greater efficiency over traditional visual screening methods (Bradshaw, 2017). While marker assisted selection (MAS) can be highly efficient, its effectiveness is generally limited to traits with simple genetic architecture and few large effect genes such as self-compatibility (Clot et al., 2020; Kaiser et al., 2021), maturity (Kloosterman et al., 2013; Gutaker et al., 2019), oblong vs. circular shape (Eck et al., 2022; Endelman et al., 2024), and disease resistance (Nie et al., 2016; Herrera et al., 2018; Prodhomme et al., 2019; Meade et al., 2020; Asano and Endelman, 2023).

For complex traits, genomic selection (GS) can be effective at increasing genetic gain and shortening breeding cycles (Slater et al., 2016; Habyarimana et al., 2017; Stich and Van Inghelandt, 2018; Endelman et al., 2018; Gemenet et al., 2020; Sood et al., 2020; Martins et al., 2023; Yusuf et al., 2024). GS estimates genome-wide marker effects for any given trait using

individuals that have been both genotyped and phenotyped within a training population (Meuwissen et al., 2001). However, even with decreasing sequencing costs, genotyping the tens of thousands of unique genetic individuals in early generations of potato breeding programs is expensive. Moreover, the successful implementation of GS depends on precise and accurate quantitative phenotypes (Yusuf et al., 2024), which remains a bottleneck, especially at the early testing stages of the breeding pipeline.

Advances in high-throughput phenotyping (HTP) are expected to alleviate some of these challenges, enhancing the efficiency and accuracy of GS (Slater et al., 2016; Hickey et al., 2019). HTP can enhance morphological trait scoring in plant breeding, accelerate the phenotyping processes, and increase precision (Shakoor et al., 2017; Hickey et al., 2019). Image-based phenotyping of tubers has been implemented to develop quantitative measures for quality traits, such as shape and color, in potatoes (Caraza-Harter and Endelman, 2020; Neilson et al., 2021; Miller et al., 2022; Feldman et al., 2024). The phenotypic data derived from these images allows for genomic prediction of quality traits (Yusuf et al., 2024).

Advances in sensor technology, machine learning, and data analytics are further enhancing the efficiency and accuracy of HTP in plant breeding (Hickey et al., 2019). Both ground-based and aerial HTP platforms have been deployed for phenotyping plants, either by directly capturing specific traits or indirectly examining phenotypic differences among plants using spectral reflectance data (Song et al., 2021). Aerial phenotyping offers the significant advantage of covering a wider expanse of field trials within a short timeframe. One such aerial platform is the unmanned aerial vehicle (UAV), which is equipped with multi-sensor cameras capable of capturing wavelengths near and beyond the visible spectrum. These wavelengths reflect different physiochemical properties related to plant structure, health, and nutrient status

45

based on canopy leaf absorption and reflectance (Alkhaled et al., 2023). Studies in several crops have reported significant improvements in prediction accuracy using HTP temporal data, both for direct predictions and within genomic selection models (Silva et al., 2022; Montesinos-López et al., 2023; Xu et al., 2023; Abdelhakim et al., 2024).

Despite its benefits, the use of temporal HTP data presents challenges due to its large volume and high dimensionality, which can complicate interpretation and increase computational demands when used directly for prediction. An efficient method for genomic selection involves using the genomic best linear unbiased prediction (GBLUP) model, where relationship matrices constructed from marker data among individuals are used as covariance in a GS model to estimate genomic estimated breeding values (GEBV) (VanRaden, 2008). This approach can be extended to incorporate multispectral data in a similar manner. Implementing a phenomic relationship matrix either as a predictor or in combination with genomic selection, improves predictions in multiple crops (Krause et al., 2019; Mróz et al., 2024; Maggiorelli et al., 2024).

We aimed to apply this combined phenomic and genomic selection method to our breeding program and extend it to predict quality traits as well as yield. Specifically, we address the following questions: (1) Can UAV-derived multispectral data predict yield, specific gravity and shape in chipping potatoes and yield, shape, and skin color in fresh market potatoes? (2) How do those predictions compare with genomic based predictions? (3) Can multispectral and genomic data be used in combination for prediction? (4) What time point is most appropriate for collecting multispectral data for phenomic prediction? (4) Which combination of spectral variables are most predictive?

3.2 Materials and Methods

3.2.1 Plant materials

We evaluated two sets of advanced potato breeding clones between 2022 and 2023 at the Sand Plains Research Farm in Becker, MN. The first set included 226 unique chipping clones while the second contained 87 unique fresh market yellow and red skinned clones. In addition to clones from the University of Minnesota (UMN) breeding program, the chips set included ten breeding clones from the University of Wisconsin Potato Breeding Program and five commonly grown varieties, namely Atlantic (Webb et al., 1978) Superior (Rieman, 1962), Lamoka (De Jong et al., 2017), Snowden, and Cascade. The fresh market clones included University of Minnesota (UMN) breeding program material and four commonly grown varieties, namely Red Norland (Johansen et al., 1959) Modoc, Chieftain (Weigle et al., 1968), Yukon Gold (Johnston and Rowberry, 1981). We evaluated these clones in an augmented block design with row/column information and repeated checks. All plots contained 15 seed pieces with 30 cm in row spacing and 91 cm between rows.

3.2.2 Phenotyping

We collected two years of phenotype data (Table 1). Yield, specific gravity, and roundness, were collected for the chips while yield, redness, lightness, and length to width ratio (L/W) were collected for the fresh market clones. Yield is reported as the total plot weight for all harvested tubers in Mg ha^{-1}. Yield was measured on a tuber basis and summed over plots using an AccuVision Dual View L.E.D. Grader System from Exeter. Specific gravity was determined by weighing a ten-tuber sample from each plot in air and then weighing it again when submerged in water. The specific gravity was then calculated by dividing the weight in air by the difference between the weight in air and the weight in water.

All other traits were measured using image analysis. Images of samples of ten harvested and washed tubers from each plot were taken using an Ortery Photosimilie 200 Lightbox with a Cannon Rebel T6i camera following the methods from Caraza-Harter and Endelman (2020). Trait measurement was performed using the potato image analysis R package, TubAR (Miller et al., 2022). Length to width ratio and roundness were calculated for each tuber and the median value was reported. Tuber skin color was quantified in terms of redness and lightness. The median tuber value for each image is used for analysis.

3.2.3 Multispectral data

Multispectral reflectance data were captured in the summers of 2022 and 2023 for five different time points using an unmanned aerial vehicle (UAV) with a Sentera 6x multispectral camera and 6x thermal sensors (Sentera, St Paul, MN). The camera is equipped with high radiometric calibration tools for correction of lighting variation and to provide a consistent reflectance surface spanning both visible and near-infrared spectral bands. Raw multispectral reflectance mean value was obtained for five separate color bands (red = 650 nm, green = 545 nm, blue = 445nm, NIR = 840 nm and RedEdge = 720 nm) per plot after image preprocessing and radiometric calibration from Sentera. Using the raw bands, eight different vegetation indices commonly used in potatoes (Alkhaled et al., 2023) were estimated. They include the Normalized Difference Vegetation Index (NDVI)(Rouse et al., 1974), Normalized Difference Red Edge (NDRE:)(Gitelson and Merzlyak, 1994a), Green Leaf Index (GLI), Chlorophyll Index Green (CIG), Green Normalized Difference Vegetation Index (GNDVI)(Gitelson and Merzlyak, 1994b), Normalized Difference Water Index (NDWI)(McFeeters, 1994), Chlorophyll Index Red Edge (CIRE)(Gitelson et al., 2012), and Transformed Chlorophyll Absorption in Reflectance Index/ Optimized Soil-Adjusted Vegetation Index (TCARI/OSAVI)(Haboudane et al., 2010).

Five flight dates were used each summer based on estimated average phenological growth stages translating to Canopy cover at 50%, Canopy cover closure, Peak Flowering, Peak flowering/ Biomass, Senescence (Table 1). The timing of phenological stage varies by genotype and market class, so an estimate was made for the whole field.

3.2.4 Genotyping data

Clones were genotyped with a combination of platforms as described in Yusuf et al. (2024). Specifically, both the SolCAP potato SNP array (Felcher et al., 2012; Vos et al., 2015) and flexseq gbs genotyping were used. A total of 181 chipping clones and 57 fresh market clones were genotyped using flexseq while 45 chippers and 30 fresh market clones were genotyped with the SNP array. The SNP array includes 21K markers while the less expensive flexseq platform includes a subset of 8k of those markers. The two genotypic datasets were merged, and imputation was used to fill in the gaps in the lower density array using the *L2H* function of the R package polyBreedR (Endelman et al., 2024). The merging and imputation of all the datasets resulted in a total of 21,531 markers which we used to carry out further analysis.

3.2.5 Statistical analysis and prediction models

3.2.5.1 Phenotypic data analysis

Within each year and market class, we accounted for spatial variation and computed best linear unbiased estimates (BLUEs) for all traits and for all multispectral data using a spatial 2D spline model, implemented with SpATS (Rodríguez-Álvarez et al., 2018), within the StageWise package (Endelman, 2023) in R ("R Core Team," 2023) based on row-column information computed as :

$$y_{ijk} = u + g_i + r_j + c_k + e_{ijk} \tag{1}$$

Where y_{ijk} is the response variable for the ith genotype in jth row and kth column, μ is the overall mean, g_i is the fixed effect for the genotypic response, r_j is the random row effect where $\mathbf{r_j} \sim N$ $(0, \sigma_r^2)$, c_k is the random column effect where $\mathbf{c_j} \sim N(0, \sigma_c^2)$, and e_{ijk} is the residual effect where $\mathbf{e_{ijk}} \sim N(0, \sigma_e^2)$.

Broad sense heritability (H^2) was computed for on a plot basis within each year and market class for all traits and for all multispectral data at each time point with genotype as random from equation (1) above. This was computed based on genetic σ_g^2 and residual σ_e^2 variance component as:

$$H_g^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2} \tag{2}$$

3.2.5.2 Relationship matrices

To explore the effect of different data types and data collection dates we generated a series of relationship matrices. Genomic relationship matrices (**G**) were first constructed as implemented in the Stagewise package (Endelman, 2023). A separate matrix was generated for each market class. Multispectral derived relationship matrices (**W**) were also computed across clone-year using the BLUEs for each spectra band and a combination of both spectra bands and indices for each of the time point individually and all time points combined. Once again, each market class was treated separately. The matrices were computed as:

$$W = \frac{ZZ^T}{n} \tag{3}$$

where n is the number of bands or bands plus indices included in the model Z is the scaled matrix that corresponds to the BLUEs of the multispectral band or indices and Z^T is the transpose of Z.

3.2.5.3 Prediction models

The estimated BLUEs for each trait, within each market class, were used as the response to fit both a genomic linear unbiased predictor model and multispectral linear unbiased predictor model using the BGLR (Pérez and de los Campos, 2014) package in R ("R Core Team," 2023). The genomic prediction model was used as a base model for comparison with the multispectral prediction models.

The genomic prediction model was fit with **G** as covariance:

$$BLUE[g_{ij}] = y_{ij} = E_j + g_i + s_{ij} \tag{4}$$

Where g_{ij} is the genotypic value for individual i in year j, E_j is the fixed effect for year j, g_i is the random effect for genotype i and s_{ij} is the residual effect with normal distribution with mean 0 and variance σ^2. The genetic effect $\boldsymbol{g} = (g_1, \dots, g_I)^T$ follows a normal distribution with $\mathbf{g} \sim N\,(\mathbf{0}, \mathbf{G}\sigma_g^2)$, where **G** is the genomic relationship matrix and σ_g^2 is the genomic variance. This model trained with equation (4) was called model **G**. The base model was further modified by incorporating the effect of genotype by environment (G×E) and model (4) was modified as:

$$BLUE[g_{ij}] = y_{ij} = E_j + g_i + gE_{ij} + s_{ij} \tag{5}$$

Where E_j, g_i and s_{ij} are defined in model (4). $gE_{ij,}$ denotes the G×E interaction term with a multivariate normal distribution $\mathbf{gE} = (gE_{11}, \dots, gE_{1I}, \dots, gE_{IJ})^T \sim N\,(\mathbf{0}, (\mathbf{Z}_g\mathbf{G}\mathbf{Z}_g^T)^{\circ}(\mathbf{Z}_E\mathbf{Z}_E^T)\, \sigma_{gE}^2)$, where **G** is the genomic relationship matrix, Z_g and Z_E are the incidence matrix for genomic and year effects, ° denotes the Hadamard product and σ_{gE}^2 is the variance component of the genotype by environment interaction. The model trained with equation 5, was denoted as model **G+GE**.

The model incorporating multispectral data was fit with **W** as covariance in a similar approach with genomic prediction:

$$BLUE[g_{ij}] = y_{ij} = E_j + wE_{ij} + s_{ij} \tag{6}$$

51

Where E_j and s_{ij} are defined in model (4). wE_{ij} is the random effect of the multispectral band and or indices of genotype i in unique year j, which follows a joint normal distribution with $\mathbf{wE} \sim N$ $(\mathbf{0}, \mathbf{W}\sigma^2_{Ew})$, where $\mathbf{W}$ is the multispectral relationship matrix for clone-year combination and σ^2_{Ew} is the multispectral variance component for the clone-year combination. Model in equation (6) was denoted as model $\mathbf{W}$.

Model given in equation (6) was further improved taking into account the effect of genotypes (g_i) as proposed by Lopez-Cruz et al. (2015) and Krause et al. (2019):

$$BLUE[g_{ij}] = y_{ij} = E_j + g_i + wE_{ij} + s_{ij} \qquad (7)$$

Where all the components of model (7) were described in the previous models and this model was denoted as model $\mathbf{G}+\mathbf{WE.}$

3.2.6 Assessment of prediction models

We evaluated the prediction performance under two types of cross validation of interest to potato breeding programs. First, we evaluated the prediction accuracy of tested lines in tested environments. Second, we evaluated the prediction performance of tested lines in untested environments (Alemu et al., 2024). For the first evaluation, we assessed model performance using prediction ability across years randomly based on a 10 folds-cross validation scheme repeated 10 times by splitting data into random train-test partitions. Each fold was partitioned with 20% of data for testing while the remaining 80% was for training. The prediction ability was evaluated as the Pearson´s correlation between the predicted value and the true value for each phenotypic trait. For the second type of cross validation, we partitioned clones into years and used the year 2023 as training set and year 2022 as testing set. Once again, the prediction

ability was measured as the Pearson´s correlation between the predicted values and the true values for each phenotypic trait.

3.2.7 Timepoint significance for prediction

To identify the time point most significant for prediction and to recommend specific time points for prediction we compared the prediction models based on multispectral matrices (**W or G +WE**) developed from the different time points and identified significant time points for prediction using the analysis of variance (ANOVA). Mean separation was carried out using least significant difference (LSD). The model was fitted as:

$$Y_{ijk}=\mu+\alpha_i+ \beta_j+\alpha\beta_{ij}+\epsilon_{ijk} \tag{8}$$

Where Y_{ijk} is the response variable corresponding to the k-th mean prediction ability of the i-th multispectral model and j-th time point. μ is the overall mean, α_i is the effect for the multispectral model, β_j is the effect for the time point, $\alpha\beta_{ij}$ is the interaction effect between the models and the time points and ϵ_{ijk} is the residual, that is assumed normally distributed with mean zero and variance σ^2.

3.2.8 Features selection and significance for prediction

Intercorrelated variables can lead to multicollinearity in regression equations resulting in underperformance of prediction models (Yu et al., 2015). Such scenarios are common for multispectral data especially when multiple bands and indices are combined across timepoints (Galán et al., 2020). We therefore carried out supervised feature selection with the combined stage multispectral data for yield of chipping potatoes using the Lasso and Elastic-Net Regularized Generalized Linear Models (GLMNET) machine learning algorithm, to identify the most important variables contributing to prediction (Friedman et al., 2010).

In addition to the internal feature selection in the GLMNET algorithm, we used two other known feature selection wrappers within the caret R package, selection by filtering (SBF) and simulation annealing (SA). Each of the features selected by the algorithm and wrappers were used to fit a new model. The set of selected features which produced the best prediction accuracies were used to fit a genomic selection model. These new models were compared to models created based on the complete set of variables without feature selection in terms of prediction ability.

3.3 Results

3.3.1 Phenotypic data analysis

The BLUEs for each trait in both market classes from q-q plots assessment were approximately normally distributed (Figure 1). We observed similar distributions for the multispectral data. The BLUEs for the traits and the multispectral data were utilized for further analysis, with the trait BLUEs assigned as response variables. For the chips, broad sense heritability for yield was between 51% and 53%, specific gravity ranged between 64% and 66%, and roundness ranged between 27% and 77% (Figure 2). Within the fresh market class clones, broad sense heritability for yield, ranged from 49% to 80%, redness ranged between 58% and 78%, lightness was between 79% and 84%, roundness was between 69% and 81%, and length-width ratio was between 48% and 65%.

Multispectral data showed low to moderate heritability across market clones and years. Heritability for bands and indices across time points within the chips clones ranged from 8% to 72% while the fresh market clones ranged between 0% to 84% (Figure S2). Heritabilities were generally lower for bands and indices from the fresh market clones in 2022 (Figure S1), with

54

blue and red bands in the earlier time points (stage 1 and stage 2) exhibiting the lowest (<1%) heritability. Blue and red bands also exhibited low heritability in later time points (stage 4 and stage 5) in 2023.

3.3.2 Multi-year genomic and multispectral prediction model

We compared all the models described above for the chips. Models featuring only multispectral bands (Figure 3) performed indistinguishably from those based on both bands and indices (Figure S3). The model combining genomic and multispectral data, **G**+**WE**, consistently outperformed the other models except for yield at stage 3 and stage 5. Both the models containing multispectral data, **W** and **G**+**WE**, outperformed the genomic only models, **G** and **G**+**GE**, for yield prediction at all time points and was highest at the combined stages. The mean prediction abilities at the combined stages for **G** and **G**+**GE** were both 0.43 while **W** and **G**+**WE** model were 0.72 and 0.74 respectively. In early time points, the models with genetic components, **G**, **G**+**GE**, and **G**+**WE**, performed better than the **W** only model for specific gravity. However, at stage 5 (senescence) and for the combined stage, model **G**, **W**, and **G**+**GE** performed similarly, although all were inferior to **G**+**WE** with a mean prediction ability of 0.36 at stage 5 and 0.39 for the combined stages. Prediction ability was generally lower for roundness. Similar to specific gravity, genetic models for roundness prediction performed better earlier in the growing season while multispectral prediction improved as the season went on, but the combination of the two was the most successful with a mean prediction ability of 0.22 at stage 5 and 0.21 for the combined stages.

prediction ability in terms of Pearson´s correlation of model predicted values to phenotypic observations.

We repeated the above analysis for the fresh market clones, to investigate the efficacy of models featuring multispectral data for prediction of quality traits. As with the chips, we found that including only bands (Figure 4) as opposed to combining bands and indices (Figure S4) did not significantly affect prediction accuracy. For yield and color traits, the **G**+**WE** model performed the best. As in the chips, yield was best predicted by the two models including multispectral data. For the combined stages the **G**+**WE** model had a prediction ability of 0.63 while that of the **W** model was 0.61. For both, redness and lightness models including the genomic relationship matrix performed best, although the difference was more pronounced for lightness across all stages. For both traits, the combined model (**G**+**WE**) with combined stages performed best, with a prediction ability of 0.43 for redness and 0.32 for lightness. For roundness, **W** was the strongest model with other models performing similarly, while for length to width ratio all models including the genomic relationship matrix performed similarly, while **W** was the worst. Overall, prediction abilities were lowest for length to width ratio.

3.3.3 Year specific genomic and multispectral prediction model

In order to evaluate our ability to predict phenotypes in untested environments, we used years as a proxy for environment. We divided our data based on the year in which the data was gathered and assigned our largest data set (2023) as the training set and our smallest (2022) as the test set. This analysis was only carried out in the chipping potatoes due to population size. Prediction ability was significant ($p<0.05$) for yield and specific gravity but not roundness (Figure 5; Figure S5). These results were consistent across time points and for all time points combined. For all traits and models, prediction ability was lower in untested environments than tested ones.

3.3.4 Significance of time points for prediction

We compared the prediction models based on multispectral matrices (**W** or **G +WE**) from the different stages and identified significant time points for prediction of yield and quality traits at harvest using analysis of variance (ANOVA) and mean separation using *LSD*. Time point was significant for all traits in both market classes. The time points which produced the best predictions varied by trait, however the combined stage model was among the best performers for all traits in all both market classes (Figure 6). Within the chips, combined analysis of all time points was the best predictor of yield. However, the earliest stage predicted yield equally well for the fresh market clones. Combined analysis was equivalent to analysis using either of the last two time points for specific gravity. Time points 4 and 5 were equally effective to combined analysis for predicting roundness in chipping clones, while stage 3 (peak flowering) was the best for roundness in fresh market clones. The last time point was similarly effective to combined analysis for lightness, and both the first and last time points were equivalent to combined analysis for predicting redness.

3.3.5 Variable importance and significance for prediction

The models using combined multispectral data across time points showed improved performance for most traits evaluated. Considering the risk of multicollinearity in multistage models (Yu et al., 2015), we evaluated the importance and contribution of each variable to our yield prediction models for the chipping germplasm using a machine learning variable importance selection algorithm and wrappers. Specifically, we used the GLMNET algorithm which provides a robust approach to coefficient shrinkage and regularization of important variables (Friedman et al., 2010).

The GLMNET internal features selection algorithm and wrappers (Table 2) selected different numbers of important multispectral variables contributing to prediction models. Performance for prediction based on the selected variables shows the wrappers performed better than the internal GLMNET features selection (Table 2). The variables selected from the 66 total variables were 47 for GLMNET, 26 for simulation annealing (SA), and 46 with selection by filtering (SF). Most of the selected multispectral variables overlapped across the two wrappers (Table S1).

Selected variables included multiple bands and indices from different time points. We compared a model including only the selected multispectral variables (**W**) and one combining genomic and selected multispectral variables (**G**+**WE**) to the genomic models (**G** and **G**+**GE**). The mean prediction ability from selected variables with SA wrapper was 0.73 and 0.75 with SF wrapper (Figure 6). These results showed no significant improvement compared with the use of the total unselected variables with 0.74 (Figure 6).

3.4 Discussion

3.4.1 Heritability of traits and spectra data

Yield and quality traits exhibited moderate heritabilities consistent with previous estimates (Yusuf et al., 2024; Miller et al., 2022), indicating a strong genetic basis for improvement. We observed variability between years for roundness in chips and yield, length to width ratio, and redness in fresh market clones. Shape is known to have a strong environmental component (Jones et al., 2021; Stefaniak et al., 2021; Yusuf et al., 2024) as does yield (Schmitz Carley et al., 2019; Agha et al., 2024; Yusuf et al., 2024) which could explain the year-to-year variation. Heritability of the multispectral bands and indices was more variable, ranging from

low to moderate across different market classes and years. One reason for this may be that cultivated potato exhibits variation for flowering time and maturity and so single flight time points did not catch all individuals at the same growth stage. Low heritability of multispectral variables has also been observed in wheat (Krause et al., 2019; Mróz et al., 2024).

3.4.2 How does phenomic prediction compare to genomic prediction?

The most dramatic benefit of implementing phenomic prediction to our prediction abilities was for yield in both market classes. This is consistent across time points and with previous studies in potato (Maggiorelli et al., 2024), wheat (Montesinos-López et al., 2017; Krause et al., 2019; Mróz et al., 2024), rye (Galán et al., 2020), corn (DeSalvio et al.,2024), and soybean (Zhu et al., 2021).

We further assessed the predictive ability of the multispectral model on traits other than yield. Phenomic models also outperformed genomic models for predicting roundness in later time points but were less successful in predicting skin color, length to width ratio, or specific gravity. It is possible that better targeting of relevant spectral bands and indices would improve our ability to use phenomic prediction for quality traits. Physiochemical traits of canopy leaves, such as starch content are known to be associated with specific wavelengths of leaf absorption (Zhou et al., 2018; Yang et al., 2022; Alkhaled et al., 2023). If leaf and tuber starch content are correlated this suggests a potential strategy for the prediction of specific gravity. In a variety of crops imaging spectroscopy during the growing season, particularly from the NIR spectrum, accurately predicts harvest traits (Su and Xue, 2021; Dallinger et al., 2023).

Trait prediction from spectral imagery could be a particularly powerful tool in potato breeding. In general, potato breeding starts with a set of F1 crosses which are evaluated visually in the first field year from single plants. Breeding programs plant tens of thousands of

individuals and select approximately 1%. Genotyping at this time point is prohibitively expensive, but phenotyping with drones would not be. For most programs sufficient seed for yield trials is not generated until field year 3. If yield could be predicted from drone imagery in year 1 it could dramatically improve early generation selection accuracy and shorten breeding cycles. However, this depends on the accuracy of year-to-year predictions from single plants in new year-environments, and phenomic prediction across environments is generally less successful than it is within environment (Montesinos-López et al., 2017).

3.4.3 Is combining genomic and phenomic prediction effective?

Genomic prediction often relies on incorporating more explanatory inputs to optimize and enhance prediction accuracy. Several studies have explored the integration of additional inputs into genomic selection (GS), such as environmental covariates (Lopez-Cruz et al., 2015; Jarquín et al., 2017), omics data (Ye et al., 2020; Perez et al., 2022) and phenomics (Montesinos-López et al., 2017; Krause et al., 2019; Mróz et al., 2024; Galán et al., 2020; Zhu et al., 2021), including spectral information, to achieve better optimization and prediction accuracy.

In this study, the integration of multispectral information through additional modeling of genotype-by-environment interactions (G×E) using spectral reflectance significantly improved genomic prediction for most of the traits considered, except for the length-to-width ratio, which experienced a reduction of 7%. The inclusion of spectral data enhanced prediction accuracy for yield by over 30%, while improvements for other traits ranged from 3% to 11% across different market classes and time points. These findings suggest that multispectral information can capture additional effects not accounted for by marker effects alone, which is particularly important in clonal crops like potatoes.

Our findings are in strong agreement with previous studies that have utilized spectral information to improve prediction accuracy in GS, particularly for yield, using relationship matrices (Krause et al., 2019; Galán et al., 2020; Mróz et al., 2024). These results also align with studies employing other linear and nonlinear modeling approaches, such as partial least squares, ordinary least squares, and functional regression, to predict yield using spectral information (Montesinos-López et al., 2017, 2023).

3.4.4 What drone flight time points are most informative?

In order to test if fewer drone flights would be sufficient, we evaluated the performance of phenomic prediction models at each growth stage. The results from the analysis of variance highlighted the significance and effectiveness of certain time points for prediction. However, for all traits the combined stage performed as well or better than individual time point analysis. The time points that were most predictive varied by trait and market class. This suggests that continuing to measure at all five time points is important to trait prediction, and it's possible adding additional time points would improve our models.

It is important to note that maturity is both highly variable and highly heritable in cultivated potato (Kloosterman et al, 2013; Guraker et al., 2019). At any particular flight time different genotypes were likely at slightly different growth stages. Therefore, even if above ground reflectance at a particular growth stage was the strongest predictor for a harvest phenotype it might not be captured in this analysis, further supporting the need to measure multiple time points for accurate prediction.

3.4.5 Can we improve prediction accuracy by selecting key variables?

Including too many non-predictive variables in a prediction model can reduce power and result in multicollinearity leading to underperformance of the model (Yu et al., 2015). This is a known problem in multispectral data especially data combined across time points (Galán et al., 2020), as well as in genomic selection where including only trait associated markers can quadruple prediction ability (Singh et al., 2023). Combining genomic and phenomic data across all time points exhibited the best predictive performance for yield; we tested if that performance could be improved by identifying the most crucial spectral variables and including only those variables in the model. Although we used multiple machine learning algorithms to define sets of predictive spectra, all subsets performed equally well to the total set. We have yet to reach the point of saturating the model and could potentially improve it with more spectral data.

3.5 Conclusion

In this study, we have established the value of using multispectral data for both phenomic and genomic selection in potatoes. By utilizing a multispectral relationship matrix, we can achieve prediction accuracy for agronomic and quality traits comparable to genomic prediction, offering a viable alternative when genetic markers are unavailable. Additionally, we have shown that integrating multispectral data into genomic selection can further enhance prediction accuracy. We found no evidence that we were overfitting the model with too many non-predictive variables, as our most accurate predictions resulted from combining data across time points and feature selection made no difference.

It's important to note that this work serves as a proof of concept for using multispectral data derived from unmanned aerial vehicles. This approach was implemented with advanced

breeding clones, where full-sib families are limited, and there is minimal Mendelian sampling within the population. Moving forward, further research will focus on applying this concept to the screening of early-generation clonal selections, where large number of entries require evaluation.

Table 1. Timing of drone flights and number of clones evaluated each year

Class	Evaluation Year	Total unique entries	Common clones	Estimated average growth stages/ Drone flight Dates	Days after planting
Chips market	2022	42	31	Canopy cover 50%: June 28	36
				Canopy cover flower: July 05	45
				Peak flowering: July 12	52
				Vine biomass: July 19	59
				Senescence: August 05	76
	2023	215		Canopy cover 50%: June 20	30
				Canopy cover flower: July 05	45
				Peak flowering: July 13	53
				Vine biomass: July 18	58
				Senescence: August 08	79
Fresh market	2022	25	18	Canopy cover 50%: June 28	36
				Canopy cover flower: July 05	45
				Peak flowering: July 12	52
				Vine biomass: July 19	59
				Senescence: August 05	76
	2023	77		Canopy cover 50%: June 20	30
				Canopy cover flower: July 05	45
				Peak flowering: July 13	53
				Vine biomass: July 18	58
				Senescence: August 08	79

Figure 1. Trait BLUEs distribution for A. Chipping clones and B. Fresh market clones.
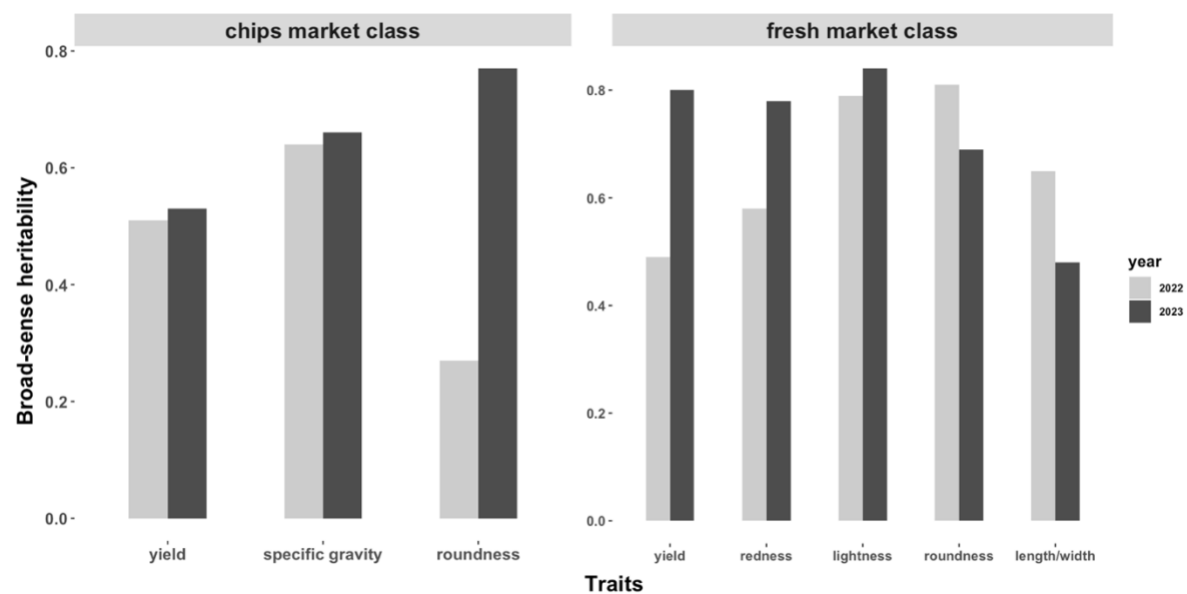
Figure 2. Broad-sense heritability for both chipping clones and fresh market clones
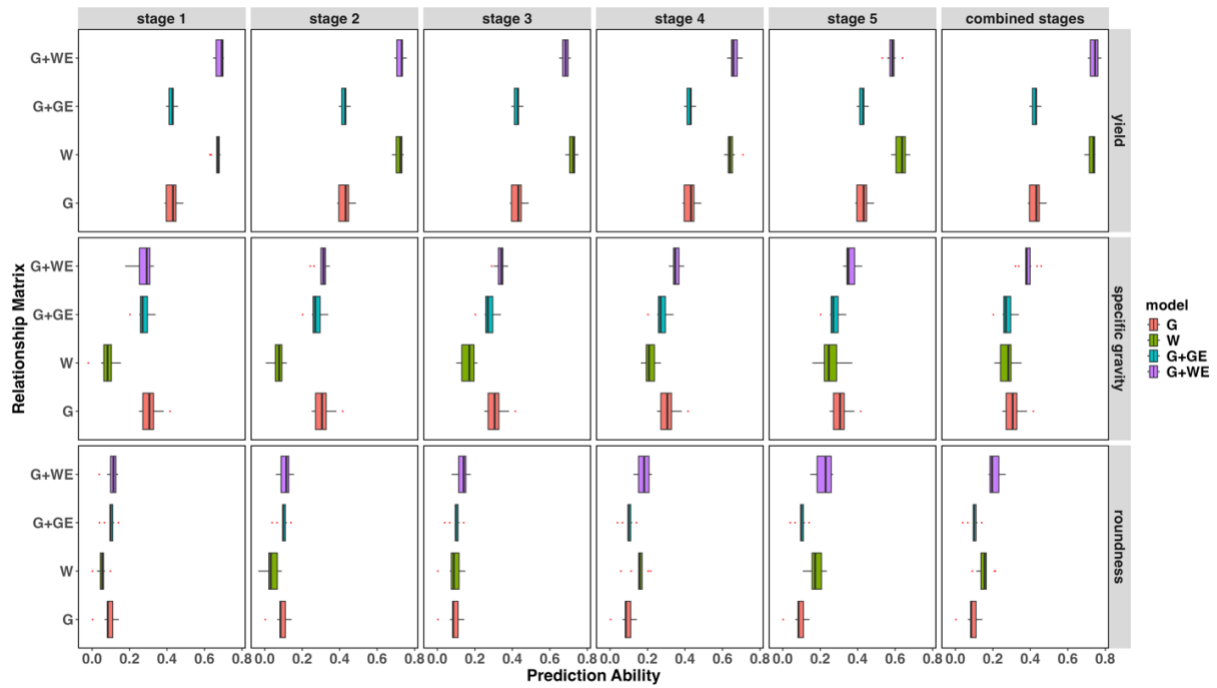
Figure 3. Comparison of models including genomic (**G**), genome by environment (**GE**), and/or multispectral (**W**) relationship matrices for chip traits across time points. Comparison is based on prediction ability in terms of Pearson´s correlation of model predicted values to phenotypic observations.
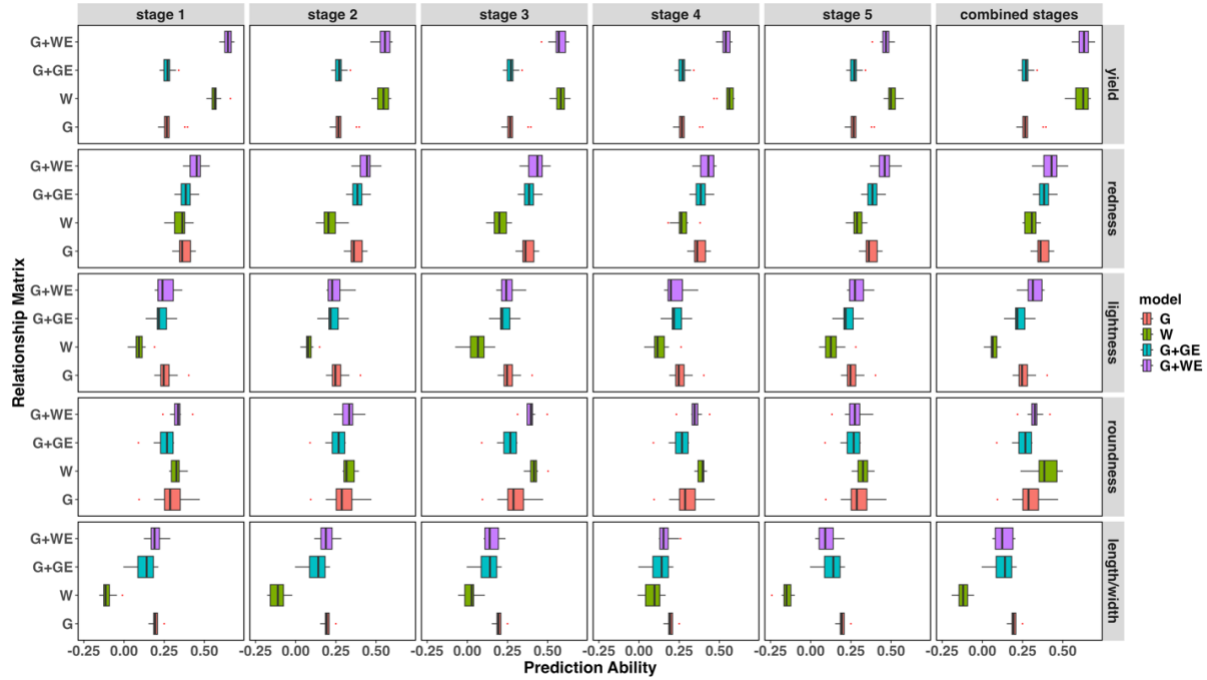
Figure 4. Comparison of models including genomic (**G**), genome by environment (**GE**), and/or multispectral (**W**) relationship matrices for traits measured in fresh market clones across time points. Comparison is based on prediction ability in terms of Pearson´s correlation of model predicted values to phenotypic observations.
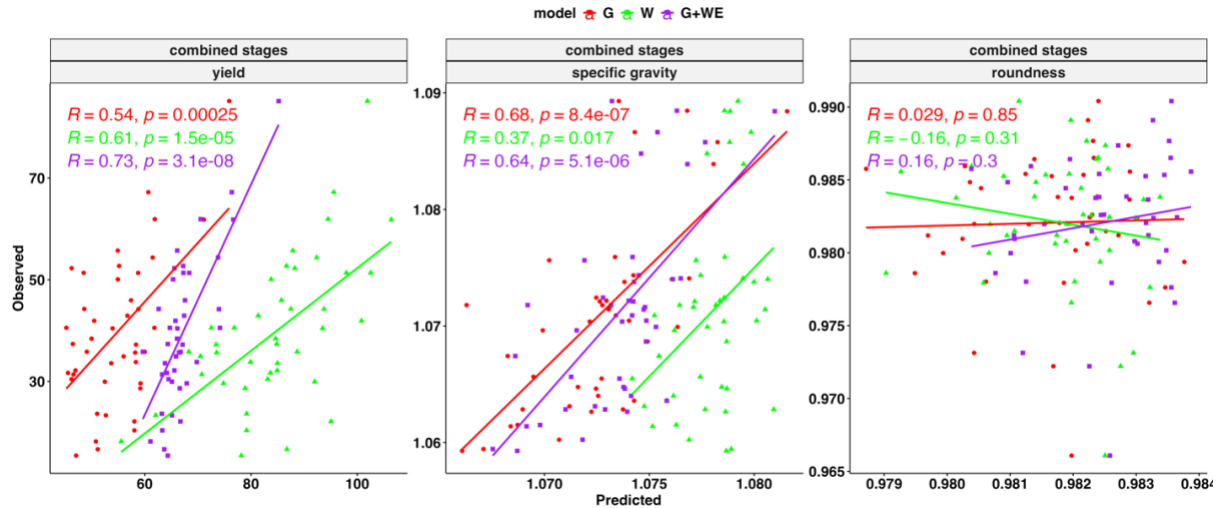
Figure 5. Comparison of year specific prediction ability for models including genome (**G**), genome by environment (**GE**), and/or multispectral (**W**) data combined across time points for chipping clones. Prediction ability is the Pearson's correlation of the model prediction to the observed phenotypic value.
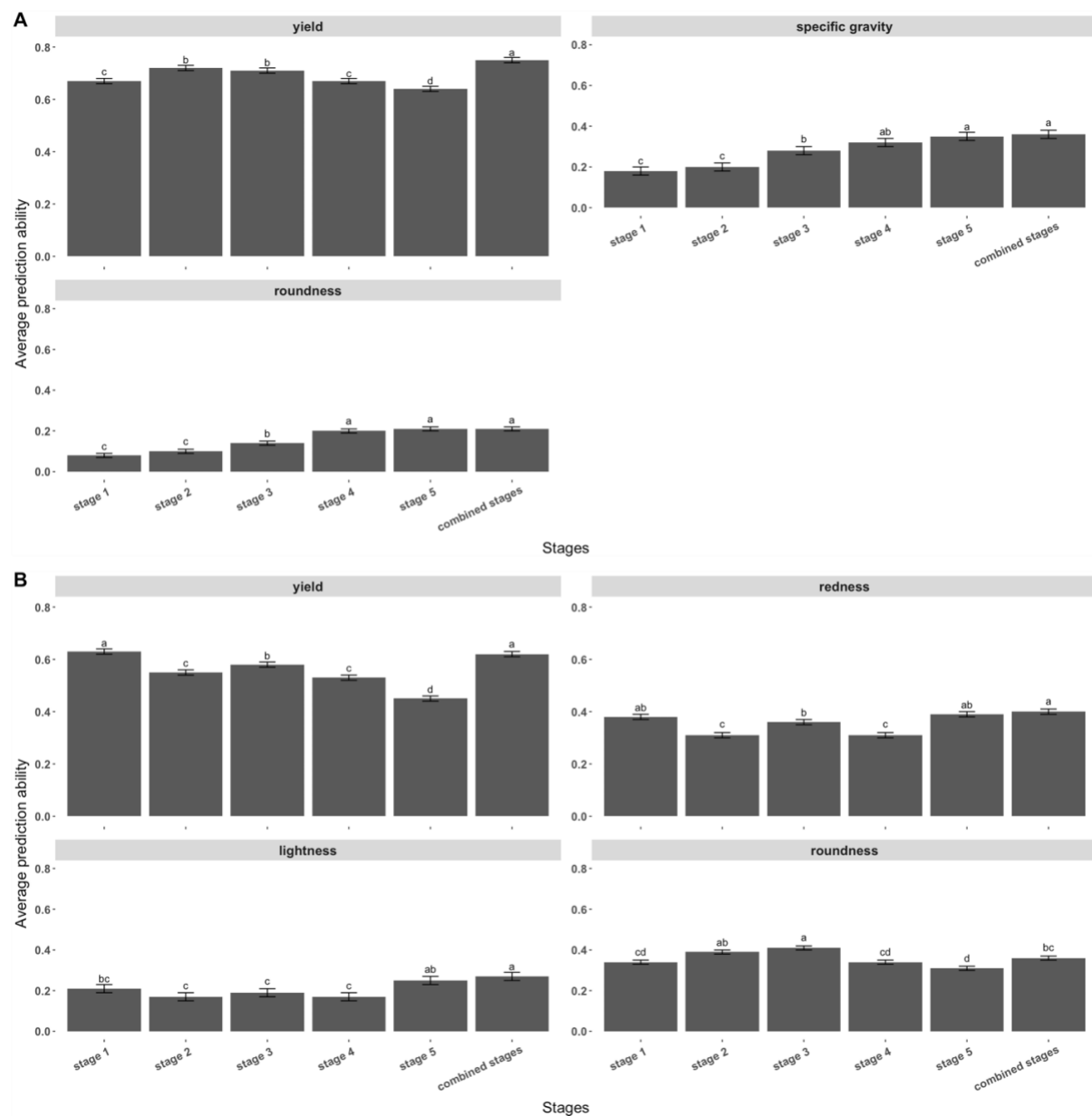
Figure 7. The influence of time points of spectral data collection on prediction ability for various traits in (A) chipping potatoes and (B) fresh market clones.

Table 2. GLMNET Prediction accuracy in terms of Pearsons correlation using selected multispectral variables

Population set	GLMNET (R)	GLMNET_SA (R)	GLMNET_SF (R)
Training	0.70	0.75	0.76
Testing	0.71	0.72	0.74
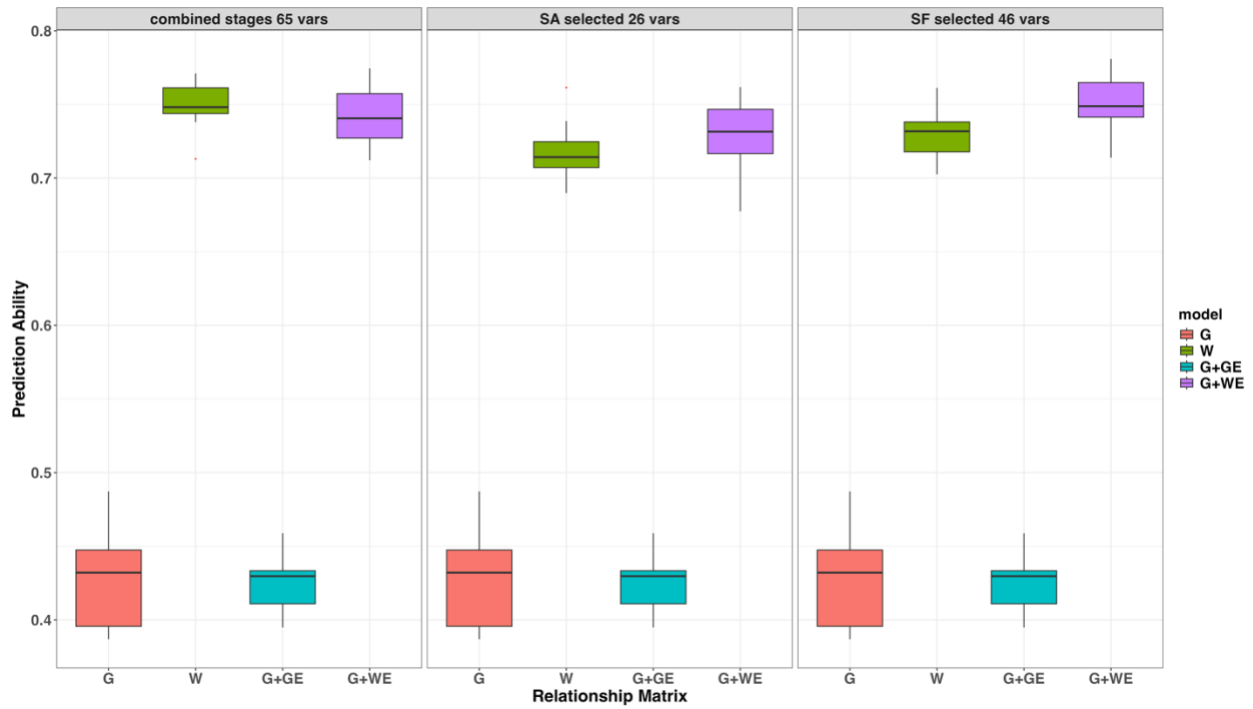
Figure 7. Prediction ability for chipping clone traits across time points using subsets of the multispectral variables.

References

Abdelhakim, L.O.A., B. Pleskačová, N.Y. Rodriguez-Granados, R. Sasidharan, L.S. Perez-Borroto, et al. 2024. High Throughput Image-Based Phenotyping for Determining Morphological and Physiological Responses to Single and Combined Stresses in Potato. J Vis Exp (208). doi: 10.3791/66255.

Agha, H.I., J.B. Endelman, J. Chitwood-Brown, M. Clough, J. Coombs, et al. 2024. Genotype-by-environment interactions and local adaptation shape selection in the US National Chip Processing Trial. Theor Appl Genet 137(5): 99. doi: 10.1007/s00122-024-04610-3.

Alemu A, Åstrand J, Montesinos-López OA, Isidro Y Sánchez J, Fernández-Gónzalez J, Tadesse W, Vetukuri RR, Carlsson AS, Ceplitis A, Crossa J, Ortiz R, Chawade A. (2024). Genomic selection in plant breeding: Key factors shaping two decades of progress. Mol Plant, 17(4):552-578. doi: 10.1016/j.molp.2024.03.007.

Alkhaled, A., P.A. Townsend, and Y. Wang. 2023. Remote Sensing for Monitoring Potato Nitrogen Status. Am. J. Potato Res. 100(1): 1–14. doi: 10.1007/s12230-022-09898-9.

Asano, K., and J.B. Endelman. 2023. Development of KASP markers for the potato virus Y resistance gene Rychc using whole-genome resequencing data. : 2023.12.20.572658. doi: 10.1101/2023.12.20.572658.

Bradshaw, J.E. 2017. Review and Analysis of Limitations in Ways to Improve Conventional Potato Breeding. Potato Res. 60(2): 171–193. doi: 10.1007/s11540-017-9346-z.

Caraza-Harter, M.V., and J.B. Endelman. 2020. Image-based phenotyping and genetic analysis of potato skin set and color. Crop Science 60(1): 202–210. doi: 10.1002/csc2.20093.

Clot, C.R., C. Polzer, C. Prodhomme, C. Schuit, C.J.M. Engelen, et al. 2020. The origin and widespread occurrence of Sli-based self-compatibility in potato. Theor Appl Genet. doi: 10.1007/s00122-020-03627-8.

Dallinger, H.G., F. Löschenberger, H. Bistrich, C. Ametz, H. Hetzendorfer, et al. 2023. Predictor bias in genomic and phenomic selection. Theor Appl Genet 136(11): 235. doi: 10.1007/s00122-023-04479-8.

De Jong, W.S., D.E. Halseth, R.L. Plaisted, X. Wang, K.L. Perry, et al. 2017. Lamoka, a Variety with Excellent Chip Color Out of Cold Storage and Resistance to the Golden Cyst Nematode. Am. J. Potato Res. 94(2): 148–152. doi: 10.1007/s12230-016-9557-x.

DeSalvio, A.J., A. Adak, S.C. Murray, D. Jarquín, N.D. Winans, et al. 2024. Near-infrared reflectance spectroscopy phenomic prediction can perform similarly to genomic prediction of maize agronomic traits across environments. The Plant Genome n/a(n/a): e20454. doi: 10.1002/tpg2.20454.

Douches, D.S., D. Maas, K. Jastrzebski, and R.W. Chase. 1996. Assessment of Potato Breeding Progress in the USA over the Last Century. Crop Science 36(6): cropsci1996.0011183X003600060024x. doi: 10.2135/cropsci1996.0011183X003600060024x.

Eck, H.J. van, M.E.P. Oortwijn, I.R. Terpstra, N.H.M. van Lieshout, E. van der Knaap, et al. 2022. Engineering of tuber shape in potato (Solanum tuberosum) with marker assisted breeding or genetic modification using StOFP20. doi: 10.21203/rs.3.rs-1807189/v1.

Endelman, J.B. 2023. Fully efficient, two-stage analysis of multi-environment trials with directional dominance and multi-trait genomic selection. Theor Appl Genet 136(4): 65. doi: 10.1007/s00122-023-04298-x.

Endelman, J.B., C.A.S. Carley, P.C. Bethke, J.J. Coombs, M.E. Clough, et al. 2018. Genetic Variance Partitioning and Genome-Wide Prediction with Allele Dosage Information in Autotetraploid Potato. Genetics 209(1): 77–87. doi: 10.1534/genetics.118.300685.

Endelman, J.B., M. Kante, H. Lindqvist-Kreuze, A. Kilian, L.M. Shannon, et al. 2024. Targeted genotyping-by-sequencing of potato and software for imputation. : 2024.02.12.579978. doi: 10.1101/2024.02.12.579978.

FAO. Crop and Livestock products. License: CC BY-NC-SA 3.0 IGO. Extracted from: https://www.fao.org/faostat/en/#data/QCL. Data of Access: 01-10-24

Felcher, K.J., J.J. Coombs, A.N. Massa, C.N. Hansey, J.P. Hamilton, et al. 2012. Integration of Two Diploid Potato Linkage Maps with the Potato Genome Sequence. PLOS ONE 7(4): e36347. doi: 10.1371/journal.pone.0036347.

Feldman, M.J., J. Park, N. Miller, C. Wakholi, K. Greene, et al. 2024. A scalable, low-cost phenotyping strategy to assess tuber size, shape, and the colorimetric features of tuber skin and flesh in potato breeding populations. The Plant Phenome Journal 7(1): e20099. doi: 10.1002/ppj2.20099.

Friedman, J., T. Hastie, and R. Tibshirani. 2010. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw 33(1): 1–22.

Galán, R.J., A.-M. Bernal-Vasquez, C. Jebsen, H.-P. Piepho, P. Thorwarth, et al. 2020. Integration of genotypic, hyperspectral, and phenotypic data to improve biomass yield prediction in hybrid rye. Theor Appl Genet 133(11): 3001–3015. doi: 10.1007/s00122-020-03651-8.

Gemenet, D.C., H. Lindqvist-Kreuze, B. De Boeck, G. da Silva Pereira, M. Mollinari, et al. 2020. Sequencing depth and genotype quality: accuracy and breeding operation considerations for genomic selection applications in autopolyploid crops. Theor Appl Genet 133(12): 3345–3363. doi: 10.1007/s00122-020-03673-2.

Gutaker, R.M., C.L. Weiß, D. Ellis, N.L. Anglin, S. Knapp, et al. 2019. The origins and adaptation of European potatoes reconstructed from historical genomes. Nat Ecol Evol 3(7): 1093–1101. doi: 10.1038/s41559-019-0921-3.

Habyarimana, E., B. Parisi, and G. Mandolino. 2017. Genomic prediction for yields, processing and nutritional quality traits in cultivated potato (Solanum tuberosum L.). Plant Breeding 136(2): 245–252. doi: 10.1111/pbr.12461.

Hardigan, M.A., F.P.E. Laimbeer, L. Newton, E. Crisovan, J.P. Hamilton, 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 U S A 114(46): E9999–E10008. doi: 10.1073/pnas.1714380114.

Herrera, M. del R., L.J. Vidalon, J.D. Montenegro, C. Riccio, F. Guzman, et al. 2018. Molecular and genetic characterization of the Ryadg locus on chromosome XI from Andigena potatoes conferring extreme resistance to potato virus Y. Theor Appl Genet 131(9): 1925–1938. doi: 10.1007/s00122-018-3123-5.

Hickey, L.T., A. N. Hafeez, H. Robinson, S.A. Jackson, S.C.M. Leal-Bertioli, et al. 2019. Breeding crops to feed 10 billion. Nature Biotechnology 37(7): 744–754. doi: 10.1038/s41587-019-0152-9.

Hoopes, G., X. Meng, J.P. Hamilton, S.R. Achakkagari, F. de Alves Freitas Guesdes, et al. 2022. Phased, chromosome-scale genome assemblies of tetraploid potato reveal a complex genome, transcriptome, and predicted proteome landscape underpinning genetic diversity. Molecular Plant 15(3): 520–536. doi: 10.1016/j.molp.2022.01.003.

Jansky, S. 2009. Chapter 2 - Breeding, Genetics, and Cultivar Development. In: Singh, J. and Kaur, L., editors, Advances in Potato Chemistry and Technology. Academic Press, San Diego. p. 27–62

Jarquín, D., C. Lemes da Silva, R.C. Gaynor, J. Poland, A. Fritz, et al. 2017. Increasing Genomic-Enabled Prediction Accuracy by Modeling Genotype × Environment Interactions in Kansas Wheat. The Plant Genome 10(2): plantgenome2016.12.0130. doi: 10.3835/plantgenome2016.12.0130.

Johansen, R.H., N. Sandar, W.G. Hoyman, and E.P. Lana. 1959. Norland a new red-skinned potato variety with early maturity and moderate resistance to common scab. American Potato Journal 36(1): 12–15. doi: 10.1007/BF02877209.

Johnston, G.R., and R.G. Rowberry. 1981. Yukon Gold: A new yellow-fleshed, medium-early, high quality table and French-fry cultivar. American Potato Journal 58(5): 241–244. doi: 10.1007/BF02853905.

Jones, C.R., T.E. Michaels, C. Schmitz Carley, C.J. Rosen, and L.M. Shannon. 2021. Nitrogen uptake and utilization in advanced fresh-market red potato breeding lines. Crop Science 61(2): 878–895. doi: 10.1002/csc2.20297.

Kaiser, N.R., S. Jansky, J.J. Coombs, P. Collins, M. Alsahlany, et al. 2021. Assessing the Contribution of Sli to Self-Compatibility in North American Diploid Potato Germplasm Using KASP™ Markers. Am. J. Potato Res. doi: 10.1007/s12230-021-09821-8.

Kloosterman, B., J.A. Abelenda, M. del M.C. Gomez, M. Oortwijn, J.M. de Boer, et al. 2013. Naturally occurring allele diversity allows potato cultivation in northern latitudes. Nature 495(7440): 246–250. doi: 10.1038/nature11912.

Krause, M.R., L. González-Pérez, J. Crossa, P. Pérez-Rodríguez, O. Montesinos-López, et al. 2019. Hyperspectral Reflectance-Derived Relationship Matrices for Genomic Prediction of Grain Yield in Wheat. G3 Genes|Genomes|Genetics 9(4): 1231–1247. doi: 10.1534/g3.118.200856.

Lopez-Cruz, M., J. Crossa, D. Bonnett, S. Dreisigacker, J. Poland, et al. 2015. Increased Prediction Accuracy in Wheat Breeding Trials Using a Marker × Environment Interaction Genomic Selection Model. G3 Genes|Genomes|Genetics 5(4): 569–582. doi: 10.1534/g3.114.016097.

Maggiorelli, A., N. Baig, V. Prigge, J. Bruckmüller, and B. Stich. 2024. Using drone-retrieved multispectral data for phenomic selection in potato breeding. Theor Appl Genet 137(3): 70. doi: 10.1007/s00122-024-04567-3.

Martins, V.S., M.H.M.L. Andrade, L.N. Padua, L.A. Miguel, C.C. Fernandes Filho, et al. 2023. Evaluating the impact of modeling the family effect for clonal selection in potato-breeding programs. Front. Plant Sci. 14. doi: 10.3389/fpls.2023.1253706.

Meade, F., S. Byrne, D. Griffin, C. Kennedy, F. Mesiti, et al. 2020. Rapid Development of KASP Markers for Disease Resistance Genes Using Pooled Whole-Genome Resequencing. Potato Res. 63(1): 57–73. doi: 10.1007/s11540-019-09428-x.

Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics 157(4): 1819–1829. doi: 10.1093/genetics/157.4.1819.

Miller, M.D., C.A. Schmitz Carley, R.A. Figueroa, M.J. Feldman, D. Haagenson, et al. 2022. TubAR: an R Package for Quantifying Tuber Shape and Skin Traits from Images. Am. J. Potato Res. doi: 10.1007/s12230-022-09894-z.

Montesinos-López, O.A., A.W. Herr, J. Crossa, and A.H. Carter. 2023. Genomics combined with UAS data enhances prediction of grain yield in winter wheat. Front. Genet. 14. doi: 10.3389/fgene.2023.1124218.

Montesinos-López, O.A., A. Montesinos-López, J. Crossa, G. de los Campos, G. Alvarado, et al. 2017. Predicting grain yield using canopy hyperspectral reflectance in wheat breeding data. Plant Methods 13(1): 4. doi: 10.1186/s13007-016-0154-2.

Mróz, T., S. Shafiee, J. Crossa, O.A. Montesinos-Lopez, and M. Lillemo. 2024. Multispectral-derived genotypic similarities from budget cameras allow grain yield prediction and genomic selection augmentation in single and multi-environment scenarios in spring wheat. Mol Breeding 44(1): 5. doi: 10.1007/s11032-024-01449-w.

Neilson, J.A.D., A.M. Smith, L. Mesina, R. Vivian, S. Smienk, et al. 2021. Potato Tuber Shape Phenotyping Using RGB Imaging. Agronomy 11(9): 1781. doi: 10.3390/agronomy11091781.

Nie, X., D. Sutherland, V. Dickison, M. Singh, A.M. Murphy, et al. 2016. Development and Validation of High-Resolution Melting Markers Derived from Rysto STS Markers for High-Throughput Marker-Assisted Selection of Potato Carrying Rysto. Phytopathology® 106(11): 1366–1375. doi: 10.1094/PHYTO-05-16-0204-R.

Ortiz, R. 2020. Genomic-Led Potato Breeding for Increasing Genetic Gains: Achievements and Outlook. Crop Breeding, Genetics and Genomics 2(2). doi: https://doi.org/10.20900/cbgg20200010.

Perez, B.C., M.C.A.M. Bink, K.L. Svenson, G.A. Churchill, and M.P.L. Calus. 2022. Adding gene transcripts into genomic prediction improves accuracy and reveals sampling time dependence. G3 (Bethesda) 12(11): jkac258. doi: 10.1093/g3journal/jkac258.

Pérez, P., and G. de los Campos. 2014. Genome-wide regression and prediction with the BGLR statistical package. Genetics 198(2): 483–495. doi: 10.1534/genetics.114.164442.

Prodhomme, C., D. Esselink, T. Borm, R.G.F. Visser, H.J. van Eck, et al. 2019. Comparative Subsequence Sets Analysis (CoSSA) is a robust approach to identify haplotype specific SNPs; mapping and pedigree analysis of a potato wart disease resistance gene Sen3. Plant Methods 15(1): 60. doi: 10.1186/s13007-019-0445-5.

R Core Team. (2022). CRAN mirrors. Retrieved May 25, 2023, from https://cran.r-project.org/mirrors.html

Rieman, G.H. 1962. Superior: A new white, medium-maturing, scabresistant potato variety with high chipping quality. American Potato Journal 39(1): 19–28. doi: 10.1007/BF02912628.

Rodríguez-Álvarez, M.X., Boer, M.P., van Eeuwijk, F.A., & Eilers, P.H.C. (2018). Correcting for spatial heterogeneity in plant breeding experiments with P-splines. Spatial Statistics, 23, 52–71. https://doi.org/10.1016/j.spasta.2017.10.003

Schmitz Carley, C.A., J.J. Coombs, M.E. Clough, W.S. De Jong, D.S. Douches, et al. 2019. Genetic Covariance of Environments in the Potato National Chip Processing Trial. Crop Science 59(1): 107–114. doi: 10.2135/cropsci2018.05.0314.

Shakoor, N., S. Lee, and T.C. Mockler. 2017. High throughput phenotyping to accelerate crop breeding and monitoring of diseases in the field. Current Opinion in Plant Biology 38: 184–192. doi: 10.1016/j.pbi.2017.05.006.

Silva, M.F. e, G.M. Maciel, R.B. Gallis, R.L. Barbosa, V.Q. Carneiro, et al. 2022. High-throughput phenotyping by RGB and multispectral imaging analysis of genotypes in sweet corn. Hortic. Bras. 40: 92–98. doi: 10.1590/s0102-0536-2022012.

Singh, V., M. Krause, D. Sandhu, R.S. Sekhon, and A. Kaudal. 2023. Salinity stress tolerance prediction for biomass-related traits in maize (Zea mays L.) using genome-wide markers. The Plant Genome. 16:4. https://doi.org/10.1002/tpg2.20385

Slater, A.T., N.O.I. Cogan, J.W. Forster, B.J. Hayes, and H.D. Daetwyler. 2016. Improving Genetic Gain with Genomic Selection in Autotetraploid Potato. The Plant Genome 9(3): plantgenome2016.02.0021. doi: https://doi.org/10.3835/plantgenome2016.02.0021.

Song, P., J. Wang, X. Guo, W. Yang, and C. Zhao. 2021. High-throughput phenotyping: Breaking through the bottleneck in future crop breeding. The Crop Journal 9(3): 633–645. doi: 10.1016/j.cj.2021.03.015.

Sood, S., V. Bhardwaj, S.K. Kaushik, and S. Sharma. 2020. Prediction based on estimated breeding values using genealogy for tuber yield and late blight resistance in auto-tetraploid potato (Solanum tuberosum L.). Heliyon 6(11): e05624. doi: 10.1016/j.heliyon.2020.e05624.

Stefaniak, T.R., S. Fitzcollins, R. Figueroa, A.L. Thompson, C. Schmitz Carley, et al. 2021. Genotype and Variable Nitrogen Effects on Tuber Yield and Quality for Red Fresh Market Potatoes in Minnesota. Agronomy 11(2): 255. doi: 10.3390/agronomy11020255.

Stich, B., and D. Van Inghelandt. 2018. Prospects and Potential Uses of Genomic Prediction of Key Performance Traits in Tetraploid Potato. Front. Plant Sci. 9. doi: 10.3389/fpls.2018.00159.

Su, W.-H., and H. Xue. 2021. Imaging Spectroscopy and Machine Learning for Intelligent Determination of Potato and Sweet Potato Quality. Foods 10(9): 2146. doi: 10.3390/foods10092146.

Tuttle, H.K., A.H. Del Rio, J.B. Bamberg, and L.M. Shannon. 2024. Potato soup: analysis of cultivated potato gene bank populations reveals high diversity and little structure. Front. Plant Sci. 15. doi: 10.3389/fpls.2024.1429279.

VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. J Dairy Sci 91(11): 4414–4423. doi: 10.3168/jds.2007-0980.

Vos, P.G., J.G.A.M.L. Uitdewilligen, R.E. Voorrips, R.G.F. Visser, and H.J. van Eck. 2015. Development and analysis of a 20K SNP array for potato (Solanum tuberosum): an insight into the breeding history. Theor Appl Genet 128(12): 2387–2401. doi: ros.

Webb, R.E., D.R. Wilson, J.R. Shumaker, B. Graves, M.R. Henninger, et al. 1978. Atlantic: A new potato variety with high solids, good processing quality, and resistance to pests. American Potato Journal 55(3): 141–145. doi: 10.1007/BF02852087.

Weigle, J.L., A.E. Kehr, R.V. Akeley, and J.C. Horton. 1968. Chieftain: A red-skinned potato with attractive appearance and broad adaptability. American Potato Journal 45(8): 293–296. doi: 10.1007/BF02850285.

Xu, S., X. Xu, Q. Zhu, Y. Meng, G. Yang, et al. 2023. Monitoring leaf nitrogen content in rice based on information fusion of multi-sensor imagery from UAV. Precision Agric. doi: 10.1007/s11119-023-10042-8.

Yang, H., Y. Hu, Z. Zheng, Y. Qiao, K. Zhang, et al. 2022. Estimation of Potato Chlorophyll Content from UAV Multispectral Images with Stacking Ensemble Algorithm. Agronomy 12(10): 2318. doi: 10.3390/agronomy12102318.

Ye, S., J. Li, and Z. Zhang. 2020. Multi-omics-data-assisted genomic feature markers preselection improves the accuracy of genomic prediction. Journal of Animal Science and Biotechnology 11(1): 109. doi: 10.1186/s40104-020-00515-5.

Yu, H., S. Jiang, and K.C. Land. 2015. Multicollinearity in hierarchical linear models. Social Science Research 53: 118–136. doi: 10.1016/j.ssresearch.2015.04.008.

Yusuf, M., M.D. Miller, T.R. Stefaniak, D. Haagenson, J.B. Endelman, A.L. Thompson, and L.M. Shannon. 2024. Genomic prediction for potato (Solanum tuberosum) quality traits improved through image analysis. The Plant Genome: e20507.

Zhou, Z., M. Jabloun, F. Plauborg, and M.N. Andersen. 2018. Using ground-based spectral reflectance sensors and photography to estimate shoot N concentration and dry matter of potato. Computers and Electronics in Agriculture 144: 154–163. doi: 10.1016/j.compag.2017.12.005.

Zhu, X., W.L. Leiser, V. Hahn, and T. Würschum. 2021. Phenomic selection is competitive with genomic selection for breeding of complex traits. The Plant Phenome Journal 4(1): e20027. doi: 10.1002/ppj2.20027.

Chapter 4

Prediction of total plant nitrogen based on multispectral reflectance for evaluation of nitrogen use efficiency in advance potato (*Solanum tuberosum*) breeding clones

Overview

Potato production in the United States predominantly occurs on sandy soils, where up to 60% of applied nitrogen (N) is lost to the environment, contributing to pollution and resource inefficiency. Accurately assessing and predicting total plant nitrogen (TPN), critical for the growth and development of different potato varieties, is essential for optimizing N management. This study utilized drone-derived multispectral reflectance data and vegetation indices (VIs) collected at five growth stages (from emergence to senescence) to predict TPN using four machine learning regression models: Ridge Regression (RR), Random Forest Regression (RFR), Elastic Network Regression (ENR), and Regression Cubist Learner (RCL).

We evaluated 24 and 20 clones each, representing chips and fresh market varieties, in replicated 15-hill plots over two growing seasons (2022 and 2023) under four N rates (recommended, 66% of recommended, 33% of recommended, and no added N). Biomass accumulation and N concentration were assessed from vine and tuber samples collected before senescence. TPN was derived from N related traits, including dry matter (DM), and total tuber yield (TTY). The findings revealed moderate to high correlation between TPN and multispectral variables (ranging from -1 to 0.9). While ENR achieved the highest prediction accuracy at individual time points, RFR and RCL outperformed when data across all time points were combined. This research

serves as the foundation for optimizing nitrogen inputs and advancing to the development of

potato clones with enhanced nitrogen use efficiency, thereby supporting sustainable agricultural

practices.

4.1 Introduction

Potato (*Solanum tuberosum* L.) ranks among the most significant non-grain food crops globally,

playing a crucial role in global food security. Its productivity and tuber quality are strongly

influenced by soil moisture and nutrient availability, particularly nitrogen (N). Nitrogen is a vital

macronutrient for potato development, supporting processes of vegetative growth, and tuber

formation (Haverkort and Verhagen, 2008). However, the excessive application of N, often

driven by the goal of maximizing yield, poses environmental risks, including nitrate leaching,

greenhouse gas emissions, and reduced soil health (Kraft and Stites, 2003). Growing N efficient

potato varieties helps minimize added N and it's lost to the environment. Breeders therefore need

to develop N efficient varieties to ensure precise nutrient management. Varieties that perform

well under high N conditions don't necessarily replicate same under low added N which

therefore necessitate the early screening for nitrogen use efficiency (Jones et al., 2021).

However, to measure N use efficiency and other component parts requires the assessment and

prediction of total plant N content. The assessment through conventional methods have mostly

relied on direct sampling of potato vines and tubers, but these approaches are time consuming,

labor intensive that involves lots of resources for sampling and prone to inconsistencies

particularly for large-scale fields.

Remote sensing (RS) technologies, combined with advancements in machine learning

(ML) algorithms, present an alternative for assessment and prediction of total N in most plants.

Images captured by unmanned aerial vehicles (UAVs) equipped with sensors across various

bandwidths provide reflectance data from sunlight reflected off plant canopies. This reflectance,

influenced by absorption processes within the leaf, reveals key physicochemical characteristics

of plants. The visible spectrum (RGB) ranges from 400-680 nm, while beyond-visible

wavelengths exceed 680 nm, these include near-infrared and red edge for assessing plant health and nutrient uptake (Zhao et al., 2020; Peng et al., 2021b).

Key insights into plant properties are derived from vegetation indices (VIs), which are formulated as mathematical relationships between spectral reflectance values. VIs, such as the widely used Normalized Difference Vegetation Index (NDVI), highlight contrasts between red light absorption (by leaf pigments) and NIR scattering, offering indirect yet reliable indications of plant health and nitrogen status Rouse et al. (1974). Additional indices, such as those developed by McFeeters (1994), Haboudane et al. (2010), and Gitelson et al. (2012), have further enhanced in-season monitoring of nitrogen dynamics in vegetation.

The use of simple regression models have been employed to predict plant nitrogen content, they often fall short due to multicollinearity in spectral data. The high correlation among spectral bands creates redundancies that basic models cannot resolve effectively (Liu & Li, 2017). Advanced ML algorithms, capable of managing non-linear and complex relationships, provide a more accurate and reliable solution for nitrogen prediction.

Numerous studies have applied ML models in conjunction with multispectral or hyperspectral data to predict N content in potato crops (Morier et al., 2015; Zheng et al., 2018; Peng et al., 2021a; Liu et al., 2022; Ye et al., 2024). Despite their promising results, these studies often face limitations due to small sample sizes and a narrow focus on widely cultivated commercial varieties. Broadening these models to include larger, more diverse potato populations would significantly enhance their generalizability and application. This in turn will provide accurate measures of nitrogen use efficiency which usually requires larger population as opposed to few varieties that have been used previously for building this prediction models. Such

advancements could mirror successes seen in other crops, facilitating the development of new nitrogen-efficient potato clones (Yamashita et al., 2020; Li et al., 2023).

Therefore, this study aims to explore the potential of multispectral data, collected at different growth stages using UAVs, to predict total plant N in a breeding program through the application of diverse machine learning (ML) models. The study is focused on diverse set of clones within the University of Minnesota (UMN) potato breeding program based on two primary market classes: chipping potatoes and fresh market potatoes

4.2. Materials and Methods

4.2.1 Plant materials and management

In order to evaluate plant's response to nitrogen, we trialed 45 clones at 4 different N levels in 2022 and 2023 at the Sand Plains Research Farm in Becker, MN. A total of twenty-four and twenty-one advanced potato breeding clones of chippers and fresh market class respectively were trialed. The clones were evaluated in a row/column design with two replications. All plots contained 15 seed pieces with 30 cm in row spacing and 91 cm between rows.

Most commonly grown varieties were also included as part of trialed clones, these includes Atlantic (Webb et al., 1978), Superior (Rieman, 1962), Lamoka (De Jong et al., 2017), Snowden, and Cascade for chips market. The fresh market clones include Red Pontiac, Red Lasoda, Red Norland (Johansen et al., 1959), Modoc, Chieftain (Weigle et al., 1968), Yukon Gold (Johnston and Rowberry, 1981).

Four levels of nitrogen were applied at 0%, 33%, 66% and 100% of the total added at 252.17 kg/ha for chips market and 0%, 33%, 66% and 100% of the total added at 218.54 kg/ha for fresh market with an initial 28 kg/ha preplant application. The field management for irrigation,

pesticides and herbicides application followed routine practices across all treatments at the trial site (Jones et al., 2021; Stefaniak et al., 2021). The trials were established in May and harvested two weeks after vine kill in September of each year

4.2.2 Sample data collection and phenotyping

Excavation of single randomly sampled plant per plot excluding row-end plants was dug out before vine kill for vines (i.e., tissue attached to a whole plan above ground) and tuber collection. The vines from the whole plant were cut separate from tubers and collected into bags weighed as fresh vine weight (FVW) for determination of biomass and N concentration. Similarly, tubers were harvested from each plant and stored in meshed bags after cleaning off soil for biomass and N concentration. The tubers were then subsampled diced and weighed as fresh tuber weight (FTW) as part of the process for determination of dry matter and N concentration.

We weighed vine and tubers samples after drying at 70 degrees Celsius for at least 72 hours as dry vine weight (DVW) and dry tuber weight (DTW) respectively. The dried samples were ground through a 1-mm sieve screen and were sent to the University of Minnesota Research Analytical Lab for determination of N concentration via combustion (Jones et al., 2021; "Total Nitrogen (Plant) | Research Analytical Laboratory,") as NVINE and NTUBER.

4.2.3 Estimation total plant N (TPN)

The TPN trait is derived as a function of N concentration and dry matter which was calculated after harvest. First, we estimated dry matter percentage (DM) for both vine and tubers as dry weight divided by fresh weight (Table 1). The DM for vine was multiplied by wet weights and plant number per plot denoted as VINEDM and DM for tuber was multiplied by fresh total tuber

yield per plot (TTY) as TUBERDM. Nitrogen content of vine and tuber were obtained by multiplying the DM for each with N concentration as NCONVINE and NCONTUBER. The total plant N (TPN) is then the sum of both vine and tuber nitrogen content (Table 1). Total tuber yield (TTY) was measured on a tuber basis and summed over plots using an AccuVision Dual View L.E.D. Grader System from Exeter. All units of measurements were converted to metric systems accordingly for ease of estimation of TPN as presented in Table. 1.

4.2.4 Multispectral data

Multispectral reflectance data were captured in the summers of 2022 and 2023 for five different growth stages using an unmanned aerial vehicle (UAV) with a Sentera 6x multispectral and 6x thermal sensors (Sentera, St Paul, MN). The camera is equipped with high radiometric calibration tools for correction of lightening variation and to provide a consistent reflectance surface spanning both visible and near-infrared spectral bands. Raw multispectral reflectance mean value was obtained for five separate color bands (red = 650 nm, green = 545 nm, blue = 445nm, NIR = 840 nm and RedEdge = 720 nm) per plot after image preprocessing and radiometric calibration from Sentera. Using the raw bands, eight different vegetation indices commonly used in potatoes (Alkhaled et al., 2023) were estimated, they include the Normalized Difference Vegetation Index (NDVI)(Rouse et al., 1974), Normalized Difference Red Edge (NDRE:)(Gitelson and Merzlyak, 1994a), Green Leaf Index (GLI), Chlorophyll Index Green (CIG), Green Normalized Difference Vegetation Index (GNDVI)(Gitelson and Merzlyak, 1994b), Normalized Difference Water Index (NDWI)(McFeeters, 1994), Chlorophyll Index Red Edge (CIRE)(Gitelson et al., 2012), and Transformed Chlorophyll Absorption in Reflectance Index/ Optimized Soil-Adjusted Vegetation Index (TCARI/OSAVI)(Haboudane et al., 2010).

Five flight dates were used each summer based on estimated phenological growth stages including Canopy cover at 50%, Canopy cover closure, Peak Flowering, Peak flowering/ Biomass, Senescence (Table 2). Obviously, the timing of phenological stage varies by genotype, N rate and market class, so an estimate was made for the whole field.

4.2.5 Statistical analysis

4.2.5.1 Trait distribution and BLUEs estimation

The best linear unbiased estimates (BLUEs) for TPN and for all multispectral bands and indices with different N levels were computed using a spatial 2D spline model, implemented with SpATS (Rodríguez-Álvarez et al., 2018), within the StageWise package (Endelman, 2023) in R ("R Core Team," 2022) based on two-stage approach as :

$$y_{ijk} = u + g_i + r_j + c_k + e_{ijk} \tag{1}$$

Where y_{ijk} is the response variable for the ith genotype in jth row and kth column, μ is the overall mean, g_i is the fixed effect for the genotypic response, r_j is the random row effect where $\boldsymbol{r_j} \sim N$ $(0, \sigma_r^2)$, c_k is the random column effect where $\boldsymbol{c_j} \sim N (0, \sigma_c^2)$, and e_{ijk} is the residual effect where $\boldsymbol{e_{ijk}} \sim N (0, \sigma_e^2)$.

Using Pearson correlation coefficients, we looked at pairwise relationships between BLUEs for TPN and BLUEs for multispectral bands and indices across N rates and years. This was to identify spectral bands or indices most correlated with the TPN traits.

4.2.5.2 Prediction of TPN with Machine learning models

To predict total plant nitrogen (TPN) based on multispectral bands and indices collected at various growth stages and their combinations across market classes, we implemented four machine learning (ML) models using the CARET R package (Kuhn, 2008): Ridge Regression

(RR) (Hoerl and Kennard, 1970), Random Forest Regression (RFR)(Breiman, 2001), Elastic

Network Regression (ENR)(Zou and Hastie, 2005), and Regression Cubist Learner (RCL)

(Birnbaum, 2014). These models were compared, and the best-performing model was selected

for hyperparameter tuning and evaluation on the test set.

The dataset was divided into 75% for training and 25% for testing. The training set was

resampled using 10-fold cross-validation with 10 repetitions to optimize model performance.

Once the best-performing model was identified, it was used to select significant predictor

variables (bands and indices) based on feature importance scores intrinsic to the model.

Predictors contributing at least 20% to the overall prediction were considered significant.

For feature selection, the training data was further partitioned into an 80%-20% split. The larger

subset was used to identify significant features, while the smaller subset was used to train a new

model using only the selected predictors. This refined model was then evaluated on the initial

test set.

Prediction accuracy was measured using coefficient of determination (R^2) as Pearson´s

correlation between the predicted and observed values.

4.3 Results

4.3.1 Distribution and correlations of traits and multispectral variables

The BLUEs for total plant nitrogen content (TPN) varied across N rates and years within

the fresh market class and but were less variable within the chipping clones (Figure 1.). The

same distribution was observed for all spectra bands and vegetation indices (Figure S1.) across

the different growth stages, years and market class. Correlation of TPN with multispectral

variables was low to moderate across different growth stages and market class. The correlation

of multispectral variables across growth stages and N rates ranged from -0.58 to 0.58 within the fresh market. Among the bands and indices, NIR and NDRE consistently exhibited the highest correlation values, as shown in Figures 2a and 2b. For chips, the correlation ranged from -0.66 to 0.70 across different N rate levels. The red band and NDWI consistently demonstrated the highest correlation across growth stages, except for stage 5, within the bands and indices (Figures 3a and 3b).

4.3.2 Evaluating prediction ability

To predict total nitrogen at harvest from multispectral images taken at a series of growth stages using machine learning models, we separated our population into training and testing set for each growth stage and all combined stages. We evaluated prediction ability for each training set with four different regression model (Figure 4a and 4b). The distribution of prediction ability within the fresh market clones was 0.003 to 0.93 across all growth stages while combined stages performed better than each of the individual growth stages. The regression models performed similarly within each growth stage. In the combined stages ENR, RFR, and RCL outperformed RR, with RFR exhibiting the highest mean prediction ability of 0.56. The prediction abilities for the chipping clones were also highly variable but much lower (Figure 3b). We did not observe the same positive effect of combining stages for the chips.

4.3.3 Variable feature selection and prediction accuracy assessment

To improve the prediction of total plant nitrogen (TPN), the best-performing regression model, after hyperparameter tuning, was used for feature selection within the training set. The feature variable retained were used to fit a new regression model, which was then applied to predict the test set (Table 3). Prediction accuracy across individual growth stages within the fresh market

was relatively consistent but showed a notable improvement at the combined stage, achieving R^2 values of 0.53 for both the training and test sets. Features selected at each growth stage and the combined stage were used for final predictions respectively. The ENR model was primarily utilized for feature selection and final prediction in both the training and test sets, except for stage 5 and the combined stage, where Random Forest Regression (RFR) was used.

For chipping clones, prediction accuracy remained consistently low across growth stages, even after feature selection. While the ENR model was used for both feature selection and final model fitting, it was outperformed by the Regression Cubist Learner (RCL) at the combined stage (Table 3). Notably, most of the selected multispectral features overlapped across growth stages and between market classes (Table S1a and S1b).

4.4 Discussion

Total plant nitrogen (TPN) in potato is a critical quantitative trait that directly influences the efficiency of nitrogen use in biomass accumulation. A more economical and timely quantification of TPN is essential for optimizing agronomic practices and supporting the breeding of cultivars with improved nitrogen use efficiency.

In this study, we utilized high-throughput remote sensing data derived from UAV-based multispectral reflectance and vegetation indices as predictors for TPN. Using various machine learning models, we analyzed TPN prediction at individual phenological growth stages and combined stages. This approach allowed us to identify the reflectance bands and vegetation indices that contribute most significantly to TPN prediction, providing valuable insights for agronomic and breeding decision-making.

4.4.1 Using multispectral data for prediction of total plant N

To model TPN prediction, we first examined its correlation with multispectral bands and vegetation indices across N rate treatments for both market classes. Our results revealed a moderate correlation between TPN and spectral bands, particularly for NIR within the fresh market class. Notably, the correlation values exhibited an upward trend as growth progressed, suggesting that spectral band values change with phenological changes.

Vegetation indices showed stronger and more consistent correlations with TPN compared to individual spectral bands, indicating that certain indices are more closely related to TPN regardless of growth stage. These findings align with previous research that identified indices such as NDVI, GNDVI, and CIRE as significantly correlated with plant health, above-ground biomass, and total yield (Morier et al., 2015; Liu et al., 2022; Alkhaled et al., 2023; Kumar et al., 2023).

In predicting TPN using multispectral reflectance bands and vegetation indices, we observed comparable accuracy across individual growth stages with the different machine learning models employed. This suggests that the choice of model may not significantly influence prediction accuracy at individual stages. However, combining data from all growth stages resulted in notably improved TPN prediction within the fresh market class. This improvement highlights the value of leveraging a combined dataset that captures variability across growth stages, providing a more comprehensive representation of spectral features.

Our findings are consistent with prior studies that employed a combination of spectral variables and diverse models to predict plant nitrogen content. For instance, Peng et al. (2021a), Li et al. (2023) and Ye et al. (2024) demonstrated improved prediction accuracy when using

combined datasets and models, including ensemble methods such as model stacking, for predicting plant nitrogen using multispectral reflectance and vegetation indices.

Prediction of TPN within the chips market class remained consistently low across all individual growth stages and the combined stages. As in the fresh market class, the best-performing models were utilized for feature variable selection and final model fitting on both training and testing datasets. Among these models, ENR demonstrated superior performance across individual growth stages, while RCL outperformed others for the combined stages. This result is not unexpected, given the limited variation observed in TPN and TDM within the chips market class, which likely constrained the models' predictive capabilities. Furthermore, the low prediction accuracy may also stem from reduced nitrogen uptake and inefficient nitrogen utilization, both of which are characteristic of this market class (Zvomuya et al., 2002; Ye et al., 2024).

4.4.2 Implications of selected multispectral features for predicting total plant N (TPN)

Evaluating prediction models across growth stages highlighted the value of using the best-performing models for feature selection, leading to potential improvements in prediction accuracy. The selected multispectral reflectance and vegetation indices varied across growth stages for both market classes, although there was significant overlap in key indices at specific stages. Notably, CIRE, NDRE, and NDVI consistently emerged as the most significant predictors across all growth stages and market classes, substantially enhancing prediction accuracy. These indices have been extensively validated in previous studies for their ability to estimate nutrient status in cereal crops, particularly wheat (Bai et al., 2019; Li et al., 2023).

The improved model accuracy observed in our results after feature selection aligns with findings from (Lebourgeois et al., 2012; Morier et al., 2015; Yamashita et al., 2020; Ye et al., 2024) all of whom demonstrated the effectiveness of multispectral or hyperspectral indices in monitoring and predicting nitrogen status. This reinforces the importance of selecting optimal spectral features for precise TPN prediction.

4.5 Conclusion

This study demonstrated the potential of using remote sensing data for predicting and quantifying total plant nitrogen in potato across different growth stages with various machine learning models. It provided critical insights into the interaction between nitrogen dynamics and multispectral data, the significance of aggregated data across growth stages for improving prediction accuracy. The findings underscore the potential of integrating remote sensing data and machine learning models to optimize nitrogen inputs for agronomic decision-making and facilitate the breeding of nitrogen-efficient cultivars. While further validations are necessary to enable widespread adoption of these high-throughput procedures, this approach offers a promising alternative to the traditional, labor-intensive process of field sampling for TPN quantification.

Table 1: Total Plant Nitrogen and related traits measured

Parameter	Abbreviation	formula
Applied N (kg/ha)	NLEVEL	Weighing scale
Fresh vine weight(g)	FVW	Weighing scale
Dry vine weight(g)	DVW	Weighing scale
Fresh tuber weight(g)	FTW	Weighing scale
Dry tuber weight(g)	DTW	Weighing scale
Total tuber fresh weight(mg/ha)	TTY	Grader
Nitrogen concentration (%)	NVINE	lab analytical estimate
	NTUBER	lab analytical estimate
Dry matter	VINEDM	DVW/FVW
	TUBERDM	(DTW/FTW) X TTY
Nitrogen content (kg/ha)	NCONVINE	NVINE X VINEDM
	NCONTUBER	NTUBER X TUBERDM
Total Plant N (kg/ha)	TPN	NCONVINE + NCONTUBER
Total dry matter (kg/ha)	TDM	VINEDM + TUBERDM

Table 2: Timing of drone flights and number of clones evaluated for different N levels

Class	Evaluation Year	Total unique entries	N rates (kg/ha)	Growth stages/ Drone flight Dates	Days after planting
Chips market	2022	20	28 96.37 170.35 252.17	Canopy cover 50%: June 28 Canopy cover flower: July 05 Peak flowering: July 12 Vine biomass: July 19 Senescence: August 05	36 45 52 59 76
	2023	20		Canopy cover 50%: June 20 Canopy cover flower: July 05 Peak flowering: July 13 Vine biomass: July 18 Senescence: August 08	30 45 53 58 79
Fresh market	2022	20	28 85.28 128.16 218.54	Canopy cover 50%: June 28 Canopy cover flower: July 05 Peak flowering: July 12 Vine biomass: July 19 Senescence: August 05	36 45 52 59 76
	2023	20		Canopy cover 50%: June 20 Canopy cover flower: July 05 Peak flowering: July 13 Vine biomass: July 18 Senescence: August 08	30 45 53 58 79

Figure 1. Distribution of total plant N across years and different N application levels as a percentage of 197 kg/ha for fresh market and 230kg/ha for chips market
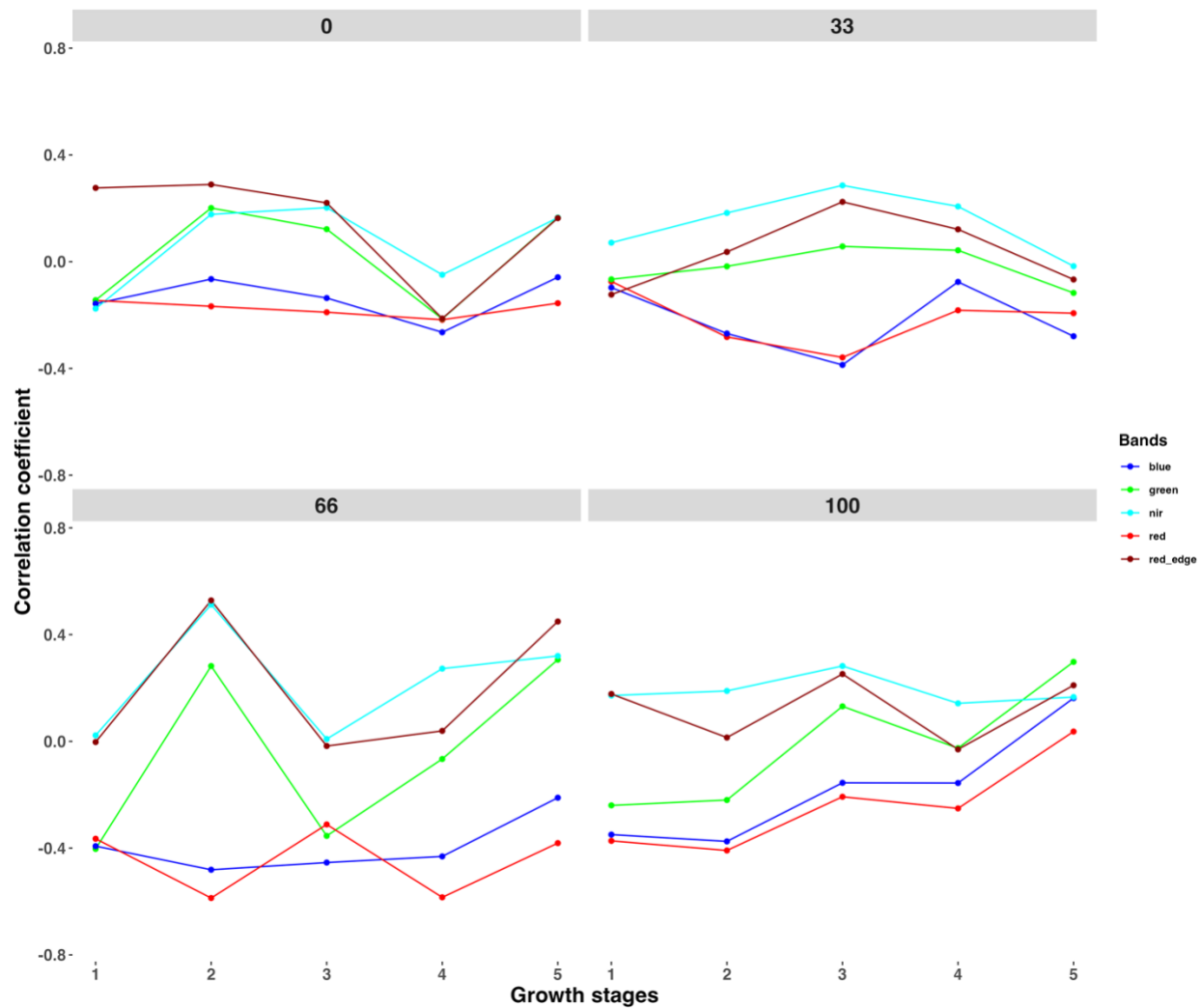
Figure 2a. Correlation of TPN with all multispectral bands across growth stages and N rates within the fresh market.
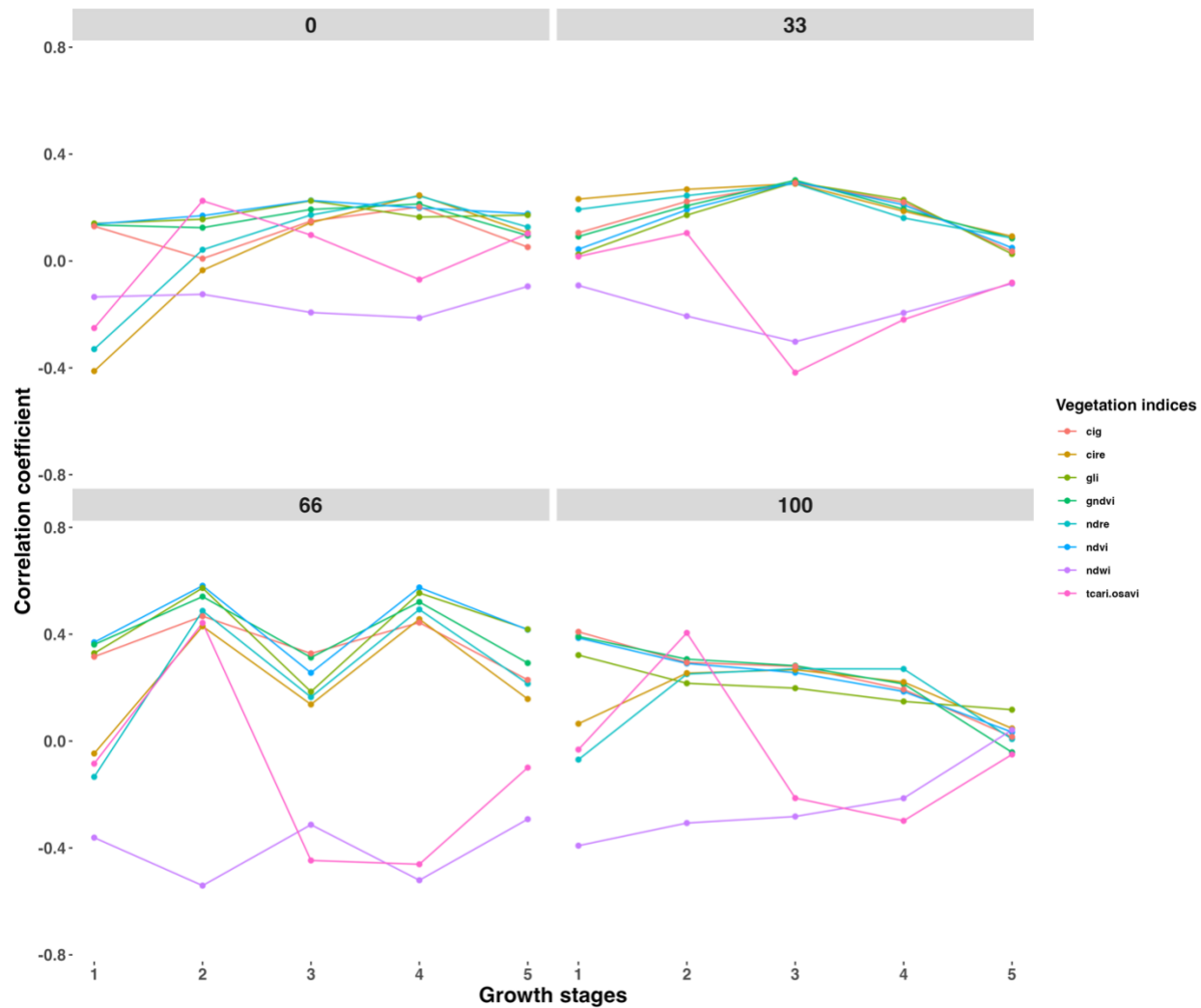
Figure 2b. Correlation of TPN with all indices across growth stages and N rates within the fresh market.
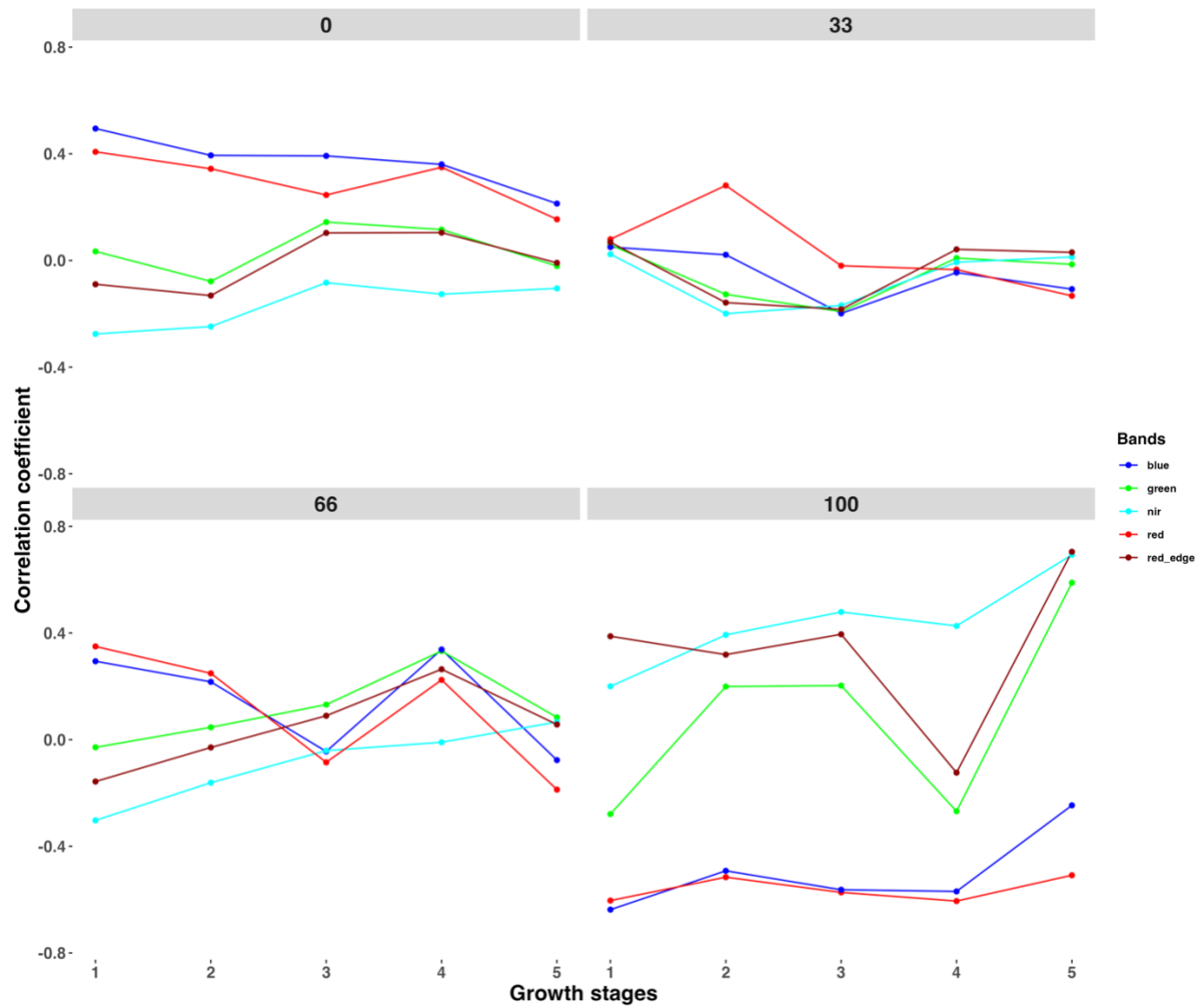
Figure 3a. Correlation of TPN with all multispectral bands across growth stages and N rates within chipping potatoes
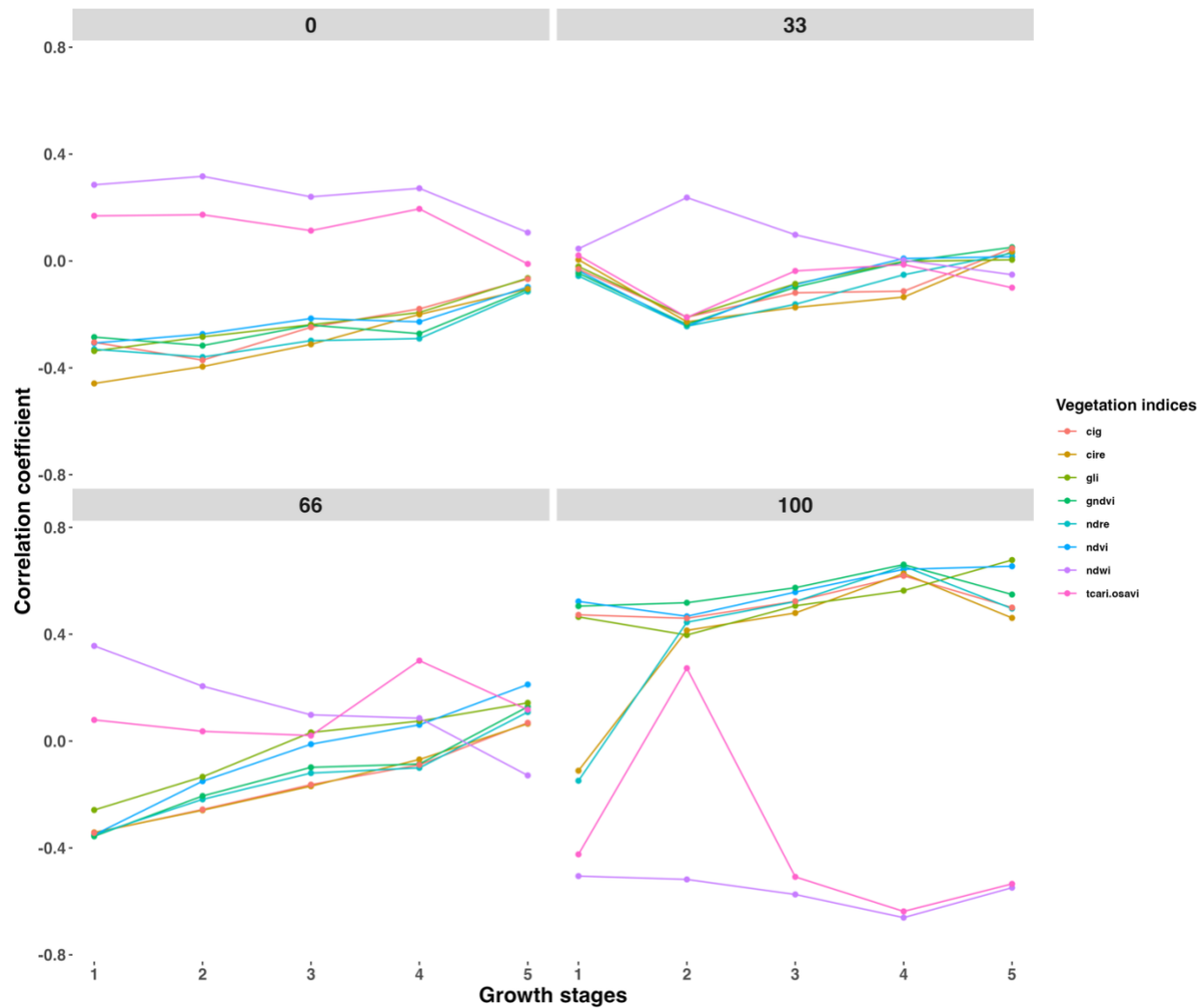
Figure 3b. Correlation of TPN with all indices across growth stages and N rates within chipping potatoes
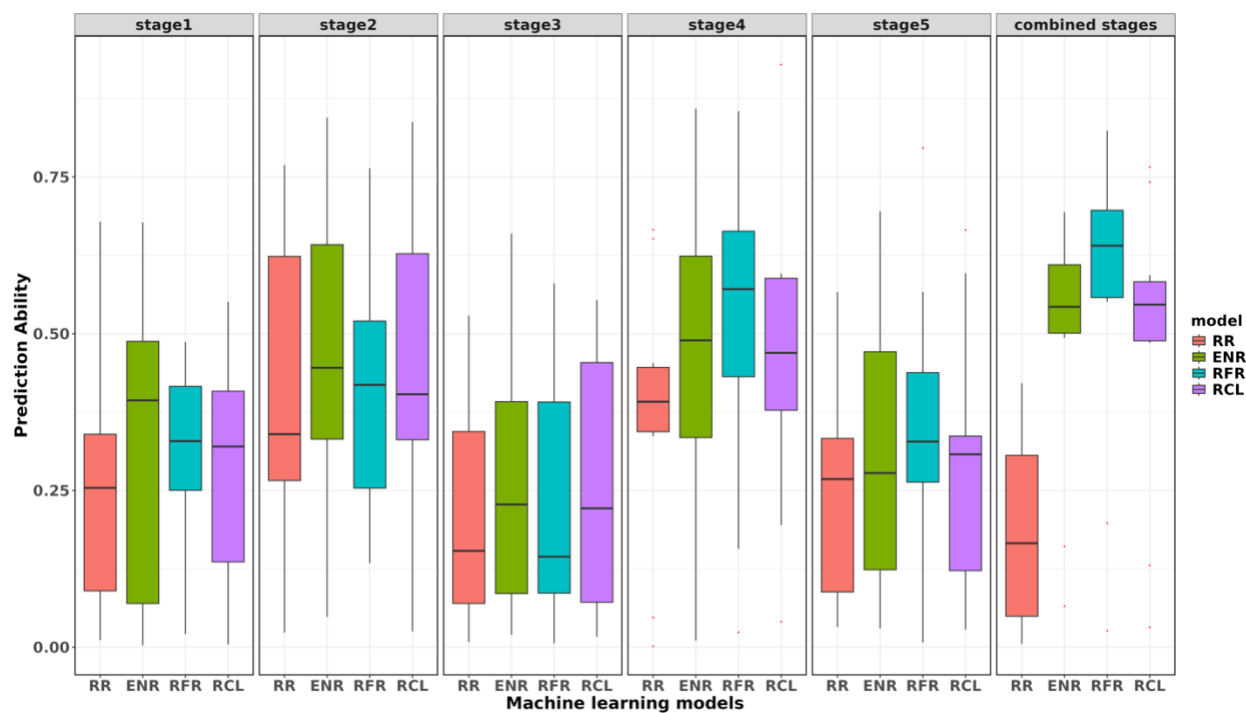
Figure 4a. Distribution of prediction ability for at different growth stages with multispectral bands and indices using different machine learning regression models within the fresh market class.

RR Ridge Regression, ENR Elastic Network Regression, RFR Random Forest Regression and

RCL Regression Cubist Learner.
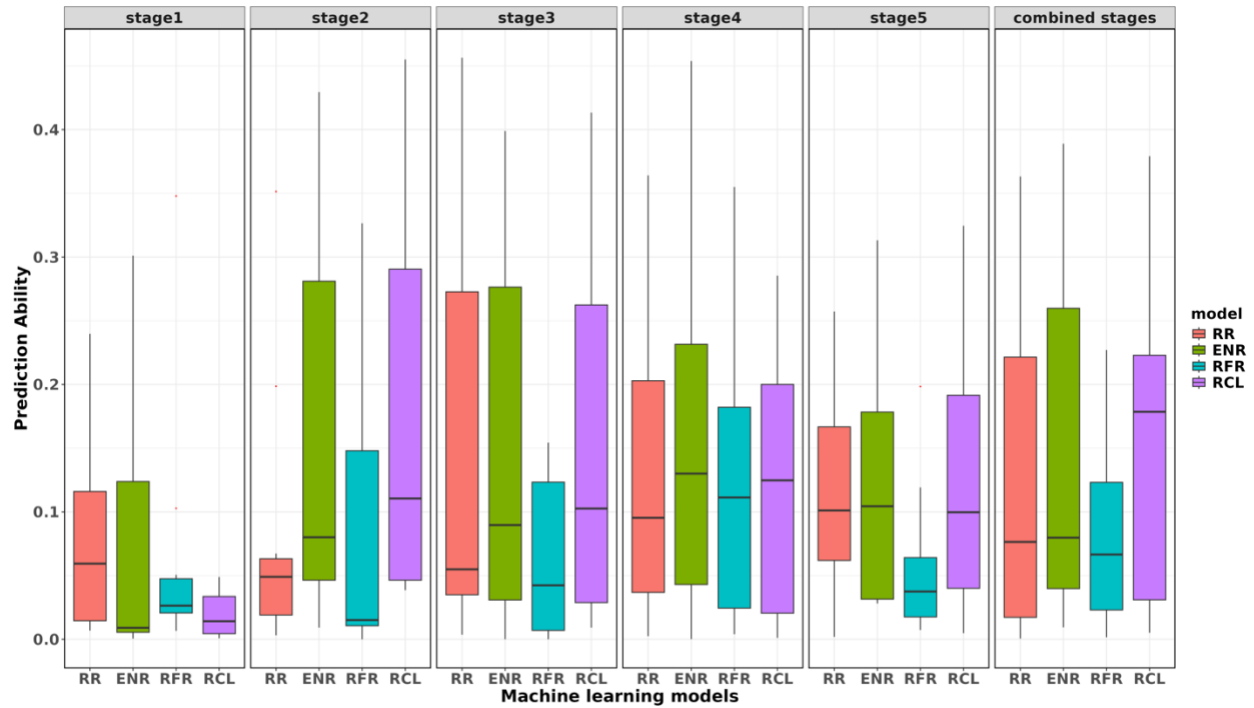
Figure 4b. Distribution of prediction ability of TPN at different growth stages with multispectral bands and indices using different machine learning regression models within the chips market class.

RR Ridge Regression, ENR Elastic Network Regression, RFR Random Forest Regression and RCL Regression Cubist Learner.

Table 3. Prediction accuracy in terms of coefficient of determination across growth stages with best performing models

	Stage 1 (R2)	Stage 2 (R2)	Stage 3 (R2)	Stage 4 (R2)	Stage 5 (R2)	Combined (R2)
Fresh market						
Training	0.26	0.45	0.22	0.44	0.33	0.53
Testing	0.39	0.47	0.42	0.47	0.38	0.53
model	ENR	ENR	ENR	ENR	RFR	RFR
Chips market						
Training	0.10	0.17	0.20	0.18	0.13	0.15
Testing	0.22	0.11	0.12	0.12	0.04	0.13
model	ENR	ENR	ENR	ENR	ENR	RCL

ENR Elastic Network Regression, RCL Regression Cubist Learner

References

Abdelhakim, L.O.A., B. Pleskačová, N.Y. Rodriguez-Granados, R. Sasidharan, L.S. Perez-Borroto, et al. 2024. High Throughput Image-Based Phenotyping for Determining Morphological and Physiological Responses to Single and Combined Stresses in Potato. J Vis Exp (208). doi: 10.3791/66255.

Agha, H.I., J.B. Endelman, J. Chitwood-Brown, M. Clough, J. Coombs, et al. 2024. Genotype-by-environment interactions and local adaptation shape selection in the US National Chip Processing Trial. Theor Appl Genet 137(5): 99. doi: 10.1007/s00122-024-04610-3.

Alkhaled, A., P.A. Townsend, and Y. Wang. 2023. Remote Sensing for Monitoring Potato Nitrogen Status. Am. J. Potato Res. 100(1): 1–14. doi: 10.1007/s12230-022-09898-9.

Asano, K., and J.B. Endelman. 2023. Development of KASP markers for the potato virus Y resistance gene Rychc using whole-genome resequencing data. : 2023.12.20.572658. doi: 10.1101/2023.12.20.572658.

Bai, Y., Y. Zhang, Y. Wang, Y. Liu, X. Zhang, et al. 2019. Using remote sensing to assess nitrogen status in wheat. Field Crops Research 240: 1–10.

Beavis, W.D., and J.P.M. van Heerwaarden. 2017. Mapping QTLs: A comparison of methods. Molecular Genetics of Extracellular Plant Pathogens. Springer. p. xx–xx

Birnbaum, L.A. 2014. Machine Learning Proceedings 1993: Proceedings of the Tenth International Conference on Machine Learning, University of Massachusetts, Amherst, June 27-29, 1993. Morgan Kaufmann.

Bourke, P.M., R.E. Voorrips, R.G.F. Visser, and C. Maliepaard. 2018. Tools for Genetic Studies in Experimental Populations of Polyploids. Front Plant Sci 9. doi: 10.3389/fpls.2018.00513.

Bradshaw, J.E. 2017. Review and Analysis of Limitations in Ways to Improve Conventional Potato Breeding. Potato Res. 60(2): 171–193. doi: 10.1007/s11540-017-9346-z.

Breiman, L. 2001. Random Forests. Machine Learning 45(1): 5–32. doi: 10.1023/A:1010933404324.

Caraza-Harter, M.V., and J.B. Endelman. 2020. Image-based phenotyping and genetic analysis of potato skin set and color. Crop Science 60(1): 202–210. doi: 10.1002/csc2.20093.

Clot, C.R., C. Polzer, C. Prodhomme, C. Schuit, C.J.M. Engelen, et al. 2020. The origin and widespread occurrence of Sli-based self-compatibility in potato. Theor Appl Genet. doi: 10.1007/s00122-020-03627-8.

Crossa, J., P. Pérez-Rodríguez, J. Cuevas, O. Montesinos-López, D. Jarquín, et al. 2017. Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. Trends in Plant Science 22(11): 961–975. doi: 10.1016/j.tplants.2017.08.011.

De Jong, W.S., D.E. Halseth, R.L. Plaisted, X. Wang, K.L. Perry, et al. 2017. Lamoka, a Variety with Excellent Chip Color Out of Cold Storage and Resistance to the Golden Cyst Nematode. Am. J. Potato Res. 94(2): 148–152. doi: 10.1007/s12230-016-9557-x.

D'hoop, B.B., M.J. Paulo, K. Kowitwanich, M. Sengers, R.G.F. Visser, et al. 2010. Population structure and linkage disequilibrium unravelled in tetraploid potato. Theoretical and Applied Genetics 121: 1151–1170. doi: 10.1007/s00122-010-1379-5.

Eck, H.J. van, P.G. Vos, J.P.T. Valkonen, J.G.A.M.L. Uitdewilligen, H. Lensing, et al. 2017. Graphical genotyping as a method to map Ny (o,n)sto and Gpa5 using a reference panel of tetraploid potato cultivars. Theor Appl Genet 130(3): 515–528. doi: 10.1007/s00122-016-2831-y.

Endelman, J.B. 2023. Fully efficient, two-stage analysis of multi-environment trials with directional dominance and multi-trait genomic selection. Theor Appl Genet 136(4): 65. doi: 10.1007/s00122-023-04298-x.

Endelman, J.B., C.A.S. Carley, P.C. Bethke, J.J. Coombs, M.E. Clough, et al. 2018. Genetic Variance Partitioning and Genome-Wide Prediction with Allele Dosage Information in Autotetraploid Potato. Genetics 209(1): 77–87. doi: 10.1534/genetics.118.300685.

Endelman, J.B., M. Kante, H. Lindqvist-Kreuze, A. Kilian, L.M. Shannon, et al. 2024. Targeted genotyping-by-sequencing of potato and software for imputation. : 2024.02.12.579978. doi: 10.1101/2024.02.12.579978.

Felcher, K.J., J.J. Coombs, A.N. Massa, C.N. Hansey, J.P. Hamilton, et al. 2012. Integration of Two Diploid Potato Linkage Maps with the Potato Genome Sequence. PLOS ONE 7(4): e36347. doi: 10.1371/journal.pone.0036347.

Feldman, M.J., J. Park, N. Miller, C. Wakholi, K. Greene, et al. 2024. A scalable, low-cost phenotyping strategy to assess tuber size, shape, and the colorimetric features of tuber skin and flesh in potato breeding populations. The Plant Phenome Journal 7(1): e20099. doi: 10.1002/ppj2.20099.

Freire, R., M. Weisweiler, R. Guerreiro, N. Baig, B. Hüttel, et al. 2021. Chromosome-scale reference genome assembly of a diploid potato clone derived from an elite variety. G3 (Bethesda): jkab330. doi: 10.1093/g3journal/jkab330.

Gitelson, A.A., S. Gritz, and M.N. Merzlyak. 2012. Chlorophyll content estimation in plant leaves using reflectance spectroscopy. Agronomy Journal 104(3): 721–727. doi: 10.2134/agronj2011.0370.

Gitelson, A.A., and M.N. Merzlyak. 1994a. In vivo optical properties of higher plant leaves. Plant Physiology 104(3): 1019–1028. doi: 10.1104/pp.104.3.1019.

Gitelson, A.A., and M.N. Merzlyak. 1994b. Remote sensing of chlorophyll a in higher plant leaves. Remote Sensing of Environment 50(3): 246–252. doi: 10.1016/0034-4257(94)90026-9.

Gutaker, R.M., C.L. Weiß, D. Ellis, N.L. Anglin, S. Knapp, et al. 2019. The origins and adaptation of European potatoes reconstructed from historical genomes. Nat Ecol Evol 3(7): 1093–1101. doi: 10.1038/s41559-019-0921-3.

Haboudane, D., C.A. Silva, D. Street, and P.J. Zarco-Tejada. 2010. Integrating remote sensing and crop growth models for irrigation management and yield forecasting. Field Crops Research 118: 150–162. doi: 10.1016/j.fcr.2010.06.007.

Habyarimana, E., B. Parisi, and G. Mandolino. 2017. Genomic prediction for yields, processing and nutritional quality traits in cultivated potato (Solanum tuberosum L.). Plant Breeding 136(2): 245–252. doi: 10.1111/pbr.12461.

Hardigan, M.A., F.P.E. Laimbeer, L. Newton, E. Crisovan, J.P. Hamilton, 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 U S A 114(46): E9999–E10008. doi: 10.1073/pnas.1714380114.

Haverkort, A.J., and A. Verhagen. 2008. Potato production in the world: A global perspective. In: Haverkort, A.J. and Verhagen, A., editors, Potato Biology and Biotechnology: Advances and Perspectives. Elsevier. p. 1–20

Herrera, M. del R., L.J. Vidalon, J.D. Montenegro, C. Riccio, F. Guzman, et al. 2018. Molecular and genetic characterization of the Ryadg locus on chromosome XI from Andigena potatoes conferring extreme resistance to potato virus Y. Theor Appl Genet 131(9): 1925–1938. doi: 10.1007/s00122-018-3123-5.

Hoerl, A.E., and R.W. Kennard. 1970. Ridge Regression: Applications to Nonorthogonal Problems. Technometrics 12(1): 69–82. doi: 10.1080/00401706.1970.10488635.

Hoopes, G., X. Meng, J.P. Hamilton, S.R. Achakkagari, F. de Alves Freitas Guesdes, et al. 2022. Phased, chromosome-scale genome assemblies of tetraploid potato reveal a complex genome, transcriptome, and predicted proteome landscape underpinning genetic diversity. Molecular Plant 15(3): 520–536. doi: 10.1016/j.molp.2022.01.003.

Jansky, S. 2009. Chapter 2 - Breeding, Genetics, and Cultivar Development. In: Singh, J. and Kaur, L., editors, Advances in Potato Chemistry and Technology. Academic Press, San Diego. p. 27–62

Jansky, S.H., and D.M. Spooner. 2018. The Evolution of Potato Breeding. Plant Breeding Reviews. John Wiley & Sons, Ltd. p. 169–214

Johansen, R.H., N. Sandar, W.G. Hoyman, and E.P. Lana. 1959. Norland a new red-skinned potato variety with early maturity and moderate resistance to common scab. American Potato Journal 36(1): 12–15. doi: 10.1007/BF02877209.

Johnston, G.R., and R.G. Rowberry. 1981. Yukon Gold: A new yellow-fleshed, medium-early, high quality table and French-fry cultivar. American Potato Journal 58(5): 241–244. doi: 10.1007/BF02853905.

Jones, C.R., T.E. Michaels, C. Schmitz Carley, C.J. Rosen, and L.M. Shannon. 2021. Nitrogen uptake and utilization in advanced fresh-market red potato breeding lines. Crop Science 61(2): 878–895. doi: 10.1002/csc2.20297.

Kaiser, N.R., S. Jansky, J.J. Coombs, P. Collins, M. Alsahlany, et al. 2021. Assessing the Contribution of Sli to Self-Compatibility in North American Diploid Potato Germplasm Using KASP^{TM} Markers. Am. J. Potato Res. doi: 10.1007/s12230-021-09821-8.

Kloosterman, B., J.A. Abelenda, M. del M.C. Gomez, M. Oortwijn, J.M. de Boer, et al. 2013. Naturally occurring allele diversity allows potato cultivation in northern latitudes. Nature 495(7440): 246–250. doi: 10.1038/nature11912.

Kraft, G.J., and W. Stites. 2003. Nitrate impacts on groundwater from irrigated-vegetable systems in a humid north-central US sand plain. Agriculture, Ecosystems & Environment 100(1): 63–74. doi: 10.1016/S0167-8809(03)00112-5.

Kuhn, M. 2008. Building Predictive Models in R Using the caret Package. Journal of Statistical Software 28: 1–26. doi: 10.18637/jss.v028.i05.

Kumar, C., P. Mubvumba, Y. Huang, J. Dhillon, and K. Reddy. 2023. Multi-Stage Corn Yield Prediction Using High-Resolution UAV Multispectral Data and Machine Learning Models. Agronomy 13(5): 1277. doi: 10.3390/agronomy13051277.

Lebourgeois, V., A. Bégué, S. Labbé, M. Houlès, and J.F. Martiné. 2012. A light-weight multi-spectral aerial imaging system for nitrogen crop monitoring. Precision Agric 13(5): 525–541. doi: 10.1007/s11119-012-9262-9.

Li, Z., X. Zhou, Q. Cheng, S. Fei, and Z. Chen. 2023. A Machine-Learning Model Based on the Fusion of Spectral and Textural Features from UAV Multi-Sensors to Analyse the Total Nitrogen Content in Winter Wheat. Remote Sensing 15(8): 2152. doi: 10.3390/rs15082152.

Liu, Y., H. Feng, J. Yue, X. Jin, Z. Li, et al. 2022. Estimation of potato above-ground biomass based on unmanned aerial vehicle red-green-blue images with different texture features and crop height. Front. Plant Sci. 13. doi: 10.3389/fpls.2022.938216.

Martins, V.S., M.H.M.L. Andrade, L.N. Padua, L.A. Miguel, C.C. Fernandes Filho, et al. 2023. Evaluating the impact of modeling the family effect for clonal selection in potato-breeding programs. Front Plant Sci 14: 1253706. doi: 10.3389/fpls.2023.1253706.

McFeeters, S.K. 1994. Estimation of wetlands and open water using Landsat Thematic Mapper data. Photogrammetric Engineering and Remote Sensing 60(3): 281–289. doi: 10.14358/PERS.60.3.281.

Meade, F., S. Byrne, D. Griffin, C. Kennedy, F. Mesiti, et al. 2020. Rapid Development of KASP Markers for Disease Resistance Genes Using Pooled Whole-Genome Resequencing. Potato Res. 63(1): 57–73. doi: 10.1007/s11540-019-09428-x.

Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics 157(4): 1819–1829. doi: 10.1093/genetics/157.4.1819.

Miller, M.D., C.A. Schmitz Carley, R.A. Figueroa, M.J. Feldman, D. Haagenson, et al. 2022. TubAR: an R Package for Quantifying Tuber Shape and Skin Traits from Images. Am. J. Potato Res. doi: 10.1007/s12230-022-09894-z.

Montesinos-López, O.A., S. Ramos-Pulido, C.M. Hernández-Suárez, B.A. Mosqueda González, F.A. Valladares-Anguiano, et al. 2023. A novel method for genomic-enabled prediction of cultivars in new environments. Frontiers in Plant Science 14. https://www.frontiersin.org/articles/10.3389/fpls.2023.1218151 (accessed 25 July 2023).

Morier, T., A.N. Cambouris, and K. Chokmani. 2015. In-Season Nitrogen Status Assessment and Yield Estimation Using Hyperspectral Vegetation Indices in a Potato Crop. Agronomy Journal 107(4): 1295–1309. doi: 10.2134/agronj14.0402.

Neilson, J.A.D., A.M. Smith, L. Mesina, R. Vivian, S. Smienk, et al. 2021. Potato Tuber Shape Phenotyping Using RGB Imaging. Agronomy 11(9): 1781. doi: 10.3390/agronomy11091781.

Ortiz, R. 2020. Genomic-Led Potato Breeding for Increasing Genetic Gains: Achievements and Outlook. Crop Breeding, Genetics and Genomics 2(2). doi: https://doi.org/10.20900/cbgg20200010.

Peng, J., K. Manevski, K. Kørup, R. Larsen, and M.N. Andersen. 2021a. Random forest regression results in accurate assessment of potato nitrogen status based on multispectral data from different platforms and the critical concentration approach. Field Crops Research 268: 108158. doi: 10.1016/j.fcr.2021.108158.

Peng, J., H. Yu, X. Wu, Y. Chen, Y. Huang, et al. 2021b. Review of unmanned aerial vehicle (UAV) applications in agriculture: A geospatial perspective. Frontiers in Plant Science 12: 674347. doi: 10.3389/fpls.2021.674347.

Pham, G.M., J.P. Hamilton, J.C. Wood, J.T. Burke, H. Zhao, et al. 2020. Construction of a chromosome-scale long-read reference genome assembly for potato. GigaScience 9(9): giaa100. doi: 10.1093/gigascience/giaa100.

Prodhomme, C., D. Esselink, T. Borm, R.G.F. Visser, H.J. van Eck, et al. 2019. Comparative Subsequence Sets Analysis (CoSSA) is a robust approach to identify haplotype specific SNPs; mapping and pedigree analysis of a potato wart disease resistance gene Sen3. Plant Methods 15(1): 60. doi: 10.1186/s13007-019-0445-5.

R Core Team. 2022. https://cran.r-project.org/mirrors.html (accessed 25 May 2023).

Rieman, G.H. 1962. Superior: A new white, medium-maturing, scabresistant potato variety with high chipping quality. American Potato Journal 39(1): 19–28. doi: 10.1007/BF02912628.

Rodríguez-Álvarez, M.X., M.P. Boer, F.A. van Eeuwijk, and P.H.C. Eilers. 2018. Correcting for spatial heterogeneity in plant breeding experiments with P-splines. Spatial Statistics 23: 52–71. doi: 10.1016/j.spasta.2017.10.003.

Rosyara, U.R., W.S.D. Jong, D.S. Douches, and J.B. Endelman. 2016. Software for Genome-Wide Association Studies in Autopolyploids and Its Application to Potato. The Plant Genome 9(2): plantgenome2015.08.0073. doi: 10.3835/plantgenome2015.08.0073.

Rouse, J.W., R.H. Haas, D.W. Deering, J.A. Schell, and J.C. Harlan. 1974. Monitoring the Vernal Advancement and Retrogradation (Green Wave Effect) of Natural Vegetation.

Silva, M.F. e, G.M. Maciel, R.B. Gallis, R.L. Barbosa, V.Q. Carneiro, et al. 2022. High-throughput phenotyping by RGB and multispectral imaging analysis of genotypes in sweet corn. Hortic. Bras. 40: 92–98. doi: 10.1590/s0102-0536-2022012.

Slater, A.T., N.O.I. Cogan, J.W. Forster, B.J. Hayes, and H.D. Daetwyler. 2016. Improving Genetic Gain with Genomic Selection in Autotetraploid Potato. The Plant Genome 9(3): plantgenome2016.02.0021. doi: https://doi.org/10.3835/plantgenome2016.02.0021.

Slater, A.T., N.O.I. Cogan, B.C. Rodoni, H.D. Daetwyler, B.J. Hayes, et al. 2017. Breeding Differently—the Digital Revolution: High-Throughput Phenotyping and Genotyping. Potato Res. 60(3): 337–352. doi: 10.1007/s11540-018-9388-x.

Slater, A.T., G.M. Wilson, N.O.I. Cogan, J.W. Forster, and B.J. Hayes. 2014. Improving the analysis of low heritability complex traits for enhanced genetic gain in potato. Theor Appl Genet 127(4): 809–820. doi: 10.1007/s00122-013-2258-7.

Song, P., J. Wang, X. Guo, W. Yang, and C. Zhao. 2021. High-throughput phenotyping: Breaking through the bottleneck in future crop breeding. The Crop Journal 9(3): 633–645. doi: 10.1016/j.cj.2021.03.015.

Spooner, D.M., J. Núñez, G. Trujillo, M. del R. Herrera, F. Guzmán, et al. 2007. Extensive simple sequence repeat genotyping of potato landraces supports a major reevaluation of their gene pool structure and classification. PNAS 104(49): 19398–19403. doi: 10.1073/pnas.0709796104.

Stefaniak, T.R., S. Fitzcollins, R. Figueroa, A.L. Thompson, C. Schmitz Carley, et al. 2021. Genotype and Variable Nitrogen Effects on Tuber Yield and Quality for Red Fresh Market Potatoes in Minnesota. Agronomy 11(2): 255. doi: 10.3390/agronomy11020255.

Stich, B., and D. Van Inghelandt. 2018. Prospects and Potential Uses of Genomic Prediction of Key Performance Traits in Tetraploid Potato. Front. Plant Sci. 9. doi: 10.3389/fpls.2018.00159.

Total Nitrogen (Plant) | Research Analytical Laboratory. https://ral.cfans.umn.edu/total-nitrogen-plant (accessed 24 December 2024).

Uitdewilligen, J.G.A.M.L., A.-M.A. Wolters, B.B. D'hoop, T.J.A. Borm, R.G.F. Visser, et al. 2013. A Next-Generation Sequencing Method for Genotyping-by-Sequencing of Highly Heterozygous Autotetraploid Potato. PLOS ONE 8(5): e62355. doi: 10.1371/journal.pone.0062355.

Vos, P.G., M.J. Paulo, R.E. Voorrips, R.G.F. Visser, and H.J. van Eck. 2017. Evaluation of LD decay and various LD-decay estimators in simulated and SNP-array data of tetraploid potato. Theoretical and Applied Genetics 130: 123–135. doi: 10.1007/s00122-016-2798-8.

Vos, P.G., J.G.A.M.L. Uitdewilligen, R.E. Voorrips, R.G.F. Visser, and H.J. van Eck. 2015. Development and analysis of a 20K SNP array for potato (Solanum tuberosum): an insight into the breeding history. Theor Appl Genet 128(12): 2387–2401. doi: ros.

Webb, R.E., D.R. Wilson, J.R. Shumaker, B. Graves, M.R. Henninger, et al. 1978. Atlantic: A new potato variety with high solids, good processing quality, and resistance to pests. American Potato Journal 55(3): 141–145. doi: 10.1007/BF02852087.

Weigle, J.L., A.E. Kehr, R.V. Akeley, and J.C. Horton. 1968. Chieftain: A red-skinned potato with attractive appearance and broad adaptability. American Potato Journal 45(8): 293–296. doi: 10.1007/BF02850285.

Xu, X., S. Pan, S. Cheng, B. Zhang, D. Mu, et al. 2011. Genome sequence and analysis of the tuber crop potato. Nature 475(7355): 189–195. doi: 10.1038/nature10158.

Xu, S., X. Xu, Q. Zhu, Y. Meng, G. Yang, et al. 2023. Monitoring leaf nitrogen content in rice based on information fusion of multi-sensor imagery from UAV. Precision Agric. doi: 10.1007/s11119-023-10042-8.

Yamashita, H., R. Sonobe, Y. Hirono, A. Morita, and T. Ikka. 2020. Dissection of hyperspectral reflectance to estimate nitrogen and chlorophyll contents in tea leaves based on machine learning algorithms. Sci Rep 10(1): 17360. doi: 10.1038/s41598-020-73745-2.

Ye, Y., L. Jin, C. Bian, J. Liu, and H. Guo. 2024. Monitoring and Optimization of Potato Growth Dynamics under Different Nitrogen Forms and Rates Using UAV RGB Imagery. Agronomy 14(10): 2257. doi: 10.3390/agronomy14102257.

Yusuf, M., M.D. Miller, T.R. Stefaniak, D. Haagenson, J.B. Endelman, et al. 2024. Genomic prediction for potato (Solanum tuberosum) quality traits improved through image analysis. The Plant Genome n/a(n/a): e20507. doi: 10.1002/tpg2.20507.

Zhao, H., X. Liu, Y. Li, H. Zhang, and J. Yang. 2020. A review of optical sensors for the measurement of soil and plant nitrogen. Sensors 20(6): 1591. doi: 10.3390/s20061591.

Zheng, H., T. Cheng, D. Li, X. Zhou, X. Yao, et al. 2018. Evaluation of RGB, Color-Infrared and Multispectral Images Acquired from Unmanned Aerial Systems for the Estimation of Nitrogen Accumulation in Rice. Remote Sensing 10(6): 824. doi: 10.3390/rs10060824.

Zou, H., and T. Hastie. 2005. Regularization and Variable Selection Via the Elastic Net. Journal of the Royal Statistical Society Series B: Statistical Methodology 67(2): 301–320. doi: 10.1111/j.1467-9868.2005.00503.x.

Zvomuya, F., C.J. Rosen, and J.C. Miller. 2002. Response of russet norkotah clonal selections to nitrogen fertilization. Amer J of Potato Res 79(4): 231–239. doi: 10.1007/BF02986355.

Appendix

Chapter 2 Supplementary Material

Table S1. Genotyping platforms with number clones

Genotyping platform	Source	Number of clones	Chips	Fresh market	comments
GeneSeek 31k array	MN		69	28	MN18 &MN19
	WI				
	ND				
	Checks				
sub total		**97**			
DarTag	MN		46	40	MN18 &MN19
	WI		2		WI17
	ND				
	Checks		1		Atlantic
sub total		**89**			
Potato 8k array	MN		1		MN
	WI		11		WI17
	ND			7	
	Checks		4	6	
sub total		**29**			
Total		215	134	81	

Table S2. Genomic estimated total genotypic values for all traits

Clone id	yield	r2_yield	sg	r2_sg	roundness	r2_round
Atlantic	352.6027	0.3077	1.0651	0.4638	0.9799	0.4605
Cascade	299.7370	0.4577	1.0599	0.4772	0.9808	0.3255
Lamoka	352.5297	0.4064	1.0645	0.3494	0.9793	0.3467
MN04844	290.0172	0.4288	1.0616	0.4624	0.9807	0.3908
MN18AF6643-10	330.4759	0.2915	1.0636	0.4492	0.9802	0.4275
MN18AF6643-12	347.6976	0.3059	1.0639	0.4717	0.9800	0.4443
MN18AF6643-13	341.1736	0.3338	1.0653	0.4772	0.9791	0.4718
MN18AF6643-7	339.7308	0.2846	1.0650	0.4409	0.9790	0.4152
MN18AF6643-9	343.9868	0.2797	1.0649	0.4465	0.9796	0.4246
MN18AF6648-10	303.2741	0.3398	1.0638	0.4951	0.9787	0.4824
MN18AF6648-14	322.4780	0.2926	1.0643	0.4296	0.9786	0.4080
MN18AF6648-8	328.3541	0.3005	1.0638	0.4409	0.9789	0.4206
MN18AF6648-9	322.0362	0.3084	1.0638	0.4522	0.9785	0.4367
MN18AF6658-5	335.4880	0.3488	1.0639	0.4791	0.9799	0.4781
MN18AF6661-2	329.5614	0.2616	1.0638	0.3923	0.9787	0.3727
MN18AF6661-9	339.7131	0.3044	1.0640	0.4492	0.9805	0.4322
MN18AF6675-2	345.7433	0.3133	1.0642	0.4664	0.9802	0.4447
MN18AF6680-12	338.7286	0.2983	1.0650	0.4383	0.9795	0.4104
MN18AF6680-8	334.2132	0.2769	1.0634	0.4315	0.9790	0.4071
MN18AF6716-2	317.4708	0.3124	1.0635	0.4418	0.9790	0.4249
MN18AF6717-10	340.6236	0.3338	1.0631	0.4664	0.9787	0.4442
MN18AF6717-2	346.8972	0.3558	1.0638	0.4931	0.9789	0.4691
MN18AF6717-3	333.6639	0.3408	1.0631	0.4815	0.9793	0.4532
MN18AF6717-6	336.5675	0.3577	1.0629	0.4916	0.9793	0.4769
MN18AF6717-7	319.4766	0.3209	1.0636	0.4309	0.9782	0.4047
MN18AF6717-9	339.4855	0.3275	1.0636	0.4664	0.9786	0.4420
MN18AF6718-1	328.5124	0.3292	1.0638	0.4596	0.9796	0.4194
MN18AF6720-2	308.4202	0.3160	1.0639	0.4764	0.9796	0.4534
MN18AF6722-2	326.4288	0.3060	1.0646	0.4625	0.9786	0.4488
MN18AF6724-5	324.4300	0.3084	1.0642	0.4283	0.9784	0.4002
MN18AF6724-8	318.2678	0.3072	1.0642	0.4243	0.9790	0.3931
MN18AF6724-9	314.3965	0.2919	1.0643	0.4177	0.9790	0.3865
MN18AF6725-1	329.0308	0.3178	1.0645	0.4215	0.9791	0.3945

MN18AF6725-2	326.9438	0.3008	1.0634	0.4409	0.9802	0.4156
MN18AF6726-2	318.6541	0.3027	1.0634	0.4495	0.9802	0.4306
MN18AF6726-4	335.8022	0.2902	1.0633	0.4400	0.9802	0.4240
MN18AF6728-7	340.3972	0.3041	1.0639	0.4358	0.9800	0.4394
MN18AF6729-1	310.7886	0.2733	1.0627	0.4044	0.9804	0.3876
MN18AF6730-3	325.8182	0.3141	1.0626	0.4526	0.9808	0.4357
MN18AF6730-5	336.4202	0.3123	1.0636	0.4541	0.9808	0.4488
MN18AF6730-6	322.1080	0.3792	1.0635	0.6347	0.9792	0.6165
MN18AF6734-1	315.5821	0.2966	1.0627	0.4452	0.9805	0.4232
MN18AF6745-4	326.5467	0.2613	1.0641	0.4060	0.9797	0.3609
MN18SR00045-1	379.5941	0.0961	1.0658	0.1048	0.9845	0.2471
MN18SR00045-3	390.9099	0.0873	1.0663	0.0931	0.9822	0.2506
MN18TX17730-8	376.4728	0.2259	1.0663	0.2261	0.9834	0.2152
MN18TX17748-1	378.6157	0.3557	1.0637	0.3390	0.9818	0.3106
MN18W17037-19	375.4862	0.3825	1.0654	0.3548	0.9860	0.3277
MN18W17037-2	375.5263	0.3610	1.0657	0.3315	0.9851	0.3029
MN18W17037-21	342.9566	0.4772	1.0669	0.4247	0.9840	0.4046
MN18W17037-24	374.0852	0.4088	1.0685	0.3831	0.9834	0.3348
MN18W17037-26	329.5604	0.5017	1.0673	0.4436	0.9841	0.4323
MN18W17037-27	357.6144	0.4639	1.0672	0.3828	0.9834	0.3900
MN18W17037-32	433.4639	0.3402	1.0634	0.3260	0.9823	0.2813
MN18W17037-34	375.1383	0.5179	1.0678	0.4271	0.9830	0.4524
MN18W17037-36	368.6880	0.3774	1.0668	0.3533	0.9859	0.3366
MN18W17037-9	334.0638	0.2941	1.0635	0.4474	0.9798	0.4339
MN18W17039-1	374.4905	0.4036	1.0686	0.3766	0.9829	0.3587
MN18W17039-12	338.8337	0.2860	1.0641	0.4211	0.9802	0.4033
MN18W17039-13	326.3960	0.2795	1.0638	0.4296	0.9797	0.4156
MN18W17039-27	378.3293	0.3961	1.0704	0.3772	0.9852	0.3507
MN18W17039-5	401.3807	0.5431	1.0713	0.4637	0.9847	0.4660
MN18W17043-11	406.6943	0.3362	1.0674	0.3046	0.9839	0.2694

MN18W17043-12	467.8000	0.4826	1.0703	0.4043	0.9830	0.4280
MN18W17043-14	349.8619	0.2756	1.0640	0.4165	0.9797	0.3760
MN18W17043-17	333.1699	0.3478	1.0640	0.4670	0.9805	0.4618
MN18W17043-2	384.6600	0.4871	1.0709	0.4187	0.9842	0.4561
MN18W17043-3	372.0544	0.4630	1.0691	0.4282	0.9856	0.4209
MN18W17043-6	409.3341	0.4921	1.0686	0.4172	0.9836	0.4499
MN18W17043-7	435.0984	0.3683	1.0684	0.3414	0.9827	0.3180
MN18W17043-8	442.1413	0.3187	1.0647	0.2955	0.9830	0.0951
MN18W17052-15	315.8313	0.4238	1.0686	0.4047	0.9874	0.3217
MN18W17052-4	378.5668	0.5256	1.0729	0.4580	0.9861	0.4905
MN18W17052-6	369.7262	0.5169	1.0708	0.4328	0.9850	0.4562
MN18W17055-4	325.1040	0.4507	1.0668	0.4220	0.9873	0.3866
MN18W17057-1	315.0761	0.4200	1.0686	0.3949	0.9860	0.3838
MN18W17057-3	327.7788	0.4537	1.0670	0.4226	0.9873	0.3865
MN18W17057-5	342.2037	0.4655	1.0670	0.3392	0.9841	0.3873
MN18W17065-5	370.1792	0.3058	1.0650	0.2806	0.9822	0.2119
MN18W17065-7	325.2802	0.2920	1.0633	0.4471	0.9793	0.4282
MN18W17092-7	331.5089	0.2918	1.0641	0.4465	0.9803	0.4042
MN19AF6866-1	330.8516	0.4141	1.0656	0.3757	0.9821	0.3768
MN19AF6866-10	331.9578	0.3763	1.0659	0.4175	0.9828	0.3875
MN19AF6866-11	335.0376	0.4333	1.0658	0.4353	0.9832	0.4059
MN19AF6866-13	332.6610	0.3734	1.0658	0.4222	0.9827	0.3952
MN19AF6866-14	337.6635	0.4400	1.0658	0.4406	0.9831	0.4134
MN19AF6866-4	316.7852	0.3054	1.0670	0.3603	0.9834	0.3445
MN19AF6867-3	350.7831	0.3724	1.0669	0.3318	0.9797	0.2576
MN19AF6869-12	335.5019	0.3228	1.0658	0.3423	0.9854	0.2963
MN19AF6869-19	307.4652	0.3168	1.0668	0.3382	0.9839	0.2848
MN19AF6869-20	336.1118	0.2333	1.0656	0.2499	0.9844	0.2110

MN19AF6869-21	321.9578	0.3329	1.0655	0.3457	0.9847	0.3003
MN19AF6869-5	324.8727	0.3120	1.0657	0.3353	0.9846	0.2862
MN19AF6869-6	304.2740	0.3473	1.0663	0.3644	0.9843	0.2414
MN19AF6869-7	324.9331	0.4364	1.0673	0.4020	0.9852	0.3922
MN19AF6892-10	383.4704	0.3051	1.0691	0.3165	0.9831	0.2712
MN19AF6892-2	365.7489	0.3076	1.0693	0.3259	0.9843	0.2804
MN19AF6892-9	392.1779	0.4010	1.0706	0.3768	0.9829	0.3601
MN19TX18010-1	380.6668	0.2659	1.0643	0.2921	0.9822	0.2428
MN19TX18014-1	321.7702	0.2534	1.0650	0.2604	0.9832	0.2150
MN19TX18014-4	345.8491	0.2321	1.0655	0.2421	0.9835	0.0922
MN19TX18031-1	390.7593	0.2494	1.0649	0.2805	0.9826	0.2337
MN19TX18032-1	360.9490	0.4203	1.0632	0.3887	0.9833	0.3933
MN19TX18032-10	328.1835	0.4138	1.0615	0.3869	0.9832	0.3935
MN19TX18032-5	378.6297	0.4040	1.0630	0.3771	0.9832	0.3779
MN19TX18032-6	363.2698	0.2862	1.0622	0.3065	0.9828	0.2730
MN19TX18032-9	376.0473	0.3176	1.0630	0.3344	0.9830	0.2965
MN19TX18054-2	369.5377	0.3291	1.0647	0.3112	0.9824	0.2943
MN19TX18093-1	499.7458	0.3897	1.0662	0.3590	0.9810	0.3493
MN19TX18120-1	377.6338	0.3216	1.0661	0.3008	0.9827	0.3096
MN19TX18171-1	366.8509	0.2075	1.0653	0.2170	0.9820	0.1825
MN19TX18171-6	369.2321	0.2156	1.0653	0.2378	0.9814	0.1940
MN19TX18188-1	341.0656	0.3916	1.0607	0.3706	0.9833	0.3614
MN19TX18211-1	414.5654	0.3414	1.0648	0.3354	0.9814	0.3227
MN19TX18212-7	375.1944	0.3729	1.0628	0.3429	0.9826	0.3299
MN19TX18212-8	403.6396	0.2748	1.0639	0.2940	0.9829	0.2433
MN19TX18260-1	371.8141	0.3529	1.0674	0.3075	0.9829	0.3168
MN19TX18280-2	396.7872	0.4044	1.0638	0.3849	0.9815	0.3168
MN19TX18304-1	447.0289	0.3769	1.0659	0.3319	0.9799	0.3358
Snowden	364.7682	0.3260	1.0661	0.2927	0.9819	0.2844

Superior	296.4383	0.4617	1.0609	0.4765	0.9808	0.3917
W17037-3	368.7165	0.2935	1.0671	0.2793	0.9834	0.2427
W17039-31	342.5609	0.2857	1.0642	0.4385	0.9800	0.4084
W17039-7	416.4603	0.3379	1.0687	0.3167	0.9832	0.2873
W17043-37	346.0288	0.3145	1.0642	0.4922	0.9801	0.4654
W17049-10	399.1712	0.2211	1.0666	0.2171	0.9814	0.1777
W17060-22	406.9902	0.3418	1.0690	0.3197	0.9835	0.2867
W17065-21	367.2267	0.2238	1.0663	0.2171	0.9824	0.1720
W17066-11	375.0221	0.2817	1.0657	0.2672	0.9820	0.1596
W17066-34	379.9569	0.2943	1.0665	0.2834	0.9825	0.2169
W17067-1	413.3053	0.2465	1.0685	0.2378	0.9824	0.1983
W17067-13	395.6945	0.2041	1.0672	0.1986	0.9825	0.1613
W17AF6670-1	399.6870	0.2413	1.0682	0.2339	0.9812	0.1970
W17AF6685-2	386.1900	0.2119	1.0656	0.2047	0.9813	0.1682

Table S3. List of clones based on selection index with equal weights

Clone ID	value	rank
MN18W17052-4	1.46465093	1
MN18W17043-12	1.46383934	2
MN18W17039-5	1.46330839	3
MN18W17039-27	1.46228449	4
MN18W17043-2	1.46209292	5
MN18W17052-6	1.46206908	6
MN18W17043-3	1.46146164	7
W17060-22	1.46116807	8
MN18W17043-7	1.46116406	9
MN19AF6892-9	1.46116139	10
W17039-7	1.46105178	11
MN18W17043-6	1.46103515	12
MN19TX18093-1	1.46079365	13
MN19AF6892-2	1.46039004	14
MN18W17043-11	1.46036571	15
MN18W17052-15	1.46033421	16
W17067-1	1.46019074	17
MN19AF6892-10	1.46006016	18
MN18W17037-36	1.46002237	19
MN18W17057-3	1.45963456	20
MN18W17037-24	1.45948549	21
MN18W17055-4	1.45943931	22

MN18W17037-19	1.45937294	23
MN18W17057-1	1.45927897	24
MN18W17039-1	1.45920063	25
MN18W17043-8	1.45919614	26
MN18W17037-2	1.45897367	27
W17067-13	1.45875334	28
MN18W17037-34	1.45874611	29
MN18SR00045-1	1.45871364	30
W17AF6670-1	1.45862252	31
W17037-3	1.45834039	32
MN19TX18260-1	1.45831525	33
MN19AF6869-7	1.4581878	34
MN18TX17730-8	1.45808172	35
MN18W17037-27	1.45799793	36
MN19TX18304-1	1.45786875	37
MN18W17057-5	1.45784468	38
MN19AF6869-12	1.45779939	39
MN18SR00045-3	1.45775986	40
W17049-10	1.4576769	41
MN18W17037-21	1.45765687	42
W17066-34	1.45765584	43
MN18W17037-26	1.45754367	44
MN19TX18120-1	1.45749881	45
MN18W17037-32	1.45749654	46
MN19TX18212-8	1.45714931	47
MN19TX18031-1	1.45708822	48
W17065-21	1.4570383	49
MN19TX18211-1	1.45702741	50
MN19AF6869-20	1.45689739	51
MN19AF6869-5	1.45665616	52
W17066-11	1.45664068	53
MN19TX18014-4	1.45653408	54
MN19AF6869-21	1.45646722	55
Snowden	1.45645185	56
W17AF6685-2	1.45644648	57
MN19AF6866-4	1.45638101	58
MN19AF6869-19	1.45623858	59
MN18W17065-5	1.45615892	60
MN19AF6866-14	1.45612599	61

MN19AF6869-6	1.45608182	62
MN19TX18010-1	1.45606777	63
MN19TX18054-2	1.45606754	64
MN19AF6866-11	1.45606577	65
MN19TX18171-1	1.45605526	66
MN19TX18032-5	1.45580681	67
MN19AF6866-10	1.45576816	68
MN19TX18280-2	1.45576289	69
MN19TX18171-6	1.45567924	70
MN19AF6866-13	1.45565102	71
MN19TX18032-9	1.45562024	72
MN19TX18032-1	1.45539668	73
MN18TX17748-1	1.45525805	74
MN19TX18212-7	1.45514785	75
MN19TX18014-1	1.4550591	76
MN19AF6866-1	1.45501732	77
MN19AF6867-3	1.45481967	78
MN19TX18032-6	1.45450426	79
Atlantic	1.45376317	80
MN18AF6675-2	1.45319575	81
MN18AF6643-9	1.45316861	82
W17043-37	1.45313059	83
MN18AF6661-9	1.45307414	84
MN19TX18188-1	1.45298512	85
Lamoka	1.45298415	86
MN19TX18032-10	1.45297089	87
MN18AF6680-12	1.45295537	88
MN18AF6643-13	1.45293267	89
W17039-31	1.45292032	90
MN18AF6730-5	1.45290005	91
MN18AF6643-12	1.45287289	92
MN18W17039-12	1.45286812	93
MN18W17043-14	1.45285264	94
MN18W17043-17	1.45279104	95
MN18AF6728-7	1.45268925	96
MN18W17092-7	1.4526169	97
MN18AF6643-7	1.45255801	98
MN18AF6658-5	1.45239354	99
MN18AF6726-4	1.452228	100

MN18AF6643-10	1.4521977	101
MN18AF6745-4	1.45203519	102
MN18W17037-9	1.45201537	103
MN18AF6717-2	1.45199933	104
MN18AF6725-1	1.45195791	105
MN18AF6725-2	1.45190636	106
MN18AF6730-3	1.45186544	107
MN18AF6718-1	1.45185286	108
MN18W17039-13	1.45184796	109
MN18AF6726-2	1.45166095	110
MN18AF6722-2	1.45149341	111
MN18AF6717-9	1.45139763	112
MN18AF6648-8	1.45137337	113
MN18AF6717-3	1.4513675	114
MN18AF6717-6	1.45132372	115
MN18AF6680-8	1.45130905	116
MN18AF6734-1	1.45130114	117
MN18AF6724-8	1.45127745	118
MN18AF6661-2	1.45122129	119
MN18AF6648-14	1.45121612	120
MN18AF6724-9	1.45120903	121
MN18AF6720-2	1.45117485	122
MN18AF6717-10	1.45116404	123
MN18W17065-7	1.45113779	124
MN18AF6730-6	1.45110267	125
MN18AF6724-5	1.45107714	126
MN18AF6729-1	1.45107264	127
MN18AF6716-2	1.45080297	128
MN18AF6648-9	1.45077689	129
MN18AF6717-7	1.45033426	130
MN18AF6648-10	1.45024051	131
MN04844	1.44979686	132
Superior	1.44963539	133
Cascade	1.44906224	134

Table S4. List of clones based on selection index with different weights

Clone ID	value	rank
MN18W17052-4	1.466167	1
MN18W17043-12	1.46613043	2
MN18W17039-5	1.4649907	3
MN18W17043-2	1.46367224	4
MN18W17039-27	1.46358748	5
MN18W17052-6	1.46340682	6
MN19TX18093-1	1.4630715	7
MN18W17043-7	1.46300534	8
MN19AF6892-9	1.4629385	9
W17060-22	1.46275515	10
W17039-7	1.46270609	11
MN18W17043-6	1.4625804	12
MN18W17043-3	1.46248257	13
W17067-1	1.46192465	14
MN18W17043-11	1.46168284	15
MN19AF6892-2	1.46157987	16
MN19AF6892-10	1.46154204	17
MN18W17037-24	1.46077769	18
MN18W17037-36	1.46065964	19
MN18W17052-15	1.46060754	20
MN18W17039-1	1.46056995	21
MN18W17043-8	1.46054641	22
W17AF6670-1	1.46038327	23
W17067-13	1.4601644	24
MN18W17037-34	1.46001077	25
MN18W17037-19	1.45986831	26
MN19TX18304-1	1.45986794	27
MN18W17057-3	1.4597972	28
MN18W17057-1	1.45976095	29
MN18W17037-2	1.45965237	30
MN18W17055-4	1.45956181	31
MN19TX18260-1	1.45951924	32
MN18SR00045-1	1.45950768	33
W17037-3	1.45940278	34
W17049-10	1.45918614	35
MN18TX17730-8	1.45908694	36
MN18SR00045-3	1.45906941	37

MN18W17037-27	1.45898312	38
W17066-34	1.45885529	39
MN18W17037-32	1.458708	40
MN19AF6869-7	1.45867947	41
MN19TX18120-1	1.4585871	42
MN18W17057-5	1.4585843	43
MN19TX18211-1	1.45842533	44
MN18W17037-21	1.4584087	45
MN18W17037-26	1.458232	46
MN19AF6869-12	1.45814273	47
MN19TX18031-1	1.45813951	48
W17065-21	1.45812558	49
MN19TX18212-8	1.45812024	50
W17066-11	1.4577625	51
W17AF6685-2	1.45773618	52
Snowden	1.45757386	53
MN19AF6869-20	1.45737117	54
MN19TX18014-4	1.4571809	55
MN18W17065-5	1.45712474	56
MN19TX18171-1	1.45707247	57
MN19AF6866-4	1.45704687	58
MN19AF6869-5	1.45702449	59
MN19TX18010-1	1.45701805	60
MN19TX18054-2	1.45696033	61
MN19TX18280-2	1.45687033	62
MN19AF6866-14	1.45683113	63
MN19TX18171-6	1.45680719	64
MN19AF6869-21	1.45676976	65
MN19AF6866-11	1.45673272	66
MN19AF6869-19	1.45673091	67
MN19AF6866-10	1.45648468	68
MN19TX18032-5	1.45642469	69
MN19AF6869-6	1.45641318	70
MN19AF6866-13	1.45637639	71
MN19AF6867-3	1.45625763	72
MN19TX18032-9	1.45623565	73
MN18TX17748-1	1.45616005	74
MN19TX18032-1	1.45590003	75
MN19TX18212-7	1.45578268	76

MN19AF6866-1	1.45577777	77
MN19TX18014-1	1.45551649	78
MN19TX18032-6	1.45495158	79
Atlantic	1.45495058	80
MN18AF6643-9	1.45431023	81
MN18AF6643-13	1.45418283	82
Lamoka	1.45418134	83
MN18AF6675-2	1.45416968	84
W17043-37	1.45410655	85
MN18AF6680-12	1.45409577	86
MN18AF6661-9	1.45392051	87
MN18W17043-14	1.45390756	88
W17039-31	1.45388778	89
MN18AF6643-12	1.45384578	90
MN18W17039-12	1.45378677	91
MN18AF6643-7	1.45377187	92
MN18AF6730-5	1.45363031	93
MN18AF6728-7	1.45360438	94
MN18W17043-17	1.45360275	95
MN18W17092-7	1.45345125	96
MN18AF6658-5	1.45329691	97
MN18AF6717-2	1.45312639	98
MN18AF6726-4	1.45301307	99
MN18AF6725-1	1.45301032	100
MN19TX18188-1	1.45300336	101
MN19TX18032-10	1.45300039	102
MN18AF6643-10	1.45297679	103
MN18AF6745-4	1.45292729	104
MN18W17037-9	1.45285375	105
MN18AF6718-1	1.45272236	106
MN18W17039-13	1.45269881	107
MN18AF6725-2	1.45262877	108
MN18AF6722-2	1.45261178	109
MN18AF6717-9	1.45248309	110
MN18AF6730-3	1.45238409	111
MN18AF6648-8	1.45235335	112
MN18AF6726-2	1.45232079	113
MN18AF6648-14	1.45226088	114
MN18AF6680-8	1.4522592	115

| MN18AF6661-2	1.45224521	116
MN18AF6717-3	1.45223843	117
MN18AF6724-8	1.45221878	118
MN18AF6717-6	1.45217849	119
MN18AF6717-10	1.45217502	120
MN18AF6724-5	1.45214907	121
MN18AF6724-9	1.45213316	122
MN18W17065-7	1.45196608	123
MN18AF6730-6	1.4519422	124
MN18AF6720-2	1.45192356	125
MN18AF6734-1	1.45180076	126
MN18AF6648-9	1.45175531	127
MN18AF6716-2	1.45164943	128
MN18AF6729-1	1.45154686	129
MN18AF6717-7	1.45131264	130
MN18AF6648-10	1.4510504	131
MN04844	1.44992127	132
Superior	1.44971015	133
Cascade	1.44901502	134
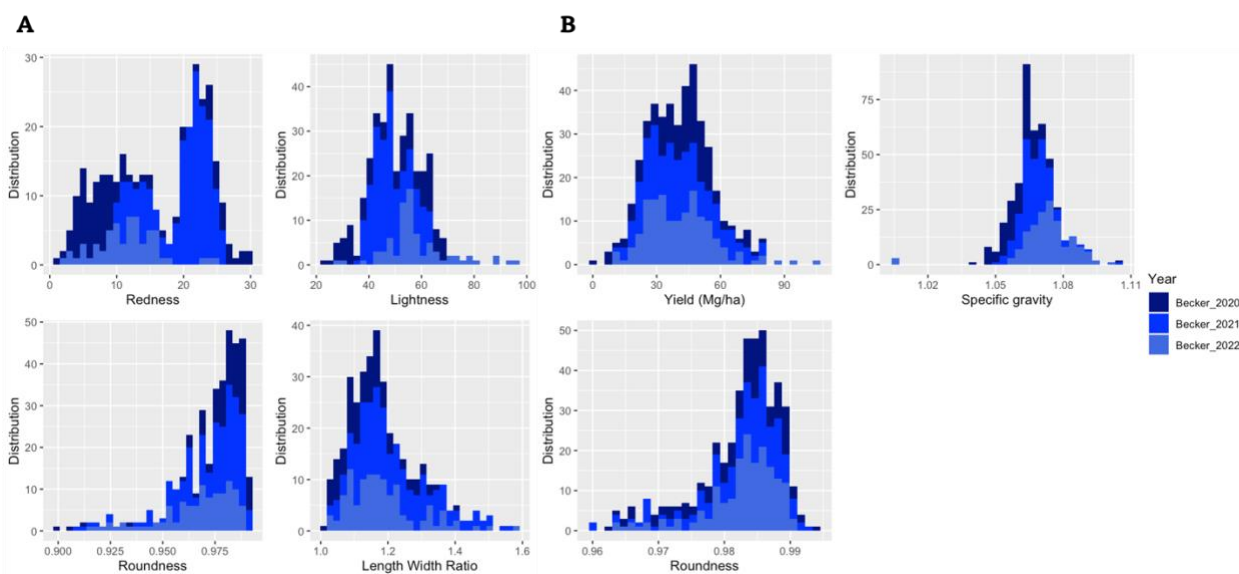
Figure S1: Distribution of all phenotypic traits A) Fresh market B) Chips.

Chapter 3 Supplementary Material

Table S1. Spectral variable features based on combined stages (1-5) versus selected stages

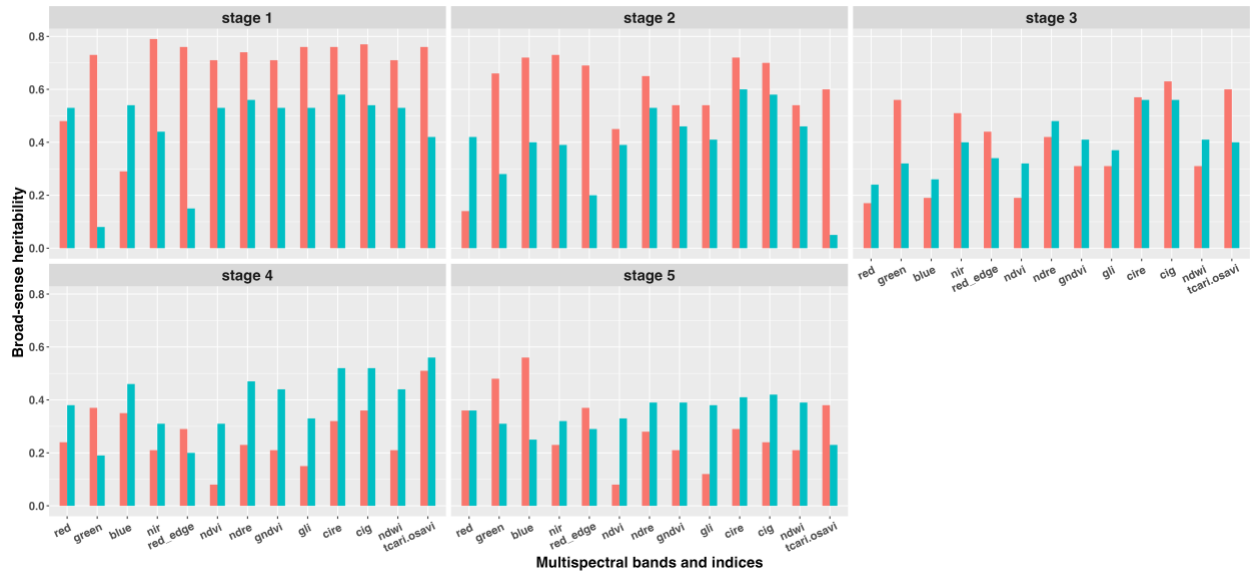| Features (Bands and Indices)	Combined stages (65)	Simulation annealing (SA, 26)	Selection by filtering (SF, 46)
Red	1,2,3,4,5	2,3,4	1,4
Green	1,2,3,4,5	2,4	1,2,3,4,5
Blue	1,2,3,4,5	2,3,4,5	1,4,5
NIR	1,2,3,4,5	2	1,2,3,4
Red Edge	1,2,3,4,5	2,3	1,2,3,4
NDVI	1,2,3,4,5	3	1,2,3,4,5
GNDVI	1,2,3,4,5	3,4,5	2,3
NDRE	1,2,3,4,5	1,4,5	1,2,3,4,5
CIRE	1,2,3,4,5	1	1,2,3,4,5
CIG	1,2,3,4,5	1	2,4
GLI	1,2,3,4,5	5	2,3,4,5
NDWI	1,2,3,4,5	2,4	2,3
TCARI.OSAVI	1,2,3,4,5	4	1,2,5

Figure S1. Broad-sense heritability of multispectral bands and indices within the chips market
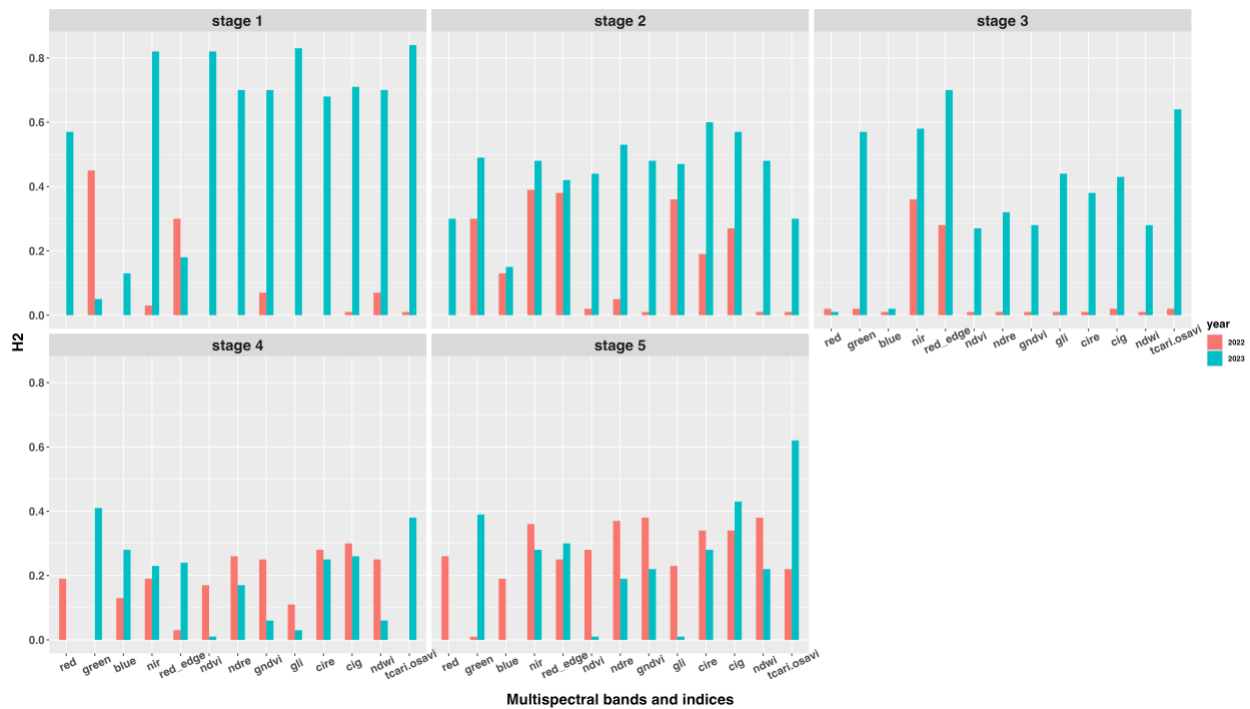

Figure **S2**. Broad-sense heritability of multispectral bands and indices within the fresh market
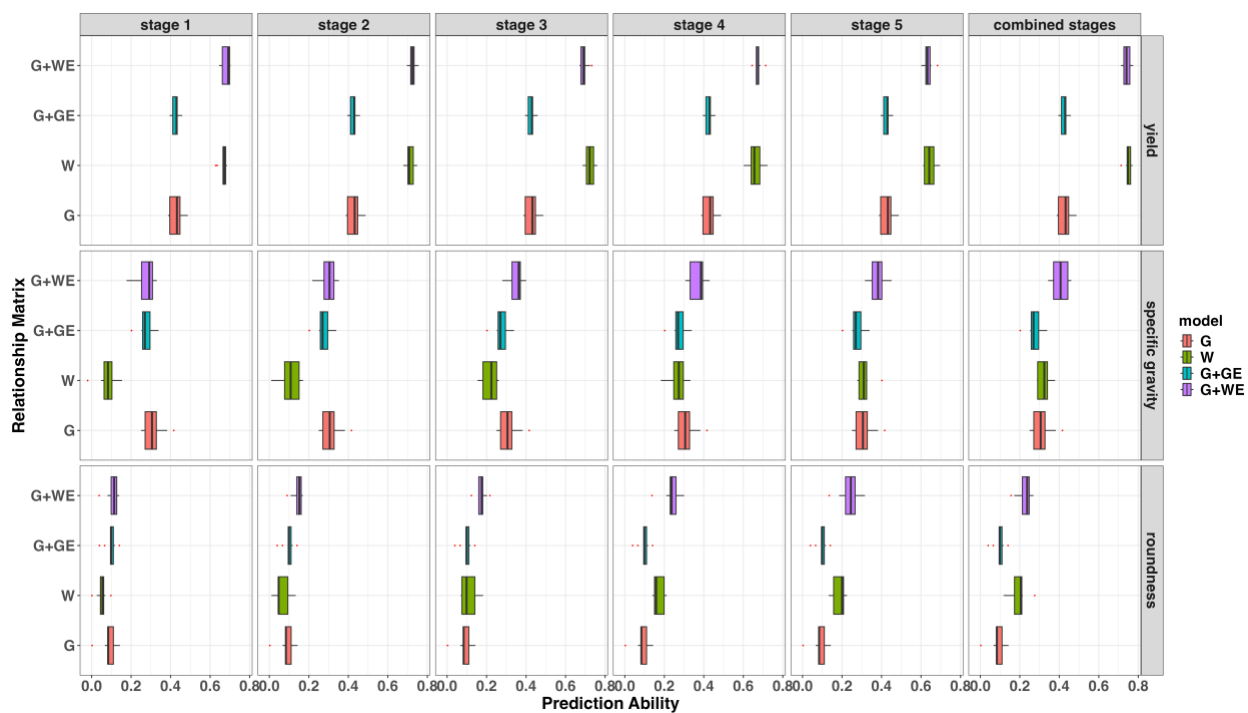
Figure S3. Prediction ability for traits within the chips market clones across growth stages using multispectral bands and Indices
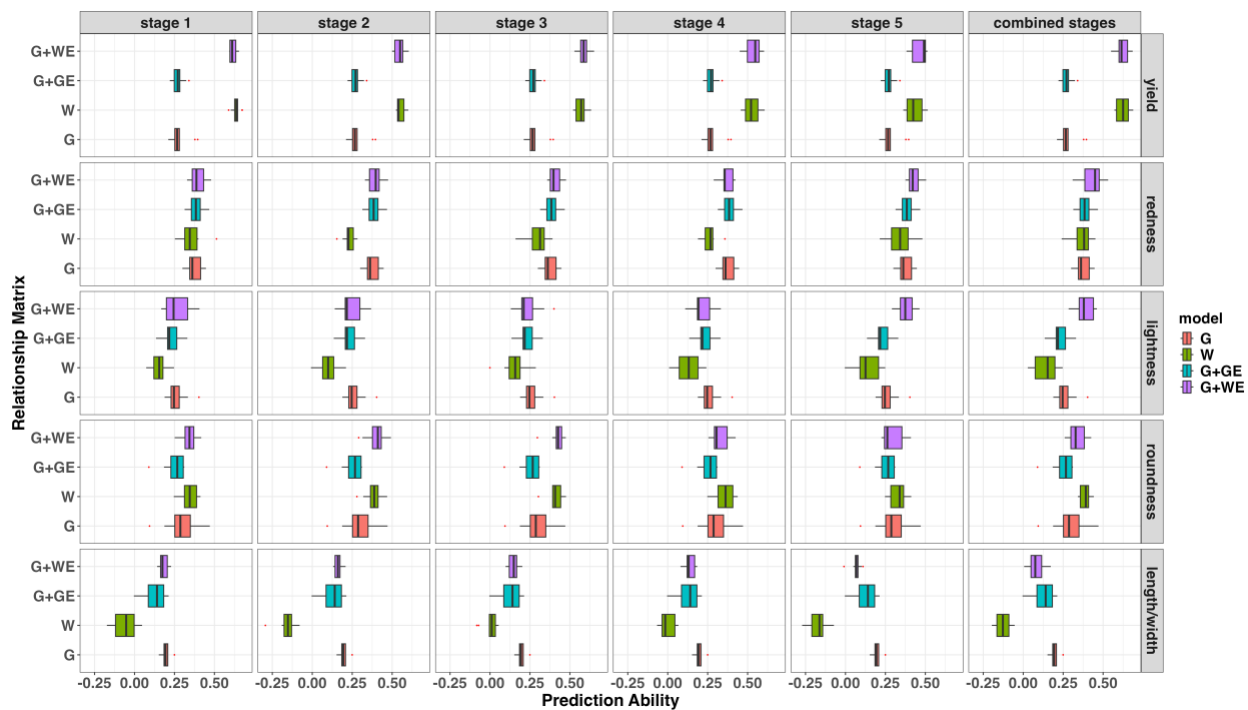
Figure S4. Prediction ability for traits within the fresh market clones across growth stages using multispectral bands and Indices
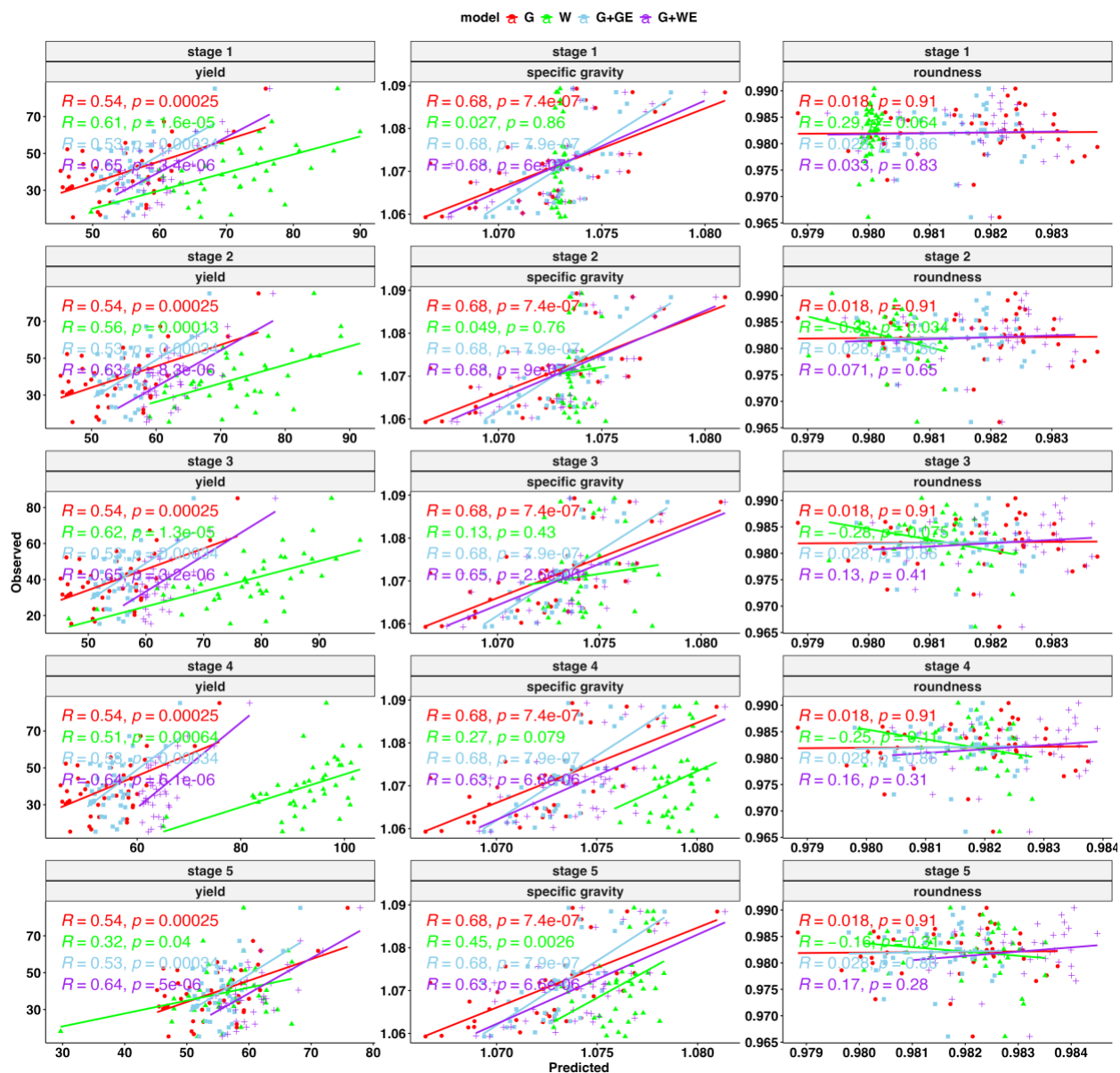

Figure S5. Prediction ability for 2022 using 2023 within the chips market clones across growth stages using multispectral bands and Indices

Chapter 4 Supplementary Material

Table S1a. Significant variable features across growth stages fresh market

Features (Bands and Indices)	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Combined stages
Red	√	√				3,5
Green	√	√				
Blue	√	√				2,3,5
NIR			√	√		3,4
Red Edge	√	√		√		1
NDVI	√			√	√	
GNDVI	√				√	
NDRE	√	√	√		√	
CIRE	√	√	√		√	
CIG	√	√			√	2
GLI		√	√	√	√	
NDWI	√					
TCARI.OSAVI					√	3

Table S1b. Significant variable features across growth stages for chips market

Features (Bands and Indices)	Stage1	Stage 2	Stage 3	Stage 4	Stage 5	Combined stages
Red					√	
Green	√		√	√	√	
Blue			√	√	√	
NIR	√	√	√	√	√	√
Red Edge			√	√	√	√
NDVI	√	√		√	√	
GNDVI	√	√				√
NDRE	√		√	√		

CIRE	√		√	
CIG			√	√
GLI	√	√		√
NDWI		√		
TCARI.OSAVI				